교육과정평가연구
The Journal of Curriculum & Evaluation
2006, Vol. 9, No. 1, pp. 275～298

가교문항 수직적 동등화 방법을 통한 중학생의 학년별 수학 학력 변화 탐색

김재철(한 남 대 학 교 교 수)
김선희(한국교육과정평가원 연구원)

≪ 요 약 ≫

본 연구에서는 중학교 3년 동안 학생들의 수학 학력 변화가 어떠한 변화양상을 보이는지 규명하고 있다. 문항 수는 학년별로 37개였으며 동등화를 위해서 1학년과 2학년 공통문항을 19개, 2학년과 3학년 공통문항을 16개 포함시켰다. 이 중에서 8문항은 3개 학년 공통문항이었다. 연구 대상은 서울에 위치한 3개의 중학교에 재학 중인 학생으로 1학년 472명, 2학년 511명, 3학년 496명 등 총 1,479명이었다. 연구문제는 다음과 같았다. 첫째, 중학교에서 학년이 올라감에 따라 정답률이 높아지는 문항과 그렇지 않은 문항의 특성은 어떠한가? 둘째, 중학교에서 학년이 올라감에 따라 수학 학력의 평균적인 변화는 어떠한 양상을 보이는가? 셋째, 수직적 동등화 과정에서 추정된 개인별 동등화 점수는 동등화 방법에 따라서 어떠한 차이가 있는가? 최종적으로 연구에 활용된 동등화 방법은 터커의 선형동등화, 능력모수동등화, 진점수동등화, 역진점수동등화 등 4가지였다. 분석과정에서 밝혀진 연구 결과 및 이를 바탕으로 내린 결론은 다음과 같았다. 첫째, 공통문항의 경우 학년이 올라감에 따라 정답률이 높아지는 경향이 뚜렷하게 나타났다. 둘째, 학년이 올라갈수록 정답률이 상승하는 문항은 그 내용이 위계성을 가지고 있어서 보다 높은 수준에서 다시 한 번 다루어지는 문항이었다. 셋째, 수직적 동등화 상황에서 점수대별 동등화 표준오차가 가장 적은 방법은 '터커의 선형동등화'였다. 넷째, 수학 학력 평가에서 공통문항을 이용하여 학년 간 수직적 동등화를 수행함으로써 학년이 올라감에 따른 절대적인 학력 변화의 측정 가능성을 확인할 수 있었다. 다섯째, 동등화 방법에 따라서 추정된 동등화 점수 간에 다소의 차이를 발견할 수 있었다. 본 연구와 관련한 향후 연구 과제를 제시하면 다음과 같다. 첫째, 동등화를 활용하여 절대적인 학력 변화를 탐색하려는 연구를 확장하여 개인별 변화를 추정하는 종단 연구를 수행할 필요가 있다. 둘째, 학년 간 변화를 도출하기 위한 방법으로써 공통문항을 이용하는 '가교검사 설계' 대신 공통피험자를 두는 '가교피험자 설계'의 활용 가능성을 연구할 필요가 있다.

주제어 : 가교문항, 수직적 동등화, 수학 학력, 변화양상

Ⅰ. 연구의 필요성 및 목적

교육은 '바람직한 인간행동의 계획적인 변화(정범모, 1971)'를 그 목적으로 하고 있다. 교육적 노력이 투입되기 전과 투입된 후 피교육자의 여러 가지 행동 특성의 변화를 우리는 흔히 '성장'이라고 정의한다. 그것은 기억력․추리력 등과 같은 일반적인 능력의 변화일 수도 있고, 사회의 한 일원으로서 원활한 생활을 영위할 수 있기 위한 경험의 습득일 수도 있다. 경우에 따라서는 지식의 축적일 수도 있지만, 이와는 대조적으로 지식을 받아들이는 방법의 습득을 의미하기도 한다. 교육을 어떻게 정의하든 그 공통된 주장은 인간을 변화․성장시키는 과정이다. 이러한 측면에서 본다면 해가 바뀜에 따라서, 또는 학년이 올라가면서 학생의 학력의 변화에 대해 관심을 가지고 이를 탐색하려는 노력은 매우 중요한 의미를 가진다.

학력의 변화를 측정하기 위해서는 먼저 학력이 무엇인지가 분명해야 한다. 사전에서는 학력을 '학교교육 등의 학습이나 훈련에 의하여 획득한 지적(知的) 적응 능력'으로 정의하고 있다(두산세계대백과사전, 2006). 학교교육을 통해 학생들은 지적 능력이 성장할 것으로 기대되며, 그것은 바로 학력의 변화를 뜻하는 것이다. 지금까지 많은 연구들은 학력을 성취도와 같은 의미로 해석하고, 두 가지 용어를 혼재하여 사용해 왔다(김동화, 김준홍, 1999). 성취도는 교육을 통해 교육목표를 얼마나 달성했는지에 대한 것으로(정구향 외, 2004), 수학과의 교육목표는 교육내용과 학생들에게 기대하는 인지적 능력, 정의적 태도 등으로 구성되어 있다(교육부, 1997). 따라서 수학에서의 성취도 평가는 학생들이 어떤 지식을 갖고 있는지, 지적 능력의 수준은 어떠한지, 정의적 태도는 어떠한지를 평가해야 한다. 하지만 본 연구는 학생들의 정의적 태도가 아닌 지적 능력을 학력이라 보기 때문에, 성취도 평가 중에서 지적 영역에 대한 평가를 통하여 학생들의 학력을 측정하려 한다. 수학에서의 지적 영역은 계산, 개념 이해, 문제해결, 추론 등을 말하며, 이러한 능력은 수학 내용을 기반으로 하기 때문에 문항의 구성은 수학 내용을 기초로 해야 한다.

학생들의 학력 변화를 탐색하려는 연구로 현재 한국교육과정평가원은 국가수준 학업성취도 평가를 실시하여 매년 변화를 탐색하는 노력을 기울이고 있다. 그러나 국가수준 학업성취도 평가에서의 추이 분석은 특정 학년에 대한 연도 간 비교에 목적이 있기 때문에 학년이 올라가면서 나타날 수 있는 학력 변화에 대해서는 해답을 제공해 주지 못한다. 이에 본 연구는 중학생을 대상으로 학년이 올라감에 따른 수학 학력 변화를 탐색하려 하며, 이를 위해 가교문항 수직적 동등화 방법을 사용하려 한다.

두 집단의 능력을 비교하기 위해서는 기본적으로 동일한 검사도구를 두 집단에 동시에 적용해야 한다. 그러나 서로 다른 학년의 학력을 비교하는 경우에는 두 집단에게 동일한 검사를 동시에 적용하는 설계는 문제점을 수반한다. 동일한 검사만을 활용한다면 학년이 올라가

면서 새롭게 배우게 되는 교육내용을 반영할 수 없는데, 수학 학력은 지식 내용을 기반으로 하기 때문에 내용과 행동 모두에 근거한 문항으로 측정되어야 한다. 그래서 학년이 서로 다른 두 집단의 능력을 비교하기 위해서는 공통문항(일명 가교문항, anchor item)과 더불어 해당 학년에서 새롭게 학습한 교육내용을 반영한 학년별 고유문항을 포함시켜야 한다. 이때 공통문항은 두 학년에서의 피험자 점수를 동일한 척도로 변환하는 데 활용된다. 이처럼 서로 다른 검사 도구를 활용한 두 집단의 점수를 동일한 척도로 변환하는 측정학적인 노력을 ‘동등화(equating)’라고 하는데, 특히 학년 간 학력을 비교하려는 본 연구에서와 같이 동등화하고자 하는 두 집단의 능력이 현저히 차이가 날 때는 ‘수직적 동등화(vertical equating)’를 사용한다. 수직적 동등화는 동일한 특성을 재고 있지만 수준이 서로 다른 집단이 치른 검사 점수의 비교를 가능하게 하는 방법이다(Linn, 1993; 부재율 & 서동엽, 2000; 성태제, 2000). 수직적 동등화라는 용어에 대해서는 학자마다 견해가 약간씩 다르다. Linn(1993)은 검사 점수를 연계하는 방법을 동등화(equating), 조정(calibration), 통계적 조절(statistical moderation), 예측(prediction), 사회적 조절(social moderation) 등 5가지로 구분하면서 수직적 동등화는 이 중에서 조정에 해당한다고 하였다(성태제, 2000). 남현우(2004)는 검사 유형 간 비교 가능성을 찾을 때 사용하는 방법을 검사동등화(equating)라고 한다면, 서로 다른 검사 간 비교 가능성을 찾을 때 사용하는 방법은 연계화(linking) 또는 척도화(scaling)라고 하였다. 부재율(2003, 2005)도 동등화와 추정은 개념상 구분된다고 하였다. 그러나 동등화와 연계화는 통계적인 절차가 다르지 않고 전통적으로 많은 연구에서 수직적 동등화라는 용어를 많이 사용하고 있기 때문에 본 연구에서도 추정이라는 용어 대신에 수직적 동등화라는 용어를 사용하고자 한다(부재율, 2005).

본 연구의 목적은 공통문항이 포함된 검사를 이용하여 중학교에서 학년이 올라감에 따른 수학 학력 변화를 탐색하는 것이다. 학년이 올라감에 따라 수학 학력이 긍정적으로 변화된다면 이를 통해서 학교 교육의 효과를 실증적으로 보여줄 수 있을 것이다. 그뿐만 아니라 특정 학년에 대한 연도 간 비교에만 관심을 가지고 있는 현재의 국가수준 학업성취도 평가를 보완할 수 있는 방안도 제공할 수 있을 것이다. 특히 본 연구를 통해서 학년이 올라감에 따른 수학 학력 변화의 측정 가능성을 확인한다면, 장기적으로는 학력의 개인별 변화양상 및 이에 대한 개인 간 차이를 설명하기 위한 종단 연구 방법도 모색할 수 있을 것이다(Willet, 1988).

본 연구에서 다루게 될 연구문제는 다음과 같다.

첫째, 중학교에서 학년이 올라감에 따라 정답률이 높아지는 문항과 그렇지 않은 문항의 특성은 어떠한가?

둘째, 중학교에서 학년이 올라감에 따라 수학 학력의 평균적인 변화는 어떠한 양상을 보이는가?

셋째, 수직적 동등화 과정에서 추정된 개인별 동등화 점수는 동등화 방법에 따라서 어떠한 차이가 있는가?

Ⅱ. 이론적 배경

1. 수학 학력의 변화

본 연구의 목적은 중학생의 수학 학력이 학년이 올라감에 따라 어떠한 변화양상을 보이는지 공통문항이 포함된 가교검사 설계로 탐색하는 데 있다. 본 연구에서와 같이 서로 다른 학년의 척도를 동일화하려는 수직적 동등화가 성공적으로 수행되기 위해서는 수학 학력을 어떻게 정의할 것인지가 매우 중요하다. 본 연구에서 수학 학력은 '수학교육을 통해 학생들이 갖게 되는 인지적 능력으로서, 수학적 힘을 구현하는 계산, 개념 이해, 문제해결, 추론 등의 능력'으로 정의할 것이며, 이것은 수학의 내용 영역인 지식을 기반으로 발휘된다.

김응태, 박한식, 우정호(1988)는 "수학교육에 있어서 지와 정의는 서로 개념상 분리되어 있는 것처럼 보이지만, 실제는 상호관련을 가지고 있고, 심지어 동시에 변화하기도 한다."고 하면서, 수학에 관련한 지적 성취가 변할 수 있음을 지적한 바 있다. 수학에서의 학력이 수학에 대한 태도와 인과관계가 있다는 것으로부터 수학 학력이 변할 수 있음을 간접적으로 지적한 연구도 있다(Poffenberger et al., 1959; Brown, 1979; Schibeci & Riley, 1986; Simpson & Troost, 1982). 또한 수학 학력의 변화를 가정하고 여러 교수 방법을 학생들에게 처치하여 실험을 실시한 연구들도 있다(조용옥, 1993; 류희찬, 이기원, 1994; 구미종, 2002; 임선순, 2000). 이들의 연구에 따르면 수학 수업에서 학생들에게 발문으로 구성된 학습자료를 제시했을 때, 형성평가에서 주관식 문항을 단계별로 제시하여 학습지도를 했을 때, 패턴 블록을 활용하여 분수학습을 했을 때, elaborative feedback과 분할노트 기법을 사용했을 때, 실험을 실시하기 전과 후에 학생들의 학력이 향상되었다. 이것은 학생들의 수학 학력이 변화 가능하다는 것을 보여주는 것이다. 또한 수학 학력을 설명하는 지적 능력의 하나인 추론 능력이 변화됨을 보여준 연구도 있었는데, 이종희, 유현주, 김선희, 이진향(2002)은 개연적 추론 지도 방안을 모색하고 이를 중학교 2학년 학생들에게 14주간 적용하여, 실험그룹 학생들이 귀납, 유추, 시각적 추론에서 통제그룹 학생들보다 높은 점수를 받아 개연적 추론 능력이 향상되었다고 한다. 이들 연구는 수학 학력의 변화에 대한 근거를 수학에 대한 태도나 교수 방법에 의해 보여준 것이다. 하지만, 이 연구들은 장기적으로 학생들의 수학 학력이 어떤 양상으로 변화하며, 학교 교육을 통해 과연 수학 학력이 긍정적으로 변화될 수 있는지에 대해 일반화할

수 있는 논거를 제시하지는 못했다.

7차 교육과정은 수학적 힘의 신장을 목표로 하고 있다. 수학적 힘은 문제를 수학적으로 형식화하고, 수학적으로 추론하며, 여러 수학적 아이디어들 사이의 관련성을 찾아내고, 수학에 대한 의사소통을 할 수 있는 능력을 말한다. 교육과정에서 추구하는 수학적 능력인 수학적 힘은 단번에 갖추어질 수 있는 것이 아니며, 학생들이 수학에 대한 개념 이해에서 출발하여 다양한 수학적 사고와 추론, 의사소통, 문제해결을 경험하면서 누적되는 것이다. 그러므로 7차 교육과정을 통해 학교교육을 받은 학생들의 수학 학력이 학년이 상승함에 따라 향상되는지에 대한 점검이 필요하다. 수학 학력의 평가는 단순히 학생들의 계산 능력이나 수학적 개념에 관한 지식에 국한하지 않고 문제해결이나 추론도 포함하여 수학적 힘을 평가하는 것이 되어야 한다. 이 연구는 7차 교육과정의 수학과 목표인 수학적 힘이 학년이 올라갈수록 학생들에게 신장되어 수학 학력 또한 학년이 상승함에 따라 향상될 것을 가정한다.

2. 수직적 동등화 방법

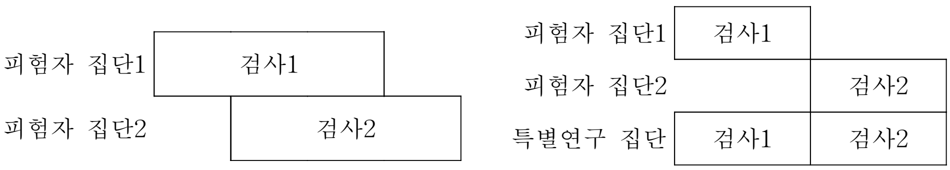
앞에서도 지적했듯이 수직적 동등화는 동일한 특성을 재고는 있지만 수준이 서로 다른 집단이 치른 검사 점수를 연계시키는 과정이기 때문에, 동일한 특성을 측정하면서 여러 동형검사를 비교하는 데 목적이 있는 검사동등화(equating)와는 다소 구분된다. 그러나 수직적 동등화 방법은 검사동등화의 그것과 대동소이하다.

본 연구에서는 학년이 올라감에 따른 수학 학력 변화를 추정하기 위해서 고전검사이론에 기반한 방법으로 터커의 선형동등화(Tucker’s linear equating), 레빈의 선형동등화(Levene’s linear equating) 및 빈도추정 동백분위동등화(equi-percentile equating)를 이용하였다. 그리고 문항반응이론에 기반한 방법으로는 능력모수동등화, 진점수동등화, 역진점수동등화를 이용하였다. 터커의 선형동등화, 레빈의 선형동등화 및 빈도추정 동백분위동등화는 Kolen이 개발한 ‘CIPE’ 프로그램을 이용하였고, 문항반응이론에 기반을 둔 능력모수동등화, 진점수동등화 및 역진점수동등화는 ‘PARSCALE(Muraki & Bock, 1998)’,「‘SCALE’, ‘CHECK’, ‘LISMIX’, ‘POLYEQUATE’(한국교육과정평가원, 2004)」프로그램을 이용하였다. 여기서 역진점수동등화(inverse true score equating)는 한국교육과정평가원의 학업성취도 평가에서 활용하고 있는 동등화 방법을 일컫는다(정구향 외, 2004; 김재철, 2005). 이에 대한 수리적인 설명은 Hambleton과 Swaminathan(1985), Kolen과 Brennan(1995), 김현철(1999), 김성훈(2000), 부재율(2000), 남현우(2001, 2003), 정구향 외(2004), 김재철(2005) 등을 참조하기 바란다.

3. 수직적 동등화를 위한 연구 설계

동등화를 위한 연구 설계로는 단일집단 설계(Single-Group Design)/무선집단 교차 설계(Counterbalanced Random-Groups Design), 동등집단 설계(Equivalent-Groups Design), 비동등집단 가교검사 설계(Anchor-Test-Nonequivalent-Groups Design) 등이 있다(Kolen & Brennan, 1995; 남현우, 2001). 피험자 집단 P와 Q의 능력 차이가 큰 수직적 동등화를 위한 설계로는 주로 [그림 1]에서 보는 바와 같이, 공통문항이 포함된 가교검사 설계와, 두 검사를 동시에 모두 치른 공통피험자를 이용한 가교피험자 설계가 이용된다.

본 연구에서는 학년이 올라감에 따른 수학 학력 변화를 탐색하기 위해서 [그림 1]의 왼쪽에서 제시한 가교검사 설계를 활용하였다. 가교검사 설계에서는 피험자 집단 P와 Q의 능력 차이가 지나치게 크다든지 공통문항 수가 전체 문항 수의 20% 이하인 경우에 동등화 오차(equating error)가 커지게 된다(남현우, 2001; Cook, Eignor & Taft, 1988). 따라서 본 연구에서는 학년 간에 수학적 능력 차이가 클 수 있기 때문에 동등화 오차를 줄이기 위해서 공통문항을 전체의 50% 내외로 구성하였다.



〔그림 1〕 가교검사 설계(좌)와 가교피험자 설계(우)

Ⅲ. 연구 방법

1. 연구 대상

본 연구에서는 서울에 위치한 A, B, C 3개의 중학교에서 1학년 472명, 2학년 511명, 3학년 496명 등 총 1,479명을 분석에 활용하였다(<표 1> 참조).

학년에 따른 학력 차이는 지역적인 특성, 교수·학습 환경 또는 가정환경 변인의 영향이 크게 작용할 수 있다. 학력 차이를 의미 있게 분석하기 위해서는 이러한 배경적인 변인을 통제할 필요가 있다. 일반적으로 한 학교 내에서 각 학년의 지역적인 특성, 교수·학습 환경 또는 가정 환경 변인은 유사하다고 볼 수 있으므로, 본 연구에서는 학력에 영향을 줄 수 있

는 배경적인 변인을 효과적으로 통제하기 위해 각 학교에서 학년별 인원을 고르게 표집하였다. 성별 분포를 보면 남학생이 697명, 여학생이 782명으로 여학생이 조금 더 많았다.

〈표 1〉 학년별 성별 학생 표집 수

학교 / 학년	A중학교	B중학교	C중학교	성		전체
				남	여	
1	222	108	142	229	243	472
2	193	116	202	248	263	511
3	199	96	201	220	276	496
전체	614	320	545	697	782	1,479

2. 검사 도구 개발

가. 수학 학력 평가의 평가틀 및 설계

본 연구에서 수학 학력을 측정하기 위한 평가문항은 내용영역과 행동영역의 틀에 따라 구성되었다. 내용영역은 7차 교육과정의 구분에 따라 수와 연산, 도형, 측정, 확률과 통계, 문자와 식, 규칙성과 함수의 6개로 구성되고, 행동영역은 계산, 개념 이해, 문제해결, 추론의 4가지로 구성되었다. 이 행동영역은 황혜정, 최승현(1999)의 평가틀을 토대로 한 것이며, 4가지 행동이 전체 문항에 유사한 비율로 들어가도록 했다. 그리고 각 학년별로 내용영역과 행동영역의 구성 비율을 유사하게 하여 각 평가틀을 동일하게 유지하려 했다. 최종적으로 개발된 문항 수는 학년별로 37개씩이고, 검사 시간은 45분이었다. 문항 형식은 일부 선다형이 포함되었지만 대부분 단답형으로 구성하였다.

문항은 크게 두 개 이상 학년에 모두 적용되는 ‘공통문항’과 특정 학년에만 적용되는 ‘비공통문항’으로 나눌 수 있다. <표 2>는 학년별 검사의 문항이 어떻게 구성되어 있는지를 보여 준다. 대문자 A, B, C는 학년 간 공통문항을, 소문자 a, b, c, d, e는 학년 고유문항을, Ⅰ, Ⅱ, Ⅲ은 문항의 내용에 해당하는 학년을 의미한다. 예컨대 A-Ⅰ은 1학년 내용으로 제작된 3개 학년 공통문항을 의미하고, e-Ⅲ은 3학년 내용으로 제작된 3학년 고유문항을 의미한다.

〈표 2〉 학년별 문항 구성 요약표

학년	검사 문항 구성			
1	A-Ⅰ	B-Ⅰ	b-Ⅰ	c-Ⅰ
2	A-Ⅰ	B-Ⅰ	C-Ⅱ	d-Ⅱ
3	A-Ⅰ	a-Ⅲ	C-Ⅱ	e-Ⅲ

* A, B, C : 학년 간 공통문항
** a, b, c, d, e : 학년 고유문항
*** Ⅰ, Ⅱ, Ⅲ : 문항 내용이 들어있는 학년

검사가 실시된 시기가 7월인 관계로, 1학년 학생을 대상으로 한 검사 문항에는 7단계의 확률과 통계, 도형, 측정 내용이 없으며, 대신 중학교 내용을 학습하는 데 필요한 6단계 수준의 내용이 2개 문항으로 삽입되어 있다. 학년별 평가내용은 <표 3>과 같다. 공통문항의 경우 해당 학년의 범위 이전의 내용영역도 포함되었다.

〈표 3〉 학년별 평가 내용

학년	시험범위	비고
1	제7차 교육과정의 6단계 + 7-가 단계 일부	공통문항은 해당 학년 이전 것도 포함
2	제7차 교육과정의 7단계 + 8-가 단계 일부	
3	제7차 교육과정의 8단계 + 9-가 단계 일부	

<표 2>, <표 3>을 토대로 <표 4>와 같은 중학교 1학년, 2학년, 3학년 검사의 이원분류표를 작성하였다.

두 학년에 동시에 시행되는 공통문항은 두 집단의 능력의 차이를 나타내는 데 목적이 있으며, 동등화를 위해서 공통문항은 전체 문항과 내용적·측정학적 특성이 동일해야 한다(Kolen & Brennan, 1995). 본 연구는 중학교 수학의 핵심 개념을 내용으로 다루고, 학생들에게 친숙하면서 복잡하지 않은 문항을 공통문항에 선정하였다. 그래야 단기적인 암기가 평가결과에 영향을 주지 않아 장기적 관점에서 수학 학력을 측정할 수 있기 때문이다. 예컨대 중학교 1학년 내용으로 구성된 공통문항이 단기적인 기억능력에 의존하는 문항이라면 장기적인 학습능력이 더 높다고 볼 수 있는 중학교 3학년 학생보다도 중학교 1학년 학생이 정답을 더 잘 맞히게 되어, 종국에는 학년이 올라갈수록 학력이 하락한다는 비현실적인 결론을 도출하게 될 우려가 있다. 또, 특정 영역 내용에 치우치지 않도록 문항을 구성하여 내용에 따른 학력의 영향력을 최소화하려 했으며, 전체 검사의 축소판이 되도록 각 요소의 수를 유지하였다.

비공통문항은 학년별로 다른 것이지만, 수학 학력을 평가한다는 검사 도구의 성격에 맞도

록 학년별로 몇 가지 조건을 동일하게 유지하려 했다. 즉, <표 2>에서 a는 B에 포함된 문항과, b는 C에 있는 문항과, 그리고 c, d, e에 있는 문항은 서로 '동일한 조건'의 문항이 되도록 했는데, 동일한 조건이란 각 부분에 포함된 문항과 대응하는 문항들이 문항의 형식, 배점, 예상난이도, 평가틀에서 동일함을 의미한다. 특히 학년 간 난이도 차이를 최소화하기 위해서 대응하는 문항들끼리 예상난이도를 동일하도록 했는데, 이때의 예상난이도는 학년이 동일한 피험자를 가상적으로 설정한 다음 그들에게 기대할 수 있는 기대정답률을 의미한다. 또, 학년별 평가틀이 동일하도록 대응하는 문항끼리 되도록 내용영역과 행동영역이 동일하게 하였다. 예컨대 내용영역을 보면, 1학년에서 함수에 관한 문항이라면 2학년에서는 일차함수, 3학년에서는 이차함수를 그 내용으로 하였다.

〈표 4〉 수학 학력 검사의 이원분류표

분류	내용	단계	행동영역	내용영역	문항수	분류	내용	단계	행동영역	내용영역	문항수
A-1	함수의 그래프	7-가	개념 이해	규칙성과 함수	2	C-2	엇각의 성질	7-나	개념 이해	도형	1
A-1	경우의 수	6-나	문제해결	확률과 통계	2	C-2	소수의 분류	8-가	추론	수와 연산	1
A-1	분수의 나눗셈	6-나	문제해결	수와 연산	1	C-2	거듭제곱의 성질	8-가	추론	수와 연산	1
A-1	정수의 계산	7-가	계산	수와 연산	1	C-2	연립일차방정식	8-가	추론	문자와 식	1
A-1	일차식의 계산	7-가	계산	문자와 식	1	C-2	연립일차방정식의 활용	8-가	문제해결	문자와 식	2
A-1	일차방정식의 풀이	7-가	계산	문자와 식	1	C-2	입체도형의 부피	7-나	추론	측정	2
B-1	소인수분해	7-가	문제해결	수와 연산	2	d-2	연립일차방정식	8-가	개념 이해	문자와 식	2
B-1	일차방정식의 활용	7-가	개념 이해	문자와 식	2	d-2	연립일차부등식	8-가	개념 이해	문자와 식	2
B-1	일차방정식의 활용	7-가	개념 이해	문자와 식	2	d-2	연립일차방정식의 활용	8-가	문제해결	문자와 식	2
B-1	정수와 유리수	7-가	문제해결	수와 연산	2	d-2	순환소수의 계산	8-가	계산	수와 연산	1
B-1	일차방정식의 활용	7-가	문제해결	문자와 식	2	d-2	다항식의 덧셈과 뺄셈	8-가	계산	문자와 식	2
B-1	최소공배수의 성질	7-가	추론	수와 연산	1	d-2	일차함수의 그래프	8-가	추론	규칙성과 함수	1
b-1	유리수의 대소 관계	7-가	개념 이해	수와 연산	1	a-3	제곱근의 성질	9-가	개념 이해	수와 연산	2
b-1	정수와 유리수	7-가	추론	수와 연산	1	a-3	무게중심의 성질	8-나	개념 이해	도형	2
b-1	교집합과 합집합	7-가	개념 이해	수와 연산	1	a-3	이등변삼각형의 성질	8-나	개념 이해	도형	2
b-1	약수	7-가	추론	수와 연산	1	a-3	확률	8-나	문제해결	확률과 통계	2
b-1	일차방정식의 활용	7-가	문제해결	문자와 식	2	a-3	경우의 수	8-나	문제해결	확률과 통계	2
b-1	함수의 그래프	7-가	추론	규칙성과 함수	2	a-3	무리수의 성질	9-가	추론	수와 연산	1
c-1	정수의 계산	7-가	개념 이해	수와 연산	2	e-3	이차함수의 꼭지점	9-가	개념 이해	규칙성과 함수	2
c-1	교집합과 합집합	7-가	개념 이해	수와 연산	2	e-3	이차함수의 그래프	9-가	개념 이해	규칙성과 함수	2
c-1	최대공약수의 성질	7-가	문제 이해	수와 연산	2	e-3	삼각형의 닮음	8-나	문제해결	도형	2
c-1	약수의 개수	7-가	계산	수와 연산	1	e-3	다항식의 곱셈	9-가	계산	문자와 식	1
c-1	이진법의 수의 계산	7-가	계산	수와 연산	2	e-3	이차방정식의 중근	9-가	문제해결	문자와 식	2
c-1	집합	7-가	추론	수와 연산	1	e-3	이차함수의 그래프	9-가	추론	규칙성과 함수	1

공통문항이든 비공통문항이든 간에 바람직하지 않은 문항은 지엽적인 개념을 포함함으로써 배운 시점에 크게 영향 받는 문항이다. 위의 원칙이 잘 지켜졌는지 확인하기 위해서 본

연구는 연구 대상이 소속된 학교의 수학교사 3명과 한국교육과정평가원 수학교육학 박사 연구원 1명에게 문항의 성격과 예상난이도의 적합성을 검토받았고, 이들 전문가의 의견을 수렴하여 문항을 수정 또는 교체하였다.

나. 검사 개발 절차와 자료 수집

검사 개발과 자료 수집은 다음의 절차에 따라 이루어졌다.

첫째, 수학 학력을 개념화함으로써 평가문항의 성격을 규정하였다.

둘째, 검사 시간을 정하고 이를 고려하여 학년별 문항 수와 학년 간 공통문항 수를 정하였다.

셋째, 문항 제작을 위한 학년별 이원분류표를 작성하였다.

넷째, 중학교에 재직 중인 3명의 교사의 도움을 받아 문항 초안을 제작하였다.

다섯째, 중학교에 재직 중인 3명의 수학 교사와 한국교육과정평가원 수학교육학 박사 1명 등 총 4명의 전문가로부터 문항 성격, 예상난이도 등의 적절성을 평가받았고, 부적합한 문항은 수정 또는 교체하였다.

여섯째, 표집학교 학생을 대상으로 2005년 7월 초순에서 중순 사이에 검사를 실시하였다. 시험 시간은 45분이었으며, 정확한 시험 시간 배정을 위해서 감독관 교사가 5분 전에 입실하도록 하였다.

일곱째, 대학원생 2명에게 채점 기준을 명확히 교육한 다음, 채점과 코딩을 하도록 하였다.

3. 분석 방법

수학 학력이 학년이 올라감에 따라서 어떻게 변화하는지 보기 위해서 터커의 선형동등화, 레빈의 선형동등화, 빈도추정 동백분위동등화, 능력모수동등화, 진점수동등화, 역진점수동등화 등 6가지 방법으로 동등화를 실시하였다. 이 중에서 레빈의 선형동등화 및 빈도추정 동백분위동등화는 부적절성이 발견되어 최종 비교 분석에서는 제외하였다. 동등화 방법으로 추정한 점수 간 비교를 용이하도록 하기 위해서 중학교 1학년에서의 평균과 표준편차가 각각 50과 10이 되도록 선형변환($aX+b$)하였고, 이때 사용된 기울기(a)와 절편(b)을 타 시점의 선형변환($aY+b$)에도 그대로 활용하였다.

Ⅳ. 연구 결과

1. 분석 자료에 대한 기초 통계

전체검사와 가교검사에 대한 학년별 기술통계치는 <표 5>와 같다. 전체 37문항에 대해서 문항별 1점을 부여하였기 때문에 원점수의 만점은 37점이었다. <표 5>에서 보는 바와 같이 전체검사와 가교검사 모두 학년이 올라갈수록 수학 학력의 평균이 상승하였고, 첨도와 왜도는 모두 음수였다. α계수는 학년에 따라서 .919에서 .939로써 매우 높았다. 즉, 학년이 올라갈수록 학력의 전반적 수준은 높아지는 경향이 있음을 알 수 있었다. 그러나 이러한 결과는 학년별로 비공통문항이 포함된 서로 다른 검사를 실시하여 얻은 결과이기 때문에 해석에 한계가 있으며, 이를 보완하기 위해서는 학년별 점수를 동일한 척도로 변환하는 절차를 수행하여야 한다.

〈표 5〉 전체검사와 가교검사에 대한 학년별 기술통계(37점 만점)

구 분	전체검사*			가교검사**			
				1·2학년 중		2·3학년 중	
	1학년	2학년	3학년	1학년	2학년	2학년	3학년
사례 수	472	511	496	472	511	511	496
평균	18.71	20.99	21.59	9.12	11.07	8.84	9.12
표준편차	9.412	9.541	8.606	4.063	4.950	4.395	4.063
최소값	1	1	1	0	0	0	0
최대값	37	36	37	16	19	16	16
첨도	-1.097	-1.041	-.913	-.933	-1.065	-1.105	-.933
왜도	-.054	-.372	-.325	-.239	-.281	-.250	-.239
α계수	.935	.939	.919	.972***	.953***	.960***	.941***

* 전체검사 만점 : 37점
** 1·2학년 가교검사 만점 : 19점, 2·3학년 가교검사 만점 : 16점
*** 전체검사와 상관

<표 6>은 37개 문항에 대한 학년별 정답률과 변별도를 비교한 결과이다. 일련번호 1에서 8까지는 3개 학년의 공통문항에 대한 것이다. 공통문항의 경우, 학년이 올라감에 따라서 정답률이 높아지는 경향이 뚜렷하게 나타남을 알 수 있다. 그러나 비공통문항의 경우, 대응하는 문항들의 정답률이 학년이 올라감에 따라서 올라가는 경향성이 공통문항의 그것과 같이 뚜렷하게 나타나지는 않았다.

2. 학년이 올라감에 따라 정답률이 높아지는 문항의 특성

학년별로 학생들이 같은 문제를 풀 수 있도록 한 공통문항은 중학교 1학년과 중학교 2학년 공통이 19문항, 중학교 2학년과 중학교 3학년 공통이 16문항이었다. 3개 학년 공통문항 8개 중에서 7개는 학년이 올라가면서 정답률이 증가했다. 그 내용을 살펴보면 6학년 수준에서 출제된 2개 문항과 일차식의 계산, 일차방정식의 풀이, 정비례의 그래프와 관련된 것이다. 이 문항들의 내용은 학생들의 학년이 올라가면서 보다 수준 높은 내용으로 학생들이 계속 접하게 되는 것이다. 예를 들어, 7학년의 일차식의 계산은 8학년의 다항식의 덧셈과 뺄셈으로, 7학년의 일차방정식의 풀이는 8학년의 연립방정식과 이차방정식의 풀이로, 7학년의 정비례의 그래프는 8학년 일차함수와 9학년 이차함수의 그래프로 연계되어 교육과정상 계속 지도되며, 이 문항의 계산 능력과 개념 이해 능력은 후속적인 수학 학습에 필수적으로 필요하다. 이런 문항에서 측정하고자 한 능력이 학년이 올라갈수록 학생들에게서 상승되고 있었던 것이다.

〈표 6〉 공통문항의 정답률 및 변별도

일련번호	1, 2학년 공통문항				2, 3학년 공통문항			
	1학년		2학년		2학년		3학년	
	정답률(%)	변별도	정답률(%)	변별도	정답률(%)	변별도	정답률(%)	변별도
1	50.2	0.761	68.5	0.702	68.5	0.702	80.4	0.634
2	34.5	0.711	52.6	0.744	52.6	0.744	66.5	0.679
3	73.9	0.374	80.2	0.303	80.2	0.303	81.0	0.222
4	29.0	0.468	33.9	0.447	33.9	0.447	42.3	0.399
5	76.7	0.495	78.7	0.527	78.7	0.527	79.6	0.472
6	51.1	0.529	43.8	0.554	43.8	0.554	34.9	0.397
7	32.6	0.538	34.1	0.492	34.1	0.492	52.6	0.508
8	53.8	0.576	65.9	0.569	65.9	0.569	66.5	0.644
9	76.7	0.510	82.2	0.531	38.4	0.616	39.7	0.607
10	46.8	0.691	29.5	0.448	45.0	0.486	37.5	0.430
11	49.4	0.586	56.9	0.639	52.6	0.653	59.5	0.643
12	56.1	0.471	61.1	0.401	74.4	0.572	71.0	0.495
13	58.1	0.708	61.8	0.755	64.2	0.761	55.2	0.733
14	67.6	0.655	64.4	0.668	64.2	0.692	62.5	0.650
15	87.3	0.366	93.3	0.348	67.1	0.493	60.3	0.303
16	63.3	0.543	76.7	0.450	20.4	0.365	22.2	0.366
17	41.1	0.560	51.7	0.603				
18	16.3	0.513	22.9	0.554				
19	52.3	0.627	48.9	0.546				
평균	잘못된 계산식	잘못된 계산식	잘못된 계산식	잘못된 계산식	잘못된 계산식	잘못된 계산식	잘못된 계산식	잘못된 계산식

반면, 학년이 올라갈수록 정답률이 하락한 1개의 문항은 $(-1)^{100} \times (-1^{100})$의 값을 구하는 것으로, 학년별로 7학년의 정답률은 51.1%, 8학년은 43.8%, 9학년은 34.9%이었다. 이것은 7단계에서 거듭제곱의 뜻을 알고 계산에 적용할 수 있는지를 묻는 문항으로 수와 연산의 내용에 해당한다. 현재 7차 교육과정의 수와 연산영역에서는 8단계에서 유리수와 소수의 관계를, 9단계에서는 무리수 개념을 다루어 수의 거듭제곱이 학년이 올라감에 따라 연계되어 지도되지 않는다. 이것은 이 문항에 필요한 계산 능력이 학생들에게 계속 활용되지 못했음을 말한다. 8단계의 문자와 식에서 지수법칙이 있어 8학년 학생들이 다시 되새겨 볼 수는 있지만 지수법칙에서는 주로 문자를 중심으로 하기 때문에, 학생들의 수의 거듭제곱에 대한 문제해결 경험은 부족하고, 그리하여 학년이 올라갈수록 그 수를 이용한 계산 능력이 향상되지 않았을 수 있다.

3. 학년별 수학 학년 변화양상

본 연구에서는 동등화 방법으로 터커의 선형동등화, 레빈의 선형동등화, 빈도추정 동백분위동등화, 능력모수동등화, 진점수동등화, 역진점수동등화 등 6가지 방법을 적용하여 학년별 변화 정도를 비교하였다.

〈표 7〉 동등화 방법에 따른 동등점수

터커의 선형동등화			레빈의 선형동등화			빈도추정 동백분위동등화			역진점수동등화		
1학년	2학년	3학년	1학년	2학년	3학년	1학년	2학년	3학년	1학년	2학년	3학년
30	30	29	30	31	30	30	30	30	24	24	24
31	31	30	31	32	31	31	32	31	29	29	28
32	32	31	32	33	32	32	33	31	31	31	31
33	33	32	33	33	33	33	34	32	33	33	33
34	34	33	34	34	34	34	35	33	34	34	34
35	35	34	35	35	35	35	36	34	36	36	36
36	36	35	36	36	36	36	36	35	37	37	37
38	38	38	38	38	38	38	38	36	38	38	39
39	39	39	39	39	39	39	39	38	39	39	40
40	40	40	40	40	40	40	40	40	40	40	41
41	41	41	41	41	41	41	41	41	41	41	42
42	42	42	42	42	42	42	42	42	43	43	43
43	43	43	43	43	43	43	43	43	44	44	44
44	44	44	44	44	44	44	44	44	45	45	45
45	45	45	45	45	45	45	45	45	46	46	46
46	46	46	46	46	46	46	45	46	47	47	47
47	47	47	47	47	47	47	46	47	47	47	48
48	48	48	48	48	48	48	47	48	48	48	49
49	48	49	49	49	49	49	48	49	49	49	50
50	49	50	50	50	50	50	49	50	19	50	50

터커의 선형동등화			레빈의 선형동등화			빈도추정 동백분위동등화			역진점수동등화		
1학년	2학년	3학년	1학년	2학년	3학년	1학년	2학년	3학년	1학년	2학년	3학년
51	50	51	51	50	51	51	50	51	20	51	51
52	51	52	52	51	52	52	51	52	21	52	52
53	52	53	53	52	53	53	52	53	22	53	53
55	53	55	55	53	55	55	53	55	23	54	54
56	55	56	56	55	56	56	55	56	24	55	55
57	56	57	57	56	57	57	56	57	25	56	56
58	57	58	58	57	58	58	57	58	26	57	57
59	58	59	59	58	59	59	58	59	27	58	58
60	59	60	60	59	60	60	59	60	28	59	59
61	60	61	61	60	61	61	60	61	29	60	60
62	61	62	62	61	62	62	61	62	30	62	62
63	62	63	63	62	63	63	62	63	31	63	63
64	63	64	64	63	64	64	63	64	32	64	64
65	64	65	65	64	65	65	64	65	33	66	66
66	65	66	66	65	66	66	66	66	34	67	67
67	66	67	67	66	67	67	67	67	35	69	69
68	67	68	68	67	68	68	68	68	36	71	71
69	67	69	69	68	69	69	69	69	37	75	75

<표 7>은 각 동등화 방법별로 상위 학년 검사의 점수를 하위 학년 검사 점수로 동등화할 때 활용된 동등점수이다. 예컨대 터커의 선형동등화에 의하면 중학교 1학년에서의 55점은 중학교 2학년에서의 53점과 동일한 점수임을 의미한다. 이때 1학년 점수는 원점수를 평균 50점, 표준편차 10점인 점수로 선형변환했을 때의 점수이다. 능력모수동등화와 진점수동등화의 경우, 원점수에서의 순위와 동등화 방법으로 추정한 점수의 순위가 바뀔 수 있기 때문에 동등점수를 표로 제시하지 않았다. 그러나 역진점수동등화의 경우는 원점수와 순위가 바뀌지 않기 때문에 동등점수를 제시하였다. 역진점수동등화 과정에서 모수추정을 위해서는 2모수 로지스틱모형을 이용하였고, 가교문항으로부터 3개 학년의 문항모수를 선형적으로 연결해 주는 기울기와 절편을 구하기 위해서는 Kim과 Lee(2004)가 제안한 '확장된 Stocking-Lord 방법'을 활용하였다.

터커의 선형동등화, 레빈의 선형동등화, 빈도추정 동백분위동등화에서의 각 점수대별 동등화 무선오차는 <표 8>과 같다. 터커의 선형동등화와 레빈의 선형동등화는 중간 점수대에서 동등화 표준오차가 더 적었던 반면 빈도추정 동백분위동등화는 높은 점수대에서 동등화 표준오차가 더 적었다. 점수대별 동등화 표준오차가 가장 적은 방법은 터커의 선형동등화였으며 레빈의 선형동등화가 그 뒤를 따랐고, 빈도추정 동백분위동등화가 가장 컸다. 이러한 결과는 부재율(2000)의 연구와 부합하는 결과이다. 이러한 이유에 근거하여 본 연구에서는 레빈의 선형동등화와 빈도추정 동백분위동등화를 수직적 동등화 결과 비교 과정에서는 제외하였다.

동등화 방법에 따라서 학년이 올라감에 따른 평균과 표준편차가 어떠한 차이가 있는지 <표 9>와 같이 비교하였다. 4가지 동등화 결과를 쉽게 비교할 수 있도록 하기 위해서 1학년

에서의 평균과 표준편차를 각각 50과 10이 되도록 선형변환하였고, 선형변환 과정에 이용된 기울기와 절편을 2학년과 3학년에도 일괄적으로 적용함으로써 2학년과 3학년에서의 척도를 1학년에서의 척도에 대한 상대적인 크기로 나타내었다. [그림 2]와 [그림 3]은 동등화 방법에 따른 학년별 평균과 표준편차의 변화를 나타낸 것이다.

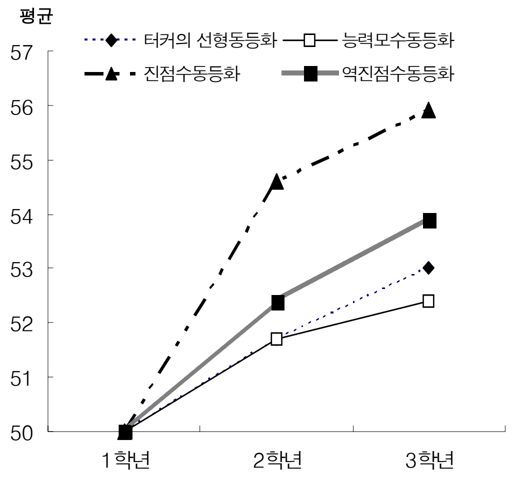
〈표 8〉 고전검사이론에 근거한 동등화 방법의 점수대별 동등화 무선오차 추정치

원점수	터커의 선형동등화		레빈의 선형동등화		빈도추정 동백분위동등화	
	중2→중1	중3→중2	중2→중1	중3→중2	중2→중1	중3→중2
0	0.3138	0.4494	0.3252	0.4713	-	-
1	0.3030	0.4336	0.3138	0.4542	0.4167	-
2	0.2924	0.4179	0.3026	0.4373	0.3618	0.6781
3	0.2819	0.4024	0.2915	0.4205	0.3424	0.7845
4	0.2716	0.3869	0.2806	0.4039	0.3594	0.8170
5	0.2615	0.3717	0.2700	0.3874	0.3484	0.6426
6	0.2517	0.3566	0.2596	0.3712	0.4029	0.5172
7	0.2420	0.3417	0.2494	0.3552	0.4641	0.5248
8	0.2327	0.3270	0.2396	0.3394	0.4771	0.5186
9	0.2237	0.3125	0.2301	0.3239	0.4489	0.4970
10	0.2150	0.2984	0.2209	0.3088	0.4182	0.5371
11	0.2068	0.2846	0.2123	0.2940	0.4229	0.6579
12	0.1990	0.2712	0.2041	0.2797	0.4224	0.7645
13	0.1917	0.2582	0.1964	0.2659	0.4017	0.6860
14	0.1850	0.2457	0.1894	0.2528	0.4491	0.6041
15	0.1789	0.2338	0.1831	0.2403	0.5388	0.5429
16	0.1735	0.2226	0.1775	0.2287	0.5623	0.5241
17	0.1689	0.2122	0.1727	0.2180	0.5261	0.5206
18	0.1651	0.2028	0.1689	0.2084	0.4714	0.4774
19	0.1622	0.1943	0.1661	0.2001	0.4434	0.4803
20	0.1603	0.1871	0.1642	0.1933	0.4528	0.4880
21	0.1593	0.1812	0.1634	0.1880	0.4351	0.4746
22	0.1594	0.1768	0.1637	0.1844	0.4103	0.4754
23	0.1604	0.1740	0.1651	0.1827	0.4374	0.4655
24	0.1624	0.1729	0.1675	0.1828	0.4340	0.4067
25	0.1653	0.1734	0.1708	0.1848	0.4180	0.3704
26	0.1691	0.1757	0.1752	0.1886	0.4118	0.3177
27	0.1738	0.1795	0.1804	0.1941	0.3854	0.2716
28	0.1792	0.1849	0.1863	0.2012	0.3574	0.3433
29	0.1853	0.1916	0.1931	0.2097	0.2951	0.3604
30	0.1920	0.1996	0.2004	0.2195	0.2434	0.2545
31	0.1993	0.2088	0.2084	0.2303	0.2376	0.1793
32	0.2072	0.2188	0.2168	0.2420	0.3213	0.1952
33	0.2154	0.2298	0.2258	0.2546	0.3128	0.2055
34	0.2241	0.2414	0.2351	0.2678	0.3676	0.2299
35	0.2332	0.2537	0.2448	0.2817	0.4864	0.2471
36	0.2425	0.2665	0.2548	0.2961	0.0000	0.1891
37	0.2521	0.2798	0.2651	0.3109	-	-
평균	잘못된 계산식	잘못된 계산식	잘못된 계산식	잘못된 계산식	잘못된 계산식	잘못된 계산식

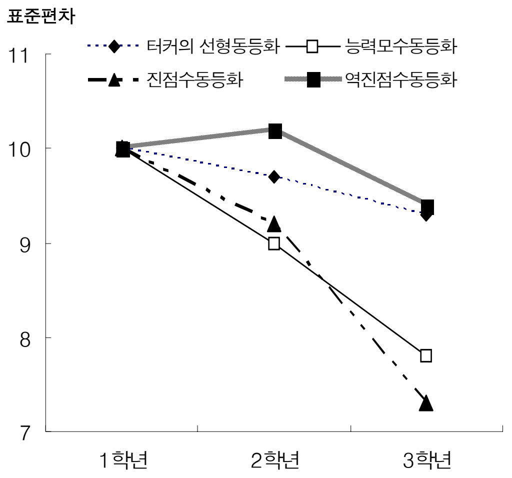
〈표 9〉 동등화 방법에 따른 학년별 평균과 표준편차

동등화 방법	구분	1학년	2학년*	3학년*
터커의 선형동등화	M SD	50.0 10.0	51.7 9.7	53.0 9.3
능력모수동등화	M SD	50.0 10.0	51.7 9.0	52.4 7.8
진점수동등화	M SD	50.0 10.0	54.6 9.2	55.9 7.3
역진점수동등화	M SD	50.0 10.0	52.4 10.2	53.9 9.4

* 1학년 척도의 평균과 표준편차를 각각 50.0과 10.0으로 선형변환했을 때 이에 대한 상대적인 점수임.



〔그림 2〕 동등화 방법에 따른 학년별 평균 변화



〔그림 3〕 동등화 방법에 따른 학년별 표준편차 변화

<표 9>, [그림 2] 및 [그림 3]으로부터 다음을 확인할 수 있다.

첫째, 4가지 동등화 방법 모두 학년이 올라갈수록 평균은 올라가고 표준편차는 작아지는 경향이 있었다.

둘째, 학년별 평균을 기준으로 4가지 동등화 방법 중 가장 민감하게 변화를 반영하는 동등화 방법은 '진점수동등화'였고, 변화를 가장 둔하게 반영하는 동등화 방법은 '능력모수동등화'였다.

셋째, 학년이 올라감에 따라서 개인 간 차이를 가장 크게 추정되는 동등화 방법은 '역진점수동등화'였고, 가장 작게 추정하는 동등화 방법은 '진점수동등화'와 '능력모수동등화'였다.

4. 동등화 방법에 따른 개인별 점수 비교

동등화 방법으로 추정된 점수들 간의 관계가 직선적인지 곡선적인지를 확인하기 위해서 중학교 3학년에서 임의의 두 가지 동등화 방법에서 추정된 점수 간에 피어슨 상관을 구하였다. 그 결과 <표 10>에서 보는 바와 같이, 4가지 동등화 방법 간에는 상관이 매우 높기는 하지만 서로 직선적인 관계는 아님을 확인할 수 있었다. 그리고 동등화 방법으로 추정된 점수가 동등화 방법에 따라서 피험자의 서열이 바뀔 수 있는지 보기 위해서 스피어만 상관을 구하였다. 그 결과 <표 10>에서 보는 바와 같이, 터커의 선형동등화와 역진점수동등화, 능력모수동등화와 진점수동등화는 동등화 방법에 따라서 추정한 피험자 점수의 순위가 전혀 바뀌지 않았으나, 터커의 선형동등화와 능력모수동등화, 터커의 선형동등화와 진점수동등화, 능력모수동등화와 역진점수동등화, 진점수동등화와 역진점수동등화는 동등화 방법에 따라서 피험자의 순위가 바뀔 수 있음을 확인할 수 있었다.

〈표 10〉 중학교 3학년의 동등화 방법에 따른 추정 점수 간 상관

동등화 방법	피어슨				스피어만			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
터커의 선형동등화(1)	1.000	-	-	-	1.000	-	-	-
능력모수동등화(2)	0.980	1.000	-	-	.989	1.000	-	-
진점수동등화(3)	0.989	0.991	1.000	-	.989	1.000	1.000	-
역진점수동등화(4)	0.995	0.987	0.985	1.000	1.000	.989	.989	1.000

동등화 방법에 따라서 추정된 개인별 점수가 통계적으로 유의미한 차이가 있는지 확인하기 위해서 중학교 3학년에서의 점수에 대한 대응표본 t검증(Paired Samples t-test)을 실시하였다. 그 결과 <표 11>에서 보는 바와 같이, 동등화 방법에 따라서 모두 차이가 있음을 확인할 수 있었다.

〈표 11〉 3학년에서의 동등화 방법에 따른 점수 차이 검증

동등화 방법 간 차이	평균 차이	표준편차 차이	t	p
터커의 선형동등화-능력모수동등화	0.611	2.244	6.067	0.000
터커의 선형동등화-진점수동등화	-2.934	2.308	-28.306	0.000
터커의 선형동등화-역진점수동등화	-0.924	0.980	-21.012	0.000
능력모수동등화-진점수동등화	-3.545	1.115	-70.786	0.000
능력모수동등화-역진점수동등화	-1.536	2.131	-16.046	0.000
진점수동등화-역진점수동등화	2.009	2.564	17.451	0.000

Ⅴ. 결론 및 논의

본 연구에서는 중학교 3년 동안 학생들의 수학 학력 변화가 어떠한 양상을 보이는지 탐색하려 하였다. 수학 학력의 변화를 추정하기 위해서는 기본적으로 3개 학년에 대해서 동일한 검사를 적용해야 하지만, 수학 학력의 변화를 제대로 평가하기 위해서는 학년이 올라가면서 새롭게 학습한 교육내용도 반영해야 한다. 이에 본 연구에서는 학년 간 공통문항과 더불어 새로운 교육내용을 반영한 학년별 고유문항도 검사문항에 포함하였다. 동일하지 않은 검사로 3개 학년 점수를 상대적으로 비교하기 위해서는 피험자 점수를 동일 척도로 변환하기 위한 측정학적인 절차를 밟아야 한다. 특히 본 연구에서와 같이 비교하고자 하는 집단의 능력수준이 이질적이고 각 집단에서 다루고 있는 교육내용의 수준도 차이가 있는 상황에서, 새로운 교육내용을 반영한 비동일 검사로 학력의 변화를 탐색하고자 할 때에는 다양한 논의가 선행되어야 한다. '수학 학력을 어떻게 정의할 것인지', '공통문항과 비공통문항의 성격을 어떻게 규정할 것인지', 그리고 '어떠한 동등화 방법을 적용할 것인지' 등이 그것이다. 본 연구는 수직적 동등화 상황에서 직면할 수 있는 이러한 선행 문제에 대한 해결 가능성을 탐색하는 데 목적이 있다.

본 연구에서는 수학 학력을 "수학교육을 통해 학생들이 갖게 되는 인지적 능력으로서, 수학적 힘을 구현하는 계산, 개념 이해, 문제해결, 추론 등의 능력"으로 정의함으로써 학습한 시기의 영향을 최소화하려 하였다. 문항 수는 학년별로 총 37개였고 시험 시간은 45분이었다. 동등화를 위해서 1학년과 2학년 공통문항을 19개, 2학년과 3학년 공통문항을 16개 포함시켰고, 이 중에서 8문항은 3개 학년 공통문항이었다. 37문항에 대한 α계수는 학년에 따라서 .919에서 .939로써 매우 높았다. 연구 대상은 서울에 위치한 세 개의 중학교에 재학 중인 학생으로 1학년 472명, 2학년 511명, 3학년 496명 등 총 1,479명이었고, 검사 시기는 2005년 7월이었다. 연구에 최종적으로 활용된 동등화 방법은 터커의 선형동등화, 능력모수동등화, 진점수동등화, 역진점수동등화 등 4가지였다.

분석과정에서 밝혀진 연구 결과 및 이를 바탕으로 내린 결론은 다음과 같다.

첫째, 공통문항의 경우 학년이 올라감에 따라 정답률이 높아지는 경향이 뚜렷하게 나타났다. 이에 비해서 비공통문항의 경우 난이도가 쉬워지는 경향성이 뚜렷하게 나타나지 않았다. 비공통문항의 정답률 증가패턴이 공통문항의 그것과 다소 차이가 있었다는 점은, 서로 다른 학년의 교육과정에 근거하는 문항의 경우 전문가에 의해서 유사 난이도라고 평가되더라도 상위 학년 교육과정에 근거한 문항이 하위 학년 교육과정에 근거한 문항에 비해서 더 어려울 수 있음을 의미한다. 이러한 현상은 수학 교과가 가지고 있는 위계성에 기인할 것이다.

예컨대 이차함수에 대한 개념 이해 능력은 일차함수의 이해를 바탕으로 하기 때문에, 각 단계에서 가장 기초적인 개념을 평가하는 문항이라 하더라도 이차함수의 난이도가 일차함수에 비해서 더 어려울 가능성이 높은 것이다.

둘째, 공통문항 중에서 학년이 올라갈수록 정답률이 상승하는 문항의 특징은 그 내용이 위계성을 가지고 있어서 보다 높은 수준에서 다시 한 번 다루어지며, 학생들이 평소에 제대로 이해하고 있는 기본적인 개념을 활용하면서 풀이 과정에서 추론 능력을 요구하는 문항이다. 이에 비해서 정답률이 하락하는 문항의 특징은 학년이 올라가더라도 다시 반복되거나 검토되지 않는 내용의 문항이다. 이것은 수학 학력이 내용을 기반으로 하는 인지적 능력임을 보여주는 것이다.

셋째, 수직적 동등화 상황에서 터커의 선형동등화, 레빈의 선형동등화, 빈도추정 동백분위 동등화 중에서 점수대별 동등화 표준오차가 가장 적은 방법은 '터커의 선형동등화'였으며 동등화 표준오차가 가장 큰 방법은 '빈도추정 동백분위동등화'였다. 이러한 결과는 '대교눈높이회원학력경시대회' 영어와 수학 검사를 이용한 수직적 동등화 방법들의 양호성 실증 비교를 통해서 도출한 부재율(2000)의 연구 결과와 그 맥을 같이 한다. 다만 문항반응이론에 근거한 동등화인 능력모수동등화, 진점수동등화, 역진점수동등화의 경우, 동등화 표준오차를 구하기 위해서 계속적인 재표집으로 문항모수치를 추정하는 절차를 반복해야 하기 때문에 문항반응이론에 근거한 동등화의 동등화 표준오차는 본 연구에서 제외하였다(부재율, 2000).

넷째, 공통문항을 이용하여 학년 간 수직적 동등화를 수행함으로써 학년이 올라감에 따른 학력 변화의 측정 가능성을 확인할 수 있었다. 김재철(2005)의 연구에서는 수학에 대한 태도 검사에서 수직적 동등화를 통한 태도의 변화 측정 가능성을 확인한 바 있다. 태도 검사의 경우 정답이 없으며, 여러 문항에 대한 일관적이고 전형적인 모습을 탐색하는 데 목적이 있기 때문에 반복 측정을 통해서 태도 수준에서 다소의 차이를 보이는 두 집단을 비교하더라도 문제가 발생하지 않는다. 그러나 학력 평가의 경우 정답이 있고, 새롭게 학습한 교육내용도 반영해야 하기 때문에 수준이 서로 다른 두 집단을 비교하기 위해서는 공통문항이 포함된 설계를 이용해야 한다. 공통문항의 정답률은 장기적인 학습 능력과 사고력보다는 그 문항에 관련한 개념을 학습한 시기에 강하게 영향 받기 때문에 학년이 올라갈수록 오히려 정답률이 낮아져서 학년이 올라가면서 절대적인 학력이 떨어진다는 결과를 얻을 수도 있다. 또한, 학년이 올라갈수록 내용이 심화되기 때문에 비공통문항의 경우도 정답률이 떨어질 수도 있다. 즉, 학력 평가의 경우 태도 검사와 달리 공통문항을 활용한 수직적 동등화는 한계가 있을 수 있다. 그러나 본 연구를 통해서 문항 제작 과정에서 공통문항과 비공통문항에 대한 기본적인 가정을 명확히 지킬 수만 있다면, 공통문항을 활용한 학력 변화의 측정이 가능함을 확인하였다.

다섯째, 4가지 동등화 방법으로 학년이 올라감에 따른 중학생의 수학 학력 변화를 추정한

결과, 변화를 가장 민감하게 반영하는 동등화 방법은 '진점수동등화'였고, 변화를 가장 둔하게 반영하는 동등화 방법은 '능력모수동등화'임을 확인할 수 있었다. 중학교 1학년의 점수를 평균 50, 표준편차 10인 점수로 척도화했을 때 진점수동등화에서는 중학교 2학년 54.6점, 중학교 3학년 55.9점이었고, 능력모수동등화에서는 중학교 2학년 51.7점, 중학교 3학년 52.4점이었다. 고전검사이론에 근거한 동등화 방법 중에서 동등화 표준오차가 작아서 가장 적합한 동등화 방법으로 제안할 수 있었던 터커의 선형동등화에서는 중학교 2학년이 51.7점, 중학교 3학년이 53.0점이었다.

여섯째, 동등화 방법에 상관없이 중학교 1학년과 중학교 2학년 사이의 변화가 중학교 2학년과 중학교 3학년 사이의 변화에 비해서 수학 학력 변화가 더 큰 것으로 나타났다. 이러한 경향성을 가장 크게 보인 동등화 방법은 '진점수동등화'로써 중학교 1학년과 2학년 사이에는 평균이 4.6점 증가한 반면, 중학교 2학년과 3학년 사이에는 평균 1.3점만 증가하였다.

일곱째, 동등화 방법으로 추정된 개인별 동등화 점수 간에는 다소의 차이를 발견할 수 있었다. 먼저 추정된 개인별 점수는 4가지 동등화 방법에 따라서 모두 통계적으로 유의미한 차이가 있었다. 그리고 4가지 동등화 방법으로 추정한 동등화 점수 간에는 상관이 매우 높기는 하지만 서로 직선적인 관계는 아님을 확인할 수 있었다. 동등화 방법으로 추정된 점수가 동등화 방법에 따라서 피험자의 서열이 바뀔 수 있는지 분석한 결과 터커의 선형동등화와 역진점수동등화, 능력모수동등화와 진점수동등화는 피험자 순위가 전혀 바뀌지 않았으나 터커의 선형동등화와 능력모수동등화, 터커의 선형동등화와 진점수동등화, 능력모수동등화와 역진점수동등화, 진점수동등화와 역진점수동등화는 동등화 방법에 따라서 피험자의 순위가 바뀔 수 있었다.

본 연구와 관련한 향후 연구 과제를 제시하면 다음과 같다.

첫째, 동등화를 활용하여 수학 학력 변화를 탐색하려는 연구를 확장하여 개인별 변화를 추정하는 종단 연구를 수행할 필요가 있다. 본 연구와 같이 특정 시점에서 여러 학년을 대상으로 공통문항을 포함한 검사를 수행하는 유사종단 연구에서는 집단적인 수준의 학력 변화만 추정할 수 있다. 이에 비해서 특정 연구 대상에게 시간적인 간격을 두고 공통문항을 포함한 검사를 반복해서 측정한다면 개인별 변화를 모형화하여 각 개인의 학력 변화의 개인간 차이를 다른 배경변인으로 설명할 수 있기 때문에 교육적으로 커다란 시사점을 얻을 수 있다(Willett, 1988; 김재철, 2002). 그 뿐만 아니라 개인별 학력 변화를 나타내는 기울기의 평균과 표준편차가 집단 간에 차이가 있는지를 분석함으로써 평준화 효과와 수준별 이동 수업의 효과 등과 같은 현안 문제에 대한 정확한 판단 근거를 제공함으로써 교육 정책 결정에 이바지할 수 있을 것이다.

둘째, 학년 간 변화를 도출하기 위한 방법으로써 공통문항을 이용하는 '가교검사 설계' 대신 공통피험자를 두는 '가교피험자 설계'의 활용 가능성을 연구할 필요가 있다. 가교검사 설

계는 공통문항을 이용하여 두 피험자 집단에서의 점수를 동일한 척도로 변환하는 데 비해서, 가교피험자 설계에서는 두 검사를 동시에 치르는 특별연구 집단(special study group)을 활용하여 두 피험자 집단에서의 점수를 동일한 척도로 변환한다. 가교검사 설계의 가장 큰 문제점은 종단 연구 상황에서 공통문항의 보안을 유지하기 어렵다는 것이다. 검사1을 피험자 집단1에서 특정 시점에 시행하고 정해진 시간이 지난 다음 집단2에게 검사2를 시행한다면, 공통문항에 대한 보안이 철저히 이루어지지 않는 경우, 피험자 집단1과 피험자 집단2의 능력수준을 심각하게 왜곡시킬 우려가 있다. 특히 동등화 설계를 3회 이상 적용할 경우 가교검사 설계는 피험자들로 하여금 공통문항의 존재 가능성을 알게 함으로써 보안의 문제점 통제가 거의 불가능하다. 하지만 가교피험자 설계는 동등화 설계의 마지막 단계에서 특별연구 집단을 대상으로 서로 다른 학년을 연결하기 위한 동등화 설계 검사를 한꺼번에 실시한다면, 피험자들에게 공통문항이 없다는 생각을 가지게 하여 이전 문제를 풀어보지 않게 할 수 있기 때문에 가교검사 설계보다 보안의 문제점을 줄일 수 있다. 요컨대 가교검사 설계뿐만 아니라 가교피험자 설계를 통한 수직적 동등화의 가능성도 확인할 필요가 있다.

셋째, 추후 연구를 통하여 터커의 선형동등화에서만 추정하였던 동등화 무선오차를 능력모수동등화, 진점수동등화, 역진점수동등화에서도 추정함으로써 동등화 결과의 양호성을 좀 더 엄밀히 비교할 필요가 있다.

참 고 문 헌

교육인적자원부(1999). **제7차 수학과 교육과정**.

구미종(2002). **패턴 블록을 활용한 분수학습에서 초등학교 4학년 아동의 학업성취도 및 태도에 관한 연구**. 석사학위논문, 이화여자대학교 교육대학원.

김동화, 김준홍(1999). 수학과 학력신장을 위한 수준별 교재의 재구성과 활용방안 - 중학교 2학년 학생중심으로 -. 경남대학교 교육문제연구소. **교육이론과 실천, 9**, 184-198.

김성훈(2000). 능력 변화과정 추정을 위한 시험자료의 동등화 방법 연구. **교육평가연구, 13**(1), 101-125.

김응태 · 박한식 · 우정호(1988). **수학교육학개론**, 서울: 서울대학교 출판부.

김재철(2002). **학생 배경변인과 수학에 대한 태도변화와의 관계 분석: 잠재변인 변화모형의 적용**. 박사학위논문, 서울대학교 대학원.

김재철(2005). 개인별 변화점수 산출을 위한 동등화 방법 비교. **교육평가연구, 18**(1), 1-26.

김현철(1999). 척도화방법들의 무작위오차 크기의 실증비교. **교육평가연구, 12**(2), 149-168.

남현우(2001). **검사동등화 방법**. 서울: 교육과학사.

남현우(2003). 2005수능의 탐구영역 내 과목간 점수조정에서 합집단 연계화방법의 적용가능성. **교육평가연구, 16**(1), 163-182.

남현우(2004). 대학수학능력시험 영역별 선택과목의 연계화 가능성 탐색. **교육평가연구, 17**(2), 93-105.

두산세계대백과사전(2005). http://100.naver.com/100.nhn?docid=185438.

류희찬 · 이기원(1994). 주관식 단계별 형성 평가 방법이 수학 성취도와 사고력에 미치는 영향. **대한수학교육학회논문집, 4**(2), 163-171.

부재율(2000). 수직적 동등화 방법들의 양호성 실증 비교-학년간 자료를 이용하여. **교육평가연구, 13**(1), 127-152.

부재율(2003). 문자점수간 비교성 확보를 위한 공통문항 이용 통계적 조정방법 탐색. **교육평가연구, 16**(1), 71-104.

부재율(2005). 가교문항 수직적 동등화를 이용한 국어, 영어, 수학 학업성취도의 학년 간 비교 연구, **교육평가연구, 18**(1), 61-80.

부재율 · 서동엽(2000). 수직적 동등화 결과 비교 · 분석 연구-단계형 수준별 교과(수학)에서-. **교육과정평가연구, 3**(1), 141-161.

성태제(2000). 다른 교과의 점수를 연결하는 통계적 조절 방법의 타당성. **교육평가연구, 13**(1). 63-99.

이종희 · 유현주 · 김선희 · 이진향(2002). **수학적 추론의 학습 지도-방안 탐색 및 현장 적용: 개연적 추론을 중심으로**. 2001년도 한국학술진흥재단 지원(KRF-2001-030-D00015) 연구보고서.

임선순(2000). **Elaborative feedback과 분할노트 기법이 수학불안, 수학태도, 시험불안, 수학 학업성취도에 미치는 효과**. 석사학위논문, 이화여자대학교 교육대학원.

정구향 · 김경희 · 김재철 · 반재천 · 민경석 · 이재기 · 박선미 · 진재관 · 조영미 · 이대현 · 이미경 · 신일용 · 김진석 · 이의갑(2004). **2003년 국가수준 학업성취도 평가 연구**. 한국교육과정평가원 연구보고 RRE 2004-16-1.

정범모(1971). 교육과 교육학. **신교육전서 1**. 서울: 배영사.

조용옥(1993). 발문으로 구성된 학습자료 개발적용을 통한 수학 학업 성취도 향상. **대한수학교육학회논문집, 3**(1), 157-173.

한국교육과정평가원(2004). **POUT, SCALE, CHECK, LISMIX, POLYEQUATE: 컴퓨터 프로그램**.

황혜정 · 최승현(1999). 수학과 평가틀에 관한 고찰. **수학교육학연구, 9**(2), 459-472.

Brown, R. M. (1979). *A Determination of Attitudes toward Mathematics and an Analysis of Factors Which are Associated with Negative Attitudes toward Mathematics of Students at an Urban Community College: A Descriptive Study*. Ph. D Dissertation, Heed University, *ED257658.*

Cook, L. L., Eignor, D. R., & Taft, H. L. (1988). A comparative study of the effects of recency of instruction on the stability of IRT and conventional item parameter estimates. *Journal of Educational Measurement, 25.*

Hambleton, R. K. & Swaminathan, H. (1985). *Item Response Theory*. Principles and applications. Boston: Kluwer.

Kim, S. & Lee, W. (2004). *IRT scale linking methods for mixed-format tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Kolen, M. J. & Brennan, R. L. (1995). *Test Equating: Method and Practices*. Springer Series in Statistics.

Linn, R. L. (1993). Linking results of distinct assessment. *Applied Measurement in Education, 6*(1), 83-102.

Muraki, E. & Bock, R. D. (1998). *PARSCALE: IRT and Test Scoring for Rating-scale Data*. Scientific Software International.

Poffenberger, T. & Norton, D. (1959). Factors in the Formation of Attitudes toward Mathematics. *Journal of Educational Research, 52.*

Schibeci, R. A. & RILEY, J. P. (1986). Influence of Students' Background and Perceptions on Science attitudes and achievement, *Journal of Research In Science Teaching, 23*(3), 177-187.

Simpson, R. D. & Troost, K. M. (1982). Influences on commitment to and learning of science among adolescent students, *Science Education, 66*(5), 763-781.

Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Rothkopf (ed.), *Review of research in education*, 345-422.

• *논문접수 : 2006년 4월 15일 / 수정본 접수 : 2006년 5월 15일 / 게재 승인 : 2006년 5월 24일*

ABSTRACT

The Calibration of the Mathematical Scholastic Ability Changes of Middle School Students through Vertical Equating Including Anchor Items

Jae-Chul Kim(Professor, Hannam University)
Sun-Hee Kim(Research Fellow, Korea Institute of Curriculum & Evaluation)

The purpose of this study is to identify the possibility of using the equating methods in order to find out the growth pattern of the mathematical scholastic ability in middle school students. The data used in this study were the results from 1,479 who consisted in 472 7th graders, 511 8th graders and 496 9th graders. The number of item was 37 in each grader, some of which were common items. The equating methods used in this study were Tucker's linear equating, ability parameter equating, true score equating and inverse true score equating. The findings of the research are as follows: First, in the case of common items, the higher the grader was, the easier the difficulty of item was. Second, if the concept contained in a certain item had hierarchy and was dealt with higher level repeatedly, the difficulty of the item became easier to the higher graders. Third, the smallest method in equating error was the Tucker's linear equating. Fourth, the changes of mathematical scholastic ability through using the equating methods were identified. Fifth, there were some differences among scale scores estimated from four equating methods. Further studies, however, are required to investigate whether it's possible to apply how to calculate individual change scores using equating methods to achievement tests. It also need investigating applicability of anchor-person design instead of anchor-test design in order to apply to the equating methods.

Key Words : anchor item, vertical equating, mathematical scholastic ability, change pattern