

연중 수학과 성취도 변화 연구: 초등학교 5학년 학생들을 중심으로¹⁾

김 선(대전석교초등학교 교사)

《 요 약 》

본 연구의 목적은 초등학교 5학년 학생들의 수학과 성취도가 연중 어떻게 변화하는지, 수학과 학습내용 영역별 성취도는 어떻게 변화하는지를 알아보는 데 있었다. 본 연구를 위해 네 개의 검사지를 동등화하였으며, 동일한 학생들에게 네 시점에 걸쳐 반복 시행하였다. 연구대상은 대전지역의 S초등학교 150명이었다.

연구결과는 다음과 같다. 첫째, 학생들의 연중 수학 성취도 평균 점수는 시점이 지남에 점차 증가했다. 그러나 평균적인 변화량이 시점1에서 시점2까지(1학기 말)는 컸으나 점차 변화량이 적었다. 둘째, 수학 성취도 점수의 분산은 시점1 대비 시점4에서 17.8% 증가했다. 셋째, 수학과 하위 학습내용 영역별 성취도 점수 평균은 시점이 지남에 따라 대체로 증가하였다. 넷째, 수준설정에 의해 판정된 우수, 보통 집단에서는 수학 성취도 점수의 평균 변화가 시점이 지남에 따라 증가했다. 반면 미달집단은 시점2까지는 조금 증가하고 시점3부터는 오히려 점차 감소했다. 학습내용 영역별 성취도점수의 평균변화도 미달 집단은 대부분의 영역에서 성취도 점수가 향상되지 못했다. 따라서 미달 집단에 대한 지도대책을 보다 체계적이고 다각적으로 검토하면서 교육기관의 책임을 다하도록 노력해야 함을 시사하고 있다.

주제어 : 성취도 변화, 학습능력 성장, 검사동등화, 수준설정

1) 박사학위논문

I. 서론

1. 연구의 필요성 및 목적

학생들은 학교 교육을 통해 다양한 측면에서 변화하고 성장한다. 시간의 경과에 따른 개인 변화의 측정은 교육평가 및 측정 분야에서 매우 중요한 주제이다. 학업성취도의 변화 정도는 좁게는 학생들의 학습을 진단하여 지도하는 데 활용되며, 넓게는 국가 교육과정의 적절성 평가 및 학교의 책무성 자료로 활용될 수 있다.

국내에서 학생의 학업성취도 변화 추이를 종단적으로 분석하기 위한 연구들이 진행됐다. 예컨대, 한국교육개발원에서는 학생들의 학업성취도를 종단적으로 평가하기 위해 ‘한국교육종단연구 2005’연구를 시행하고 있다(류한구 등, 2005). ‘한국교육종단연구 2005’ 연구에서는 표집 학생들을 대상으로 중학교 1학년부터 3학년까지 매학년 성취도 평가를 시행했다. 또한 각 학년에서 얻은 검사 점수를 성취도 점수의 변화로 보기 위해 학년별 척도점수를 비교할 수 있는 공통 척도를 개발했다. 이러한 척도개발을 통해 학생들의 학업성취도가 연간 얼마만큼 변화하는지를 분석할 수 있게 되었다. 그러나 이 연구는 학생들의 연중 학업성취도의 변화가 아닌 연간 학업성취도 변화를 알아보기 위한 것이다.

전체적인 학업성취도 변화 추이보다는 개인수준에서 변화 추이를 본 국내 연구들(권재술 등, 1999; 김선옥, 2005; 김성훈·강상진·김성호, 1999; 이성원, 2004)이 있다. 그런데, 이들 대부분은 두 시점 간 결과를 비교하거나, 평균 기술, 종속 표본 t검정, 상관계수 산출에 그쳐 단순히 시점 간, 집단 간 차이만을 보는 경우가 많았다. 학생들의 학업성취도의 연중 변화를 알아보기 위해 동등화 과정을 거친 검사지를 개발하고, 수준설정을 하여, 여러 시점에서 학생들의 능력을 측정하여 학생들이 어느 정도 그리고 어떤 양상을 보이면서 변화하는지에 대한 연구는 많지 않아 보인다.

국가수준에서도 국가수준 학업성취도 평가를 통해 학생들의 학업성취도 변화를 파악하고자 하고 있다. 하지만 국가수준 학업성취도 평가는 연간 변화 추이(정은영 등, 2009; 조지민 등, 2007)를 파악하는 데 활용될 뿐, 학생들의 연중 학업성취도 변화에 대해서는 정보를 제공하지 못하고 있다. 학급담임이 주요 교과를 모두 가르치는 우리나라 초등학교 상황에서 담임교사가 1년 동안 학생들을 가르치는 데 중요한 정보는 학생들이 얼마나 성장하고 있는가, 즉 학생들의 학업성취도가 한 학년도 내에서 얼마만큼 변화하고 있는가에 관한 것이라고 할 수 있다. 그러나 지금까지 국내외에서 학생들의 연중 학업성취도 변화를 살펴보는 연구는 많지 않아 보인다. 따라서 본 연구에서는 학생들의 연간이 아닌 연중 학업성취도 변화에 초점을 두어 일선 교사들이 교수-학습에 보다 직접적으로 시사 받을 수 있는 정보를 탐구하고자 한다.

물론 일선 학교에서는 대부분 연중(즉, 한 학년도 동안) 1학기에 중간평가와 기말평가, 2학기에 중간평가와 학년말평가, 총 4회의 평가를 실시하여 학생들의 성적향상 정도 및 성취목표 달성정도를 알아보고자 한다. 그런데 문제는 각 시험의 난이도와 검사내용이 다르기 때문에 네 번 시행된 시험의 점수 차이를 학생들의 학업성취도가 변화한 것으로 해석하기 어렵다는 점이다. 학생들의 학업성취도가 얼마나 변화하였으며, 학생들이 어느 성취수준에 도달해 있는지를 정확히 알기 위해서는 검사의 난이도를 통제한 후 학생들의 학업성취도 변화를 파악해야 한다. 그렇게 해서 얻어진 점수들 간의 차이는 학업성취도 변화 정도를 나타낼 뿐만 아니라, 일정 점수 이상을 획득하면 학생들의 성취수준이 어느 정도인지에 대해서도 파악할 수 있기 때문이다. 동등화 과정을 거친 검사를 개발하여, 수준을 설정한 후, 연중 학업성취도가 어떻게 그리고 얼마나 변화하는지를 파악하는 것은 교수학습 과정에 교사에게 매우 중요한 정보를 제공하기 때문에 이러한 연구가 이루어질 필요가 있다.

일년 동안 학생들이 배워야 하는 교과와 내용은 하위 학습내용 영역으로 구성되어 있으며, 교사가 각 학습내용 영역에서 학생의 변화 정도를 파악하는 것이 의미 있는 일이다. 어떤 학습내용 영역에서 보다 많은 변화가 있고 어떤 학습내용영역에서 학생들의 향상이 어려운지를 파악할 수 있기 때문이다. 또한 어떤 학습내용 영역에 대해서 학생 간 편차가 크거나 작은지를 파악할 수 있기 때문이다.

각 학습내용 영역에서 변화 정도를 알아보려면, 학습 내용이 위계적으로 구성되어 있는 교과가 연구 교과로 적합하다 할 수 있다. 그렇다고 해서 성장을 알아보기 위해 모두 엄격한 위계가 필요한 것은 아니다. 다른 영역을 순차적으로 학습해나가는 것도 성장으로 볼 수 있다. 본 연구에서는 학습내용이 상대적으로 위계적으로 구성되어 있으면서 주지과목인 수학과를 선택하여 학생들의 학업성취도 변화에 대해 연구하고자 한다.

또한 측정 시점에 따른 수학 학업성취도 점수들의 평균적 변화와 그 점수들의 분산 패턴에 관한 연구는 학생들의 수학 학업성취도 변화 정도에 대한 정확한 측정에 중요할 뿐만 아니라 학생들의 학습지도에도 중요한 정보이다. 따라서 본 연구에서는 하위 학습내용 영역별 성장 정도 및 패턴에 대해서도 전체집단 및 능력집단별로 나누어 연구하고자 한다.

본 연구의 목적에 따른 연구문제를 구체적으로 진술하면 다음과 같다.

첫째, 초등학교 5학년 학생들의 연중 수학 성취도는 어떻게 변화하는가?

둘째, 초등학교 5학년 학생들의 연중 수학과 하위 학습내용 영역별 성취도는 어떻게 변화하는가?

본 연구는 학생들의 연중 수학 성취도가 총점 및 하위 학습내용 영역 점수에서 어떻게 변화하는지를 전체 집단 및 능력집단별로 나누어 탐구함으로써 교수학습에 시사점을 제공하는 데 의의가 있다. 본 연구는 한 초등학교에서 이루어졌던 결과를 토대로 한 사례연구이다.

Ⅱ. 이론적 배경

1. 성장에 대한 개념 정의

학교교육에서 개인의 인지적 성장을 측정하기 위해서는 학생들이 학교에서 학습한 것이 무엇이며, 그것을 얼마나 성취했는가를 측정할 수 있는 내용으로 구성된 동형 검사형을 개발할 필요가 있다. 학교 교육과정과 밀접히 관련되어 있는 검사를 개발할 때, '성장'에 대한 개념적 정의를 분명히 하는 것이 매우 중요하다. 성장에 대한 두 가지 개념적 정의가 있다. 첫째는 전체 영역 내 성장(domain growth)이고, 둘째는 학년 간 성장(grade to grade)이다(Kolen & Brennan, 2004, p. 376). 전체 영역 내 성장이란 모든 학년의 전체 영역(domain)을 포괄하는 시험에서의 변화를 의미한다. 검사는 해당 학년에서 배우는 내용뿐만 아니라 다른 학년에서 배우는 내용까지 모두를 포함하고 있다. 따라서 검사 문항은 일반적으로 매우 길며, 문항의 내용 및 난이도도 매우 다양하다. 동일한 시험지를 모든 학년에게 실시하여, 원점수를 얻고 이를 척도점수로 변환한다. 이처럼 성장이란 '모든 학년의 내용을 포함하는 전체 영역에서 개인점수의 변화'라고 정의할 경우, 실제로 이를 반영한 검사를 시행하기가 어렵다. 검사가 너무 길고, 학생들의 수준에 안 맞는 문항들이 많기 때문이다.

또 다른 정의는 성장이란 '특정 학년에 배웠던 내용에서 개인점수의 변화'라고 정의하는 것이다. 그래서 '학기 초와 다음 학년 학기 초(예, 3학년 3월 시험과 4학년 3월 시험)에 시험을 실시하여 얻어진 점수 변화가 곧 성장이다'라고 개념화하는 것이다.

그런데 대상학년이 전체 학년이 아닐 경우, 성장에 대한 정의를 수정할 필요가 있다. 본 연구에서는 단일 학년 내용영역을 모두 포함하여 검사를 제작하여, 학년 교육과정 진도와 함께 학생들의 학업성취도가 어느 정도 성장하였는지를 밝히고자 하였다. 따라서 본 연구에서의 성장이란 모든 학년이 아닌 단일 학년 내에 포함된 영역에서의 변화를 의미한다. 다만, 본 연구에서는 성장과 때로는 변화라는 용어를 혼용하여 쓰기로 한다.

2. 학업성취도 변화에 관한 선행 연구

학생들의 학업성취도 변화에 대한 학년 간 종단적 연구에서 발달점수의 평균과 분산의 패턴에 대한 연구가 수행되어 왔다. 국외의 경우 연간(year-to-year) 혹은 학년 간(grade-to-grade) 성장을 평가하기 위해 발달점수 척도(developmental score scale)를 개발하고 활용하는 연구가 활발히 진행되어 왔다(Kolen, 2006). 초등학생들의 "연간" 학업성취도의 변화 패턴과 관련하여, Hoover(1984)는 발달점수 척도는 학년이 올라갈수록 학년 내 분산(within-grade

variability)은 증가하고 학년 간 평균 성장은 점점 감소한다고 주장했다. 그러나 이러한 주장은 일부 연구들(Andrews, 1995; Hoover, 1984; Yen, 1986)에서 확인되었을 뿐 여전히 후속 연구의 필요성이 제기되고 있다(Kolen, 2006). 이러한 연구의 필요성은 “연중” 변화 연구에도 그대로 적용된다. 즉, 학년 간 성장에 따른 발달점수의 평균과 분산의 패턴이 학년 내 변화와 성장에서도 여전히 성립할지에 대해 알아볼 필요가 있다. 학생들의 학업성취도 변화의 패턴이 학년 내에 어떤 양상을 보이는지를 살펴보는 것은 학년 밀착형 지도 방안의 관점에서 바라볼 때 현장에 매우 유용한 정보이다. 왜냐하면 교사는 연간 교육과정 진도표에 의해 교육활동을 계획하고 지도하기 때문이다.

외국의 연구에 따르면, 학년 간 성장은 학년이 증가할수록 감소한다고 알려져 있다(Kolen & Brennan, 2004). 즉, 3학년, 4학년 간의 학업성취도 평균 성장에 비해 4학년, 5학년 간 학업성취도 평균 성장이 감소한다는 것이다. 반면 분산은 3학년보다는 4학년이, 4학년보다는 5학년이 커진다고 한다. 그런데 이러한 현상이 한 ‘학년 내에서도’ 나타나는지 알아보는 것은 의미 있는 일이다. 성적 향상이 보다 많이 나는 학습내용 영역과 그렇지 않은 영역이 있을 경우, 보다 효율적 교수 학습을 위해 교육과정 및 교육시간 운영에 참조할 수 있기 때문이다.

학년 간 성장을 측정하고자 하는 연구에서 문항반응이론에 의한 점수 척도의 분산이 지속적으로 증가 혹은 감소하는지에 대해 논란이 많았다. Omar(1998)는 IRT를 이용했을 때 학년 증가와 함께 분산이 점차 줄어든다고 하였다. 반면 Becker와 Forsyth(1992)는 고등학생용 시험에서 학년 간 성장의 분산이 줄어들지 않는다고 하였다. Williams, Pommerich, & Thissen(1998)은 학년 간 성장의 분산이 줄어드는 이유가 결합최대우도(joint maximum likelihood) 방법에 의한 능력 추정 방법 때문이라고도 하였다. 이처럼 외국에서는 학업성취도의 향상도 형태와 분산에 대해 여러 논쟁이 있었다.

국내에서도 학생들의 학업 성취도의 변화를 살펴보기 위한 연구들이 진행되어 왔다. 김성훈 등(1999)은 중학생들의 수학 월말고사들을 사용하여 다층모형에 의한 수학 능력의 발달과정을 추정하고자 하였다. 4회의 월말고사를 5회째 실시된 가교검사에 동등화한 결과, 학생들의 능력이 계속 줄어들거나 혹은 계속 줄어들거나, 늘었다가 줄어드는 현실적으로 해석이 어려운 발달 궤도가 추정되었다. 이러한 결과는 동등화 결합이나 능력 추정을 위한 자료의 불충분에 기인한다고 보았다. 후속연구에서 김성훈(2000)은 1개월 더 늘어난 3월에서 7월까지의 매 월말고사를 9월의 가교검사에 동등화 시켰다. 그런데 이 두 연구에서 사용한 월말고사는 특성상 시험범위가 매달 배운 내용으로 제한되어 있어서 연중에 걸쳐 다뤄지는 전체 하위영역에서의 학생들의 변화 양상을 파악하기 어렵고, 또한 능력집단별로 하위영역에서의 변화 양상이 다를 수 있는데 이러한 부분에 대해서는 다루어지지 못했다.

임현정(2006)은 학업성취도의 추이 분석을 통해 특정 교육과정이나 교육정책이 시행된 기간 동안에 일어난 학업성취도의 변화 정도를 파악할 수 있으며, 이를 통하여 교수·학습 활동의 개

선, 교육과정의 개발, 교육정책의 수립에 필요한 정보를 제공함으로써 교육 전반의 질 개선에 기여할 수 있다고 하였다. 그의 연구는 가교검사가 없는 연도별 검사 자료를 상호비교가 가능하도록 동등화하고 국가 수준에서 학업성취도의 연도별 추이를 측정학적으로 가늠해 볼 수 있는 방법론을 제시하였다는 데 의의가 있으나, 매년 검사도구가 다르고, 피험자도 다른 학교수준의 반복측정으로, 개인 수준에서 하위영역에서의 변화를 다루지 못했다.

박정(2008)은 TIMSS 2003년 자료를 통한 분석에서 학생수준 및 학교수준 변수들이 능력 집단에 따라 영향을 달리 준다고 보고하였다. 송미영(2010)은 2009년 학업성취도 평가 자료를 이용한 분석에서 교육맥락변수가 수학교과에서 학생의 학업성취수준에 따라 다르게 작용한다고 보고하였다. 이들 연구는 성취수준에 따라서 학업성취도 변화 패턴이 달라지는가를 알아봐야 함을 시사하고 있다.

지금까지 학생의 학업성취도 변화와 관련한 국내외 연구들을 고찰한 결과, 학생들의 '연중' 학업성취도가 전체점수 및 하위 영역별로 어느 정도이며, 또한 능력 집단수준에 따른 변화 양상에 대한 탐구가 부족함을 알 수 있다. 따라서 본 연구에서는 특정 학년 내에서 학생들의 성적 향상 정도의 평균 및 분산이 전체 및 하위 영역에서, 그리고 집단별로 어느 정도인지에 대해 수학과를 중심으로 탐구하고자 한다.

Ⅲ. 연구방법

본 연구의 목적을 달성하기 위해 초등학교 5학년 교육과정 내용을 모두 포함하는 수학과 검사지 4개(가형, 나형, 다형, 라형)를 개발하였다. 4개 검사 간 난이도를 조정하기 위해서 별도의 동등화용 집단에게 해당 교육과정을 모두 이수한 시점인 학년말 2월에 시험을 실시하였다. 동등화용 집단은 5학년 27학급 약 2,500명으로 네 개의 검사를 순차적으로 섞은 후 학급별로 학생들에게 앓은 순서대로 배포하였다. 즉 학생들은 자리배치 순서대로 가, 나, 다, 라 형을 받아서 시험을 치르게 되었다. 이렇게 하여 얻어진 동등화 과정을 거친 4개의 검사는 연구대상 집단인 초등학교 5학년 5개 반의 학생들에게 일정한 간격을 두고 4회 시행되었다. 그리고 학생들의 성취수준을 알아보기 위해 가형을 기본 검사형으로 하여 미달, 보통, 우수의 세 수준으로 나누는 두 개의 분할점수를 산출하였다.

이렇게 해서 얻어진 자료를 통해 학생들의 수학 성취도 점수의 평균적인 변화 정도, 분산 등이 어느 정도 다른지를 알아보았다. 본 연구에서 사용한 연구방법 및 절차에 대해 구체적으로 기술하면 다음과 같다.

1. 연구대상

본 연구를 수행하기 위해 4번 반복 측정한 연구대상은 초등학교 5학년에 재학한 대전지역의 S초등학교 150명의 학생이었다. 대전지역은 대도시이나, S초등학교는 아파트 밀집지구가 아닌, 개인 주택 지구로 비교적 인구 이동이 적어 전출입 학생의 수가 1년에 전체 학생수의 3% 이내로 동일한 집단에 대해 지속적인 변화를 관찰하기 위한 연구를 수행하는데 적절하다고 판단되었다.

연중 수학적성취도 변화를 평가하기 위해 4개 동등화 과정을 거친 검사형의 시행 시기는 1차는 2009년 3월 말, 2차는 2009년 7월 중순, 3차는 2009년 12월 중순, 4차는 6학년 진급 후 2010년 3월 말이었다. 이와 같이 검사의 시행시기를 결정한 이유는 우리나라 2학기제의 학제 운영 방식의 특징을 반영한 것으로 검사 시행 시기 간의 간격이 3차까지는 여름 방학을 제외하고 약 3.5개월이다. 4차의 시기는 처음 시행시기로부터 1년(12개월)이 되는 시점에 시행을 하였다.

2009년 3월부터 2010년 3월까지 4번의 검사를 시행하는 동안 결석이나, 공적인 학교행사 참여로 인한 결시, 조퇴, 전출, 전입으로 인해 한번이라도 검사를 치르지 못한 학생들 25명을 제외하고 연구대상은 최종 125명이었다.

2. 연구도구

본 연구를 위해서 초등학교 5학년용 수학적성취도 검사 4종을 개발하였다. 수학 성취도 검사의 내용 범위는 초등학교 5학년 교육과정에 포함된 수와 연산, 측정, 확률과 통계, 규칙성과 함수, 도형, 문자와 식의 6개 학습내용 영역을 포함하였다. 즉, 초등학교 5학년 학생들이 1년 동안 배워야 하는 내용을 모두 포함하여 검사를 제작하였다. 4종의 검사는 동일한 검사 명세서에 의하여 초등학교 5학년 담임교사 6명이 각 하위 학습내용 6개 영역에 대해서 하나의 영역씩 담당하여 10문항씩 출제하였다. 문항검토와 예비시행을 통해 하위 학습내용 영역별로 5문항씩 선제하였다. 최종 4개의 검사형은 6개의 학습내용 영역별로 5문항씩 총 30개의 5지-선다형 문항으로 구성되었다. 초등학교 5학년 학교 교육과정 운영에서 각 학습내용 영역은 연간 교육과정 진도표에 의해 지도하며, 그 자세한 내용은 [그림 1]과 같다.

학습내용 영역이 다루어지는 시기와 관련하여 학습 내용의 양을 알아보기 위해 학습 지도 단원수를 알아보았다. <표 1>은 학습내용 영역별 지도 단원 수인데, 표에 나타났듯이 학습내용 영역별로 학생들이 학습해야 할 단원의 수가 매우 다르다. 특히 '수와 연산' 영역은 총 7단원에 걸쳐 지도하도록 되어 있다. 학습내용 영역별 단원 수가 다름에도 학습내용 영역별 문항수를 동일하게 하는 이유는 각 학습내용 영역별 변화를 알아보기 위해서는 일정한 수의 문항이 필요하기 때문이다. 따라서 확률과 통계 영역처럼 단원 수는 비록 1단원에 불과하지만 이 영역에서 출

제된 문항은 5문항이 되도록 하여 이 학습내용 영역에서 학생의 성취도 변화를 측정할 때 신뢰도를 높이고자 하였다.

영역 \ 월	3	4	5	6	7	9	10	11	12	2
수와 연산										
규칙성과 함수										
도형										
측정										
문자와 식										
확률과 통계										

[그림 1] 초등학교 5학년 수학과 학습내용 영역별 연간지도 계획

네 개 검사형은 동일한 검사 개발 지침 및 내용명세표에 의해 제작하였다. 초등학생들의 수업 시간이 40분임을 감안하여 문항 수는 30문항으로 제한하였다. 각 검사에서 만점은 30점이다.

성취수준의 변화를 보기 위해 기준설정과 검사동등화를 수행하려면 기본 검사형이 필요하였다. 네 개 검사형은 개발 과정에서 가능한 유사하도록 개발하였기 때문에 어떤 검사형을 기본 검사형으로 사용하든 상관은 없었지만 가형을 기본 검사형으로 선정하였다.

본 연구를 위해 개발된 4개 검사형을 BILOG-MG 프로그램(Zimowski, Muraki, Mislevy, & Bock, 2003)을 이용하여 문항분석을 한 결과 전체적으로 정답률은 대체로는 .5 ~ .9 사이에, 변별도는 .3 ~ .5 사이에 분포되어 있는 문항이 많았다. 각 검사형의 신뢰도를 구한 결과 Cronbach α 가 가형이 .853, 나형은 .875, 다형은 .842, 라형은 .829로 모두 높게 나왔다. 그러나 하위 학습내용 영역별로는 하위 문항수가 5문항인 관계로 Cronbach α 가 .5 이하인 영역들도 있었다.

<표 1> 초등학교 5학년 수학과 학습내용 영역별 지도 단원 수 및 신뢰도

영역	지도 단원 수			신뢰도			
	1학기	2학기	합계	가형	나형	다형	라형
수와 연산	4	3	7	.336	.679	.604	.519
규칙성과 함수	1	0	1	.512	.539	.401	.429
도형	1	2	3	.368	.605	.534	.506
측정	1	1	2	.569	.498	.578	.566
문자와 식	1	1	2	.619	.719	.677	.596
확률과 통계	0	1	1	.595	.546	.469	.453
합계	8	8	16	.853	.875	.842	.829

3. 검사동등화와 기준설정

네 개 검사형 간 난이도 차이를 통제하기 위해 검사동등화를 하였다. 검사동등화를 위한 자료는 무선동등집단과 공통문항 비동등집단을 혼합하여 수집하였다. 검사동등화용 자료는 검사동등화만을 위한 별도의 자료로 본 연구에서 분석에 사용한 자료와는 별개로 본 연구의 초점이 아니므로 간단히 기술한다. 검사동등화와 기준설정에 관한 자세한 절차는 김선, 김성훈, 반재천(2009)의 연구에 기술되어 있다.

검사동등화를 위해 3모수 로지스틱 문항반응이론(IRT)에 의해 문항 모수를 추정하였다. 문항모수 추정은 BILOG-MG 프로그램을 이용했다. 검사동등화 기법으로 IRT를 이용한 진점수 검사동등화 방법을 사용했다. 네 개 검사형 간의 IRT 진점수 동등화를 위해 가형을 기본형으로 하였으며, 기본형인 가형에 나, 다, 라 형을 총점 수준과 하위 학습내용 영역별로 동등화했다. 모든 동등화 과정은 컴퓨터 프로그램 EQUIRT (Kim, 2008)를 사용했다.

학생들을 우수, 보통, 미달과 같은 성취수준으로 분류하기 위해 기준설정 기법을 사용하였다. 본 연구에서는 국내외에서 많이 사용되고 있는 앵고프 방법과 북마크 방법을 사용하여 가형에 대해 두 개의 분할점수를 찾았다. 첫 번째 분할점수는 미달/보통을 구분하는 점수이고, 두 번째 분할점수는 보통/우수를 구분하는 점수이다. 기준설정을 위한 패널은 14명의 초등학교 교사 및 수학전문가 집단으로 구성하였으며, 두 가지 기준설정 방법에 동일 패널을 사용하였다. 기준설정을 위해 일반적으로 하나의 방법만을 선택하지만, 기준설정 방법에 따라 분할점수가 다르기 때문에 본 연구에서는 두 방법을 모두 적용하여 두 방법에서 얻어진 분할점수의 평균을 사용하였다. 그렇게 함으로써 기준설정 방법으로 인해 연구결과가 달라지는 등의 영향을 최소화하고자 하였다.

4. 자료 분석 방법

본 연구에서의 자료분석은 SPSS 18.0 프로그램을 이용하여 분석하였다. 첫 번째 연구문제와 관련한 통계적 방법들을 차례로 제시하면 다음과 같다. 학생들의 시점별 평균 성적 변화량을 알아보기 위해 점수산출 방법에 따라 평균 증분값과 효과계수를 산출하였다. 평균 증분값은 시점 간 평균값 차이를 단순계산 하였으며, 효과계수의 산출 공식은 다음 식 3.1과 같다.

$$\text{효과계수} = \frac{\bar{X}_{t+1} - \bar{X}_t}{S_{pooled}}, \quad t = 1, 2, 3 \quad (\text{식 3.1})$$

$$S_{pooled} = \sqrt{\frac{(N_{t+1} - 1)S_{t+1}^2 + (N_t - 1)S_t^2}{N_{t+1} + N_t - 2}}$$

위에서 S_{pooled} 는 연속된 두 시점의 통합 표준편차이며, t 는 시험 시행 시점을 이야기하며, $t+1$ 은 시점 t 의 바로 다음 시점을 말한다. 통합 표준편차로 연속한 두 시점 간 평균차이를 나눠 줌으로써 효과계수를 산출하였다. Cohen(1988)에 따르면, 효과계수값이 .2면 “작음,” .5면 “중간,” .8이면 “큼” 으로 간주된다.

또한 학생들의 수학 성취도가 능력집단에 따라 어떻게 변화되는지 알아보았다. 이를 위해 네 번째 마지막 시점에서 얻은 성취도 점수에 의해 분류된 성취수준을 기준으로 3개의 미달, 보통, 우수 집단으로 나눈 후, 이들 집단이 각각 시점 1부터 시점 4까지 수학 성취도 점수의 평균적인 변화가 어떻게 되는지 분석했다. 시점 4를 기준으로 집단을 나눈 이유는 모든 과목을 다 이수한 내용을 바탕으로 한 시험이므로, 이 시험결과를 통해 집단 구분하는 것이 타당하기 때문이다. 학생들의 수학 성취도의 변화를 통계적으로 검증하기 위해서는 일요인 반복측정설계 방법으로 분석하였고, 성취수준 집단별의 시점별 평균차이를 통계적으로 검증하기 위해서는 스플릿-플랏 방법을 이용하였다. 집단간 평균차이가 유의한 경우에는 Bonferroni 방법으로 사후검증했다.

학생들의 수학 성취도 점수의 시점별 분산 변화를 알아보기 위해서 시점별로 분산을 산출하였다. 시점별로 관찰된 분산 차이가 통계적으로 유의한지에 대한 분산차이 검증법(임인재·김신영·박현정, 2003, p. 246)을 사용하였다. 본 연구에서는 반복 측정을 했기 때문에 각각 얻어진 두 분산은 서로 독립적이지 않으며, 종속표집으로 t -분포를 이용한다.

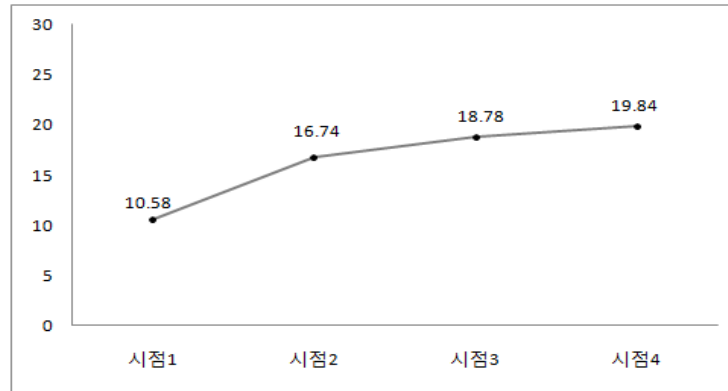
두 번째 연구문제와 관련하여 초등학교 5학년 수학과 학습내용 영역별로 점수의 평균적 변화를 일요인 반복측정설계법으로 검증하고, Bonferroni 방법으로 사후검증했다. 또한 최종 성취 수준에 비추어 미달, 보통, 우수의 세 집단으로 분류하여 각 집단의 학습내용 영역별 성취도 점수의 평균 변화를 계산하여 그래프로 제시하였다.

IV. 연구결과

1. 초등학생들의 연중 수학 성취도 변화

가. 수학 성취도 점수의 평균 변화

초등학교 5학년 학생들의 수학 성취도 점수의 평균 변화는 시간이 지남에 따라 증가 하는 모습을 보였다. 학생들은 학교 교육과정에 의해 학습을 하게 되므로, 5학년 수학과 교육과정 전 과정의 내용을 모두 담고 있는 검사의 특성상 시간이 지남에 따라 학생들의 수학 성취도 점수도 올라가는 것은 예상할 수 있는 결과였다.



[그림 2] 수학 성취도 점수의 시점별 평균 변화

그러나 시점에 따른 평균 성적 변화가 일정한 것은 아니었다. [그림 2]는 각 시점에서 평균값을 그래프로 나타낸 것이다. 수학 성취도 점수는 시점1과 시점2 사이의 평균 변화의 기울기가 가장 컸고, 이후 점차 평균 변화의 기울기가 완만했다.

〈표 2〉에서 시점 간 차이점수의 평균은 시점2-시점1은 6.17, 시점3-시점2는 2.04, 시점4-시점3은 1.06이었다. 이 결과로 보면 학생들의 수학성취도의 평균변화 크기는 초기에는 컸지만 증분량이 점차 감소($2.04/6.17=.33$, $1.06/2.04=.51$)했다. 일요인 반복측정설계를 통한 유의도 검증결과 $F=99.248$ 로 5% 수준에서 유의했다. Bonferroni 사후검증법을 이용하여 평균 차이에 대한 유의도 검증 결과 시점4-시점3간에는 통계적으로 유의하지 않았다. 즉, 시점1부터 시점3까지는 평균차이가 통계적으로 유의하게 증가했으나, 시점3과 시점4간에는 통계적으로 유의한 차이가 없었다.

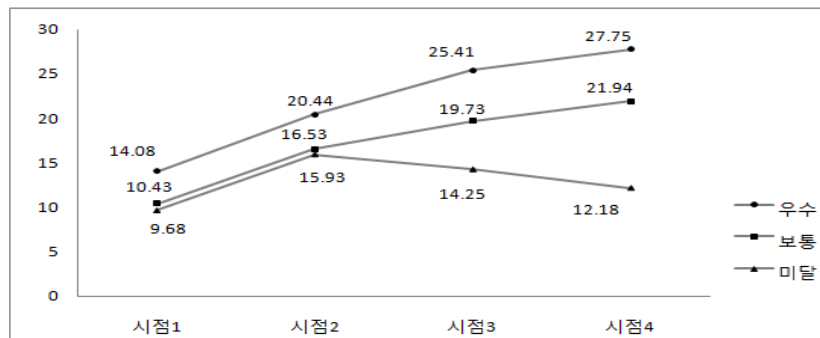
효과계수는 통계적 유의성이 아닌 실제적으로 어느 정도 의미 있게 평균이 변화했는지를 알아보기 위한 지수다. 효과계수를 보면 시점2와 시점3을 제외하고 모두 .8이상으로 Cohen의 효과크기 크기의 기준을 적용했을 때 효과가 컸다. 효과크기에서도 시점이 지남에 따라 학생들의 평균적인 성장은 줄어들고 있다는 것을 보여주었다. 또한 시점4와 시점1간에는 효과계수는 1.71로 크게 나왔다.

〈표 2〉 시점 간 평균 차와 효과 크기

시점2-시점1		시점3-시점2		시점4-시점3		시점4-시점1	
평균차	효과크기	평균차	효과크기	평균차	효과크기	평균차	효과크기
6.17**	1.16	2.04**	0.39	1.06	0.20	9.27**	1.71

$F(3,122)=99.248^{**}$

** $p<.01$



〔그림 3〕 최종 성취수준 집단별 평균 변화

나. 성취수준 집단별 수학 성취도의 평균 변화

학생들의 수학 성취도가 능력집단(성취수준)에 따라 어떻게 달라졌는지를 알아보았다. 이를 위해 먼저 최종 수학 성취도 점수를 분할점수에 따라 우수, 보통, 미달로 구분하였다. 이렇게 분류된 각 집단별 수학 성취도 점수의 평균적인 변화를 분석한 결과는 [그림 3]과 같다.

최종 성취수준이 우수한 집단과 보통인 집단은 모든 점수산출 방법에서 시점이 지남에 따라 평균이 증가했다. 그러나 최종 성취수준이 미달인 집단은 평균 점수를 보면 시점1에서 9.68, 시점2에서 15.93으로 평균이 증가하였다가 시점3에서 14.25로, 시점4에서는 12.18로 평균이 감소했다. 이 결과는 성취수준 집단별로 성취도 향상 패턴이 다를 수 있음을 나타내고 있다. 스피릿 플랏 방법으로 평균차이를 검증한 결과 시점과 성취수준간 상호작용 효과가 $F=16.72$ 가 5% 수준에서 유의하였다(〈표 3〉 참조). 따라서 성취수준 집단별 향상 패턴이 다름이 통계적으로도 뒷받침되고 있다. 또한 주효과인 성취수준과 시점도 5% 수준에서 통계적으로 유의하였다. Bonferroni 방법을 이용한 사후검증 결과 시점간에는 모두 평균차이가 통계적으로 유의하였으며, 시점별로 나누어 볼 때 시점3과 시점4에서만 성취수준별 평균차이가 통계적으로 유의했다.

〈표 3〉 시점과 성취수준에 의한 평균차이 검증

분산원	df	SS	MS	F	사후검증
집단간					
성취수준	2	3201.36	1600.68	49.67**	미달 < 보통 < 우수
오차	122	3931.86	32.23		
집단내					
시점	3	3749.76	1249.92	80.83**	시점1 < 시점2 < 시점3, 시점4
시점×성취수준	6	1551.66	258.61	16.72**	시점1, 2 : 미달 < 우수 시점3, 4 : 미달 < 보통 < 우수
오차	366	5659.88	15.46		

* $p<.05$, ** $p<.01$

다. 수학 성취도 점수의 분산변화

수학 성취도 점수의 시점별 분산을 점수산출 방법에 따라 제시한 결과는 <표 4>와 같다. 시점 2에서 시점1보다 분산이 커지고 시점3에서는 다시 시점2보다는 분산이 작아지며, 시점4에서는 분산이 다시 커졌다. 시점1 대비 시점4에 분산의 증가 비율은 17.82%이다.

<표 4> 수학 성취도 점수의 시점별 분산

시점	시점1	시점2	시점3	시점4	시점1에 대한 시점4의 분산 비율(%)
분산	27.01	29.64	26.28	31.83	17.82

시점별로 관찰된 분산 차이가 통계적으로 유의한지에 대한 분산차이 t 검증 결과는 <표 5>와 같다. 시점3과 시점4에서는 분산 차가 통계적으로 유의미했으나 시점1과 시점2간의 분산은 통계적으로 차이가 없었다. 또한 시점2와 시점3간의 분산도 통계적으로 유의미한 차이가 없었다. 그런데 제 1종 오류를 통제하기 위해 Bonferroni 방법으로 유의도 수준을 낮춘다면, 시점3과 시점4간 차이도 통계적으로 유의한 것은 아니었다.

<표 5> 수학 성취도 점수의 시점 간 분산 차이 검증

통계치	시점1-시점2	시점2-시점3	시점3-시점4	시점4-시점1
t	0.7491	0.7454	2.0206*	1.3259

* $p < .05$

2. 학생들의 연중 수학과 학습내용 영역별 성취도 변화

가. 학습내용 영역별 성취도 점수의 평균 변화

초등학교 5학년 수학과 학습내용 영역별 성취도 점수가 평균적으로 어떻게 변화하는지 분석한 결과는 <표 6>에 제시되어 있다. 수학과 전 영역의 성취도 점수는 시점이 지남에 따라 평균 변화량이 대체로 증가하였다. 수와 연산 영역과 규칙성과 함수 영역의 경우, 시점3이 가장 높은 평균 점수인 4.36, 3.76을 나타내었고, 시점4에서는 이보다 조금 낮은 4.09와 3.34로 나타났다.

시점1에서 가장 높은 평균점수를 보인 영역은 수와 연산 영역으로 평균이 3.11이었으며, 가장 낮은 평균점수를 보인 영역은 도형 영역으로 평균이 .71이었다. 시점4에서도 가장 낮은 성취를 보인 영역이 도형 영역으로 평균이 2.72였다.

〈표 6〉 수학과 학습내용 영역별 성취도 점수의 평균 변화

영역	시점1(a)	시점2(b)	시점3(c)	시점4(d)	F	사후검증
수와 연산	3.1100	3.6636	4.3577	4.0901	29.12**	a<b<c,d
측정	1.7200	2.6836	3.3326	3.5959	52.12**	a<b<c,d
확률과 통계	1.1400	2.2531	2.7115	3.2566	71.70**	a<b<c<d
문자와 식	1.3400	2.1227	2.3582	2.7895	31.263**	a<b,c,d b<d, c<d
규칙성과 함수	2.5400	3.6995	3.7605	3.3416	25.77**	a<b,c,d, c>d
도형	0.7100	1.8201	2.2160	2.7230	85.06**	a<b<c<d

* p<.05, ** p<.01

나. 성취수준 집단별 학습내용 영역의 성취도 변화

최종 성취수준에 따라 우수, 보통, 미달 집단으로 나누어서 수학과 학습내용 영역별 성취도 점수 평균이 〈표 7〉부터 〈표 9〉까지 제시되어 있다.

〈표 7〉의 우수 집단은 모든 학습내용 영역에서 성적이 평균적으로 증가했다. 도형 영역의 경우 시점1에서는 .92였으나, 시점4에서는 4.39로 점수가 크게 증가했다. 그러나 규칙성과 함수 영역의 경우에는 시점3까지는 4.49까지 증가했으나, 시점4에서는 3.91로 감소하여 시점1의 3.75보다 조금 높았다. 규칙성과 함수를 제외한 모든 영역에서 평균변화가 통계적으로 유의했다.

〈표 7〉 '우수' 집단의 수학과 학습내용 영역별 성취도

영역	시점1(a)	시점2(b)	시점3(c)	시점4(d)	F	사후검증
수와 연산	3.7500	4.0061	4.8821	4.9571	6.26**	-
측정	2.0000	3.4028	4.4090	4.9193	21.21**	a<b<c<d b<d,
확률과 통계	1.5000	2.3489	3.9834	4.6904	24.07**	a<c,d b<c,d
문자와 식	2.1700	3.0320	4.0739	4.8986	12.26**	a<c,d b<d, c<d
규칙성과 함수	3.7500	4.2506	4.4895	3.9142	1.31	-
도형	0.9200	2.8931	3.1758	4.3947	22.87**	a<b,c,d b<d

* p<.05, ** p<.01

이러한 현상은 <표 8>의 최종 성취수준이 보통인 집단 학생들의 수학과 학습내용 영역별 성취도에서도 나타났다. <표 8>에서 보면, 규칙성과 함수 영역의 성취도 점수가 시점3에서 3.96이고 시점4에서는 3.55로 낮아졌다.

<표 8> '보통' 집단의 수학과 학습내용 영역별 성취도

영역	시점1(a)	시점2(b)	시점3(c)	시점4(d)	F	사후검증
수와 연산	3.0500	3.7979	4.5489	4.4930	39.57**	a<b<c,d
측정	1.6600	2.5410	3.6447	4.0874	70.00**	a<b<c<d
확률과 통계	1.1400	2.2724	2.7796	3.6756	75.22**	a<b,c,d b<d,c<d
문자와 식	1.3200	2.0262	2.4112	3.0882	32.51**	a<b,c,d b<d,c<d
규칙성과 함수	2.5200	3.6354	3.9677	3.5551	21.87**	a<b,c,d c>d
도형	0.7500	1.6649	2.2739	2.9629	78.25**	a<b<c<d

* $p<.05$, ** $p<.01$

최종 성취수준이 미달인 학생들의 수학과 학습내용 영역별 성취도 점수의 평균적인 변화는 <표 9>에서와 같이 수와연산 영역을 제외한 다섯 가지의 내용 영역에서 비슷한 평균 변화의 패턴을 보였다. 시점2에서 평균점수가 가장 높고 시점3과 시점4에서는 다시 평균점수가 낮아졌다. 수와 연산 영역의 경우에는 시점3에서 3.72로 가장 높은 평균점수를 보였으며, 시점4에서 2.84로 시점1보다도 더 낮아진 평균점수를 보였다. 시점1보다 시점4가 낮은 성취를 보인 영역은 수와 연산 영역이었으나 시점4에서 가장 높은 성취를 보인 학습내용 영역이 또한 수와 연산 영역이었다.

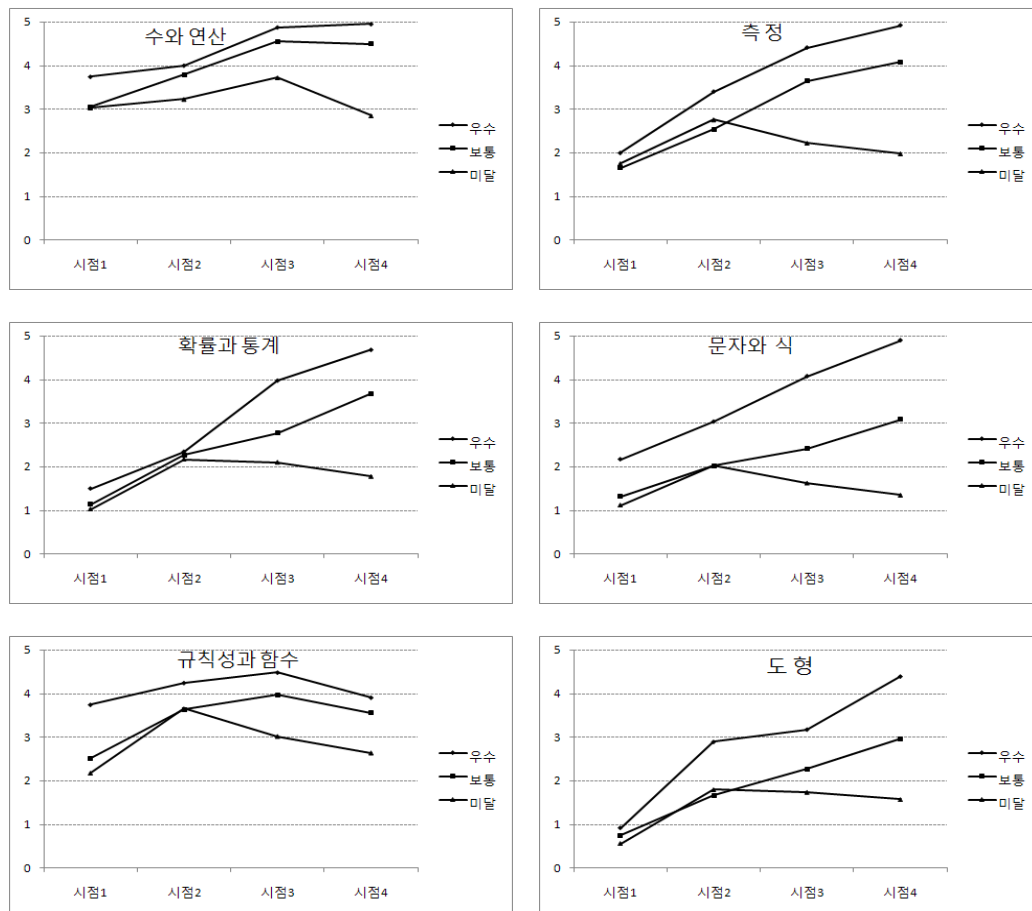
[그림 4]에서 나타났듯이 최종 성취수준에 의해 분류된 집단 간 비교에서 학생들의 능력집단에 따라 학습내용 영역별 성취도 변화 패턴이 다소 다르게 나타났다. 미달 집단의 학습내용 영역별 성취도는 거의 변화가 없거나 오히려 학기 초보다 낮아진 것으로 분석되었다. 우수 집단은 보통 집단이나 미달 집단보다 모든 학습내용 영역에서 시점이 지남에 따라 성취도 점수가 많이 향상되었다.

특히 다른 영역에 비해 문자와 식 영역에서는 [그림 4]에서와 같이 우수 집단 학생들의 시점 4 성취도 평균은 4.89로 보통 집단의 3.08, 미달 집단의 1.35보다 다른 영역에 비해 큰 차이를 보이면서 향상되었다. 반면에 규칙성과 함수 영역에서는 시점4 성취도 평균이 우수 집단은 3.91, 보통 집단은 3.55, 미달 집단은 2.64로 평균 차이가 가장 적었다.

〈표 9〉 ‘미달’ 집단의 수학과 학습내용 영역별 성취도

영역	시점1(a)	시점2(b)	시점3(c)	시점4(d)	F	사후검증
수와 연산	3.0300	3.2307	3.7285	2.8479	2.65	-
측정	1.7600	2.7611	2.2275	1.9869	3.58*	a<b
확률과 통계	1.0300	2.1744	2.1046	1.7769	6.45**	a<b,c
문자와 식	1.1200	2.0261	1.6296	1.3511	3.95*	a<b
규칙성과 함수	2.1800	3.6540	3.0217	2.6433	7.25**	a<b, b>d
도형	0.5600	1.8019	1.7428	1.5755	14.44**	a<b, c, d

* $p<.05$, ** $p<.01$



〔그림 4〕 최종 성취수준 집단의 수학과 학습내용 영역별 성취도 변화

학습내용 영역에서 능력집단별 성장 정도가 다른 것은 교과와 진도와도 관련이 있어 보인다. [그림 5]와 같이 동등화된 네 가지의 검사형은 초등학교 5학년 수학과 학교교육과정 진도계획에 의하여 교수 학습활동이 이루어지는 상황에서 시점3까지는 3.5개월의 간격으로 시행하였으며, 시점4는 다음 학년도의 시작에 시행하였다.

위에서 기술한 바와 같이 규칙성과 함수 영역은 최종 성취수준별 집단으로 나누어 성취도 변화를 분석한 결과 성취수준 집단 간 차이가 시점4에서 가장 작은 것으로 분석되었다. 이는 규칙성과 함수 영역이 검사 시행 시점1 이전에 이미 학습이 되고 그 이후에는 한 번도 학습이 되지 않은 것과 관련이 있어 보인다. 즉, 규칙성과 함수 영역은 모든 성취수준 집단에서 시점3과 시점4에서 성취도 평균이 하락하는 모습을 나타내었는데 이는 학습이 학기 초에 한번만 이루어기 때문에 학습내용에 대한 망각이 작용한 것으로 해석된다. 성취수준 집단 간 차이가 비교적 크게 나타난 문자와 식 영역은 시점2 직전과 시점4 직전에 학습이 이루어짐을 알 수 있다.

월 영역	3	4	5	6	7	9	10	11	12	2	3
수와연산											
규칙성과함수											
도형											
측정											
문자와 식											
확률과 통계											
<div>가형시행</div> <div>나형시행</div> <div>다형시행</div> <div>라형시행</div>											

(그림 5) 학습내용 영역별 연간 지도 계획과 검사시행 시기

수와 연산 영역은 시점3까지는 비교적 전 기간에 걸쳐 꾸준히 학습이 이루어지도록 계획이 되어 있다. 이는 [그림 4]의 그래프에서 나타났듯이 시점3까지는 대체로 꾸준히 향상되다가 시점4에서는 우수 집단은 조금 향상되고, 보통, 미달 집단은 오히려 점수가 낮아지는 결과와 관련이 있다. 확률과 통계 영역에서 우수 집단의 경우 시점2와 시점3간의 기울기가 급격한 것을 볼 수 있었다. 이는 시점3 이전에 확률과 통계를 학습한 것과 관련이 있기 때문이다.

위의 내용을 종합해 볼 때, 학습능력이 보통 이상인 집단의 경우 교육과정 진도 운영계획과 관련하여 학습내용 영역을 학교에서 배우고 검사를 시행한 경우 성취도 점수가 향상됨을 알 수 있었다. 반면 학습능력이 미달인 집단의 경우는 학교 교육과정 진도와 큰 관련이 없이 대부분의 영역에서 성취도 점수가 향상되지 못했다. 전체적인 평균변화가 시점이 뒤로 갈수록 줄어들고, 효과크기도 줄어드는 것은 학생들이 시험을 치르게 되는 내용 영역이 시점이 뒤로 갈수록 대부분

학습된 상태이기 때문이었다. 즉, 시점3과 시점4 간에는 효과크기가 가장 작았는데, 이는 시점3에서 학생들은 5학년 교육과정의 대부분을 학습했기 때문이었다는 것을 알 수 있었다.

V. 결론 및 논의

본 연구의 목적은 초등학교 5학년 학생들의 연중 수학성취도는 어떻게 변화하며, 학습내용 영역별 변화는 어떠한지를 알아보는 데 있었다. 다음에서는 연구문제별로 결론을 제시하고, 이에 대한 논의를 기술한다.

첫 번째 연구문제와 관련한 주요 결론은 다음과 같다. 학생들의 연중 수학 성취도 점수의 평균적인 변화는 초기에는 평균적인 변화가 컸으나 점차 성적증가의 크기가 완만했다. 최종 성적에 따라 분류된 성취수준(우수, 보통, 미달) 집단별로 살펴보았을 때 우수, 보통인 집단은 시점이 지남에 따라 평균이 증가하였고, 미달집단은 시점2까지는 다소 증가하다가 이후에는 성적향상이 이루어지지 않고 오히려 점차 감소하였다. 또한, 수학 성취도 점수의 분산은 시점1 대비 시점4에서 증가해서 학생들 간 격차가 오히려 학기말에 커지는 경향이 있었다.

두 번째 연구문제와 관련한 주요 결론은 다음과 같다. 수학과 전 영역의 성취도 점수는 시점이 지남에 따라 평균 변화량이 대체로 증가했다. 평균 변화량의 패턴은 영역에 따라 공통점도 있고 차이점도 있었다. 수와 연산 영역과 규칙성과 함수 영역의 경우, 시점3에 평균점수가 가장 높았고, 시점4에서는 이보다 조금 낮아졌다. 도형 영역은 모든 시점에서 다른 영역에 비해 평균점수가 가장 낮았다. 확률과 통계, 문자와 식 영역은 시점이 지남에 따라 평균이 올라갔다. 또한 학습능력이 우수와 보통인 학생집단의 경우 학습내용 영역에서 성취도 점수가 향상되었지만, 학습능력이 미달인 학생집단의 경우는 대부분의 영역에서 성취도 점수가 향상되지 못했다. 세 집단 간 차이를 비교해 보면, 집단 간 차이가 가장 큰 영역은 문자와 식이며, 집단 간 차이가 가장 적은 영역은 규칙성과 함수였다.

본 연구의 결과와 관련하여 논의 및 시사점을 몇 가지 제시하면 다음과 같다. 첫째, 수학 성취도 점수는 시점1과 시점2 사이의 평균 변화의 기울기가 가장 컸고, 이후 평균 변화의 기울기가 완만해졌다. 외국의 연구 결과에 의하면 IRT와 써스톤 척도를 이용한 학년 간 연구에서는 학년이 올라갈수록 학생들 성적의 평균 증가가 완만해진다고 밝혀졌다(Andrews, 1995; Bock, 1983; Seltzer, Frank, & Bryk, 1994; Williams, Pommerich, & Thissen, 1998; Yen, 1985, 1986). 반면, Becker & Forsyth(1992)는 고부담 시험을 검토한 결과 성적평균이 완만히 증가한다는 증거가 없다고 보고했다. 다만, 예외적으로 학령발달척도(developmental scale)는 임의적으로 학년 간 성장의 크기를 상수가 되도록 강제로 고정

(Kolen, 2006, p. 178)하기 때문에 학년 간 평균크기가 일정하다고 하였다. 본 연구는 비록 학년 내 성장 결과를 보여주고 있지만 학생들 성적의 평균 증가가 완만했다. 따라서 외국의 선행연구와 일정부분 비슷한 결과를 산출했다. 그렇지만 하나의 중요한 차이점이 있다. 본 연구에서 사용한 검사지는 한 개 학년에 맞춰서 만들어진 것이다. 따라서 학년말이 되면 학생들이 교육내용을 완료함에 따라 동 학년 내에서는 성적향상이 제한받을 수밖에 없다. 물론 다음 학년으로 진급했을 때는 학생들이 새로운 내용을 학습하므로 성장의 크기가 다시 커질 수 있을 것이다. 그렇지만 한 학년 내에서는 천정효과(ceiling effect)로 인해 성적평균의 증가는 시점이 지남에 따라 둔화된 것으로 보인다. 학년단위로 교육과정이 운영되는 학교현장을 감안할 때 한 개 학년 내에서는 초기에는 성적변화가 크다가 학기말에 가까워지면서 성적변화의 크기가 줄어드는 현상은 다른 학년에서도 반복될 가능성이 있을 것으로 예측된다.

둘째, 본 연구에서 발견된 우수, 보통인 집단의 변화패턴은 교육현장에서, 학교교육과정이 정상적으로 운영되었을 때 기대할 수 있는 변화 패턴과 유사하였다. 그러나 미달 집단의 변화패턴은 교육 현장에 중요한 시사점을 제공하고 있다. 일반적으로 초등학교에서 중간고사와 기말고사를 실시할 때 시험범위는 매우 제한적이며, 검사 간 난이도가 조정되지 않았다. 따라서 중간고사나 기말고사에서 성취도가 낮게 나올 경우, 해당 학년의 전체 교육과정 중 어디에서부터 학생들이 교육내용을 이해하지 못하고 있는지를 알기 어렵다. 그런데 이 연구 결과에 따르면, 초등학교 5학년에서 성취수준이 미달로 판정된 학생들은 1년의 어떤 시점에서부터 학업성취도가 향상되지 않는지를 보여주었다. 즉, 1학기말까지는 성취수준이 미달로 판정된 학생들의 학업성취도가 다소 향상되었으나 그 이후로는 학업성취도가 더 이상 향상되지 않았음을 보여줬다. 따라서 성취수준이 미달로 판정된 학생들의 학업성취도가 지속적으로 향상하기 위해서는 이들 학생 수준에 적합한 시험, 예를 들어 기초학력 향상도 평가(반재천 등, 2010)를 지속적으로 시행하여 학생들의 수준을 정확히 진단하고 피드백을 제공하면서 학생을 지도해야 함을 시사하고 있다.

셋째, 본 연구결과에 따르면, 수학 성취도 점수의 분산은 시점2에서는 시점1보다 커지고 시점3에서는 다시 시점2보다 분산이 작아지고 다시 시점4에서는 시점2에서보다 분산이 커졌다. IRT를 이용한 연구들 중 집단 내 분산의 크기가 학년이 증가함에 따라 오히려 줄어든다는 일부 연구가 있다(Andrews, 1995; Hoover, 1984; Yen, 1986). 그러나 Hoover(1984)는 집단 내 분산이 줄어드는 척도는 문제가 있다고 지적하였다. 그 이유는 학년이 증가함에 따라 성적이 낮은 학생 수가 성적이 높은 학생 수보다 점차 빨리 증가하기 때문에 집단 내 분산은 점차 커져야 한다고 주장하였다. 시뮬레이션을 이용한 연구에서 IRT 방법을 적용했을 때 학년이 증가함에 따라 분산이 감소하는 것은 검사가 다차원적(Yen, 1985)이거나 검사들의 측정오차가 다르기 때문(Camilli, 1988)이라고 하였다. 그 외 많은 연구에서 IRT를 적용했을 때 집단 내 분산이 줄어드는 것은 아니라고 밝혔다(Becker & Forsyth, 1992; Bock, 1983; Camilli, Yamamoto, & Wang, 1993; Seltzer, Frank, & Bryk, 1994; Williams, Pommerich,

& Thissen, 1998; Yen & Burket, 1997). 본 연구는 비록 학년 내 연구 상황이지만 전체적으로는 IRT를 적용했을 때 집단 내 분산이 줄어들지 않는다는 연구결과와 유사하다. 오히려 본 연구결과 집단 내 분산이 학년 내에서는 커졌음을 알 수 있다. 다만, 시점3에서 분산이 감소한 것은 검사 시행시기와 관련이 있는 것으로 해석된다. 우리나라 교육과정은 여름방학과 겨울방학이 약 40일이고 2학기제로 운영된다. 따라서 시점1의 검사 시행 시기는 3월로 학생들이 배운 내용이 많지 않기 때문에 대부분의 점수가 매우 낮으며, 점수 분산이 작았다. 시점2의 검사 시행 시기는 7월 중순으로 1학기 교육과정을 모두 학습한 후였다. 따라서 학생들의 능력차이가 3월 말보다는 커져서 시점2의 분산이 시점1의 분산보다 커졌다. 그런데 검사 시행 시기가 12월 중순인 시점3에서 분산이 시점2보다 작아진 것은 정확히 설명하기는 어렵다. 다만, 일선 학교에서는 중하위 정도의 능력을 가진 학생들을 대상으로 여름방학을 이용하여 보충학습지도를 적극적으로 실시하게 된다. 따라서 이러한 학생들의 학습능력이 조금 향상되면서 높은 점수를 받은 학생들의 점수와 격차가 조금은 줄어서, 시점3에서 분산이 줄어든 것으로 보인다. 시점4의 검사 시행 시기는 정확히 시점1의 검사 시행 시점으로부터 1년이 되는 다음 학년도의 3월 말이었다. 학생들은 겨울방학을 지냈고, 진급을 하였기 때문에 학습내용에 대한 기억과 망각이 학생들 개개인마다 다르게 작용하여 점수의 격차가 커진 것으로 보인다.

넷째, 학습내용 영역별로 분석한 내용을 종합해 보면, 최종 성취기준이 보통 이상인 학생집단의 경우 학교교육과정 진도 운영계획과 관련하여 학습내용 영역을 학교에서 배우고 나서 검사를 시행한 시점에서 성취도 점수가 향상되는 것을 볼 수 있었다. 그러나 최종성취 기준이 미달인 학생 집단의 경우에는 교육과정 진도와 상관없이 대부분의 영역에서 성취도 점수가 향상되지 못했다. 특히 성취수준 집단 간 차이가 가장 큰 문자와 식 영역은 기본적으로 문장 이해력과 문제 해결 계획 수립을 할 수 있는 능력이 있어야 문제를 해결할 수 있다. 따라서 미달집단 학생들을 지도할 때 접근이 쉬운 단순 계산 훈련만이 아니라 수학적 논리력과 문제해결력을 함께 키울 수 있는 통합적인 지도를 해야 한다.

또한 정책적 관점에서 보면, 지금까지 학습부진학생 지도를 위해 많은 예산과 지원을 했음에도 불구하고 이와 같이 학습능력이 전혀 향상되지 않는다고 할 때는 분명 무엇인가 제대로 학습부진학생에 대한 지원이 이루어지지 못하고 있음을 의미한다. 물론 연구대상이 한 학교의 5학급에 한정되어 있지만, 본 연구대상 학교는 교육복지투자 대상학교로 다른 학교보다 학습부진학생에 대한 지원이 많았으며, 학교에서도 실제로 방과 후나 방학 등을 이용하여 다양한 학습 지원 프로그램을 진행하였다. 그뿐만 아니라 상담소와 연계하여 정서적인 지원도 많았던 학교이다. 그럼에도 불구하고 성취기준이 미달로 분류된 학생들은 학교에서 어떤 방식으로 지도를 해야 할 것인지 다시 검토해 볼 필요가 있다.

본 연구에서는 초등학교 5학년 수학과목을 대상으로 연구를 실시하였다. 추후연구로 다른 교과 및 학교 급에서의 학생들의 성장 정도에 대해 더욱 탐구해 볼 필요가 있다.

참 고 문 헌

- 권재술, 최병순, 권치순, 양일호, 이경호, 김지나(1999). 초중고 학생들의 과학 지식 성취 수준 추이 분석을 위한 종단적 연구. **한국과학교육학회지**, 19(2), 185-193.
- 김선, 김성훈, 반재천(2009). 검사동등화를 이용한 북마크 방법과 앵고프 방법의 비교. **교육평가연구**, 22(4), 1035-1054.
- 김선옥(2005). 유아의 읽기에 영향을 미치는 변인 연구: 단기 종단적 접근. **유아교육연구**, 25(1), 129-146.
- 김성훈(2000). 능력 변화과정 추정을 위한 시험자료의 동등화 방법 연구. **교육평가연구**, 13(1), 101-125.
- 김성훈, 강상진, 김성호(1999). 다층모형을 적용한 능력 변화과정의 측정. **교육평가연구**, 12(2), 169-193.
- 류한구, 김양분, 현주, 김일혁, 강상진, 김현철, 박성호, 민병철(2005). **한국교육종단연구 2005(I) -예비조사 보고서-**(연구보고 RR 2005-18). 서울: 한국교육개발원.
- 박정(2008). 교육 맥락적 변수의 능력집단별 학업성취도의 영향력 분석. **교육평가연구**, 21(3), 23-41.
- 반재천, 김선, 정혜선, 오경운, 고명숙, 권오정, 김유진(2010). **2010 기초학력 향상도 평가: 스텝업 (Step-Up) 개발**(대전교육 2011-31). 대전: 충남대학교 교육연구소, 대전광역시교육청.
- 송미영(2010). 학교 교육개선을 위한 학생의 학업성취수준 결정요인 분석. 한국교육과정평가원, 한국교육평가학회(편), **2009년 국가수준 학업성취도 평가 전수 결과분석: 학력 차이 요인 탐색과 대응방안**(pp. 71-98). 서울: 한국교육과정평가원.
- 이성원(2004). 교양영어에 대한 학생들의 만족도와 영어 학업성취도의 상관관계에 대한 종단 연구. **영어교육연구**, 16(1), 203-222.
- 임인재, 김신영, 박현정(2003). **교육·심리·사회 연구를 위한 통계방법**. 서울: 학연사.
- 임현정(2006). 수능모의고사를 이용한 학업성취도의 종단적 분석 모형 탐색. **교육평가연구**, 19(2), 281-300.
- 정은영, 최인봉, 김희경, 김소영, 유진은, 남민우, 김도남, 김혜숙, 박가나, 이봉주, 권점례, 최원호, 이인호, 조보경, 송민영(2009). **국가수준 학업성취도 평가 결과 추이: 2003년~2008년 초등학교 6학년**(RRE 2009-8-1). 서울: 한국교육과정평가원.
- 조지민, 김명화, 최인봉, 송미영, 김수진, 남민우, 박종훈, 박은아, 김민정, 도종훈, 고정화, 정은영, 최원호, 김미경(2007). **국가수준 학업성취도 평가 연구: 2003~2006년 변화 추이**(RRE 2007-1-1). 서울: 한국교육과정평가원.

- Andrews, K. M. (1995). *The effects of scaling design and scaling method on the primary score scale associated with a multi-level achievement test*. Unpublished doctoral dissertation, The University of Iowa.
- Becker, D. F., & Forsyth, R. A. (1992). An empirical investigation of Thurstone and IRT methods of scaling achievement tests. *Journal of Educational Measurement*, 29, 341-354.
- Bock, R. D. (1983). The mental growth curve reexamined. In D. J. Weiss (Ed.), *New horizons in testing* (pp. 205-209). New York: Academic Press.
- Cammili, G. (1988). Scale shrinkage and the estimation of latent distribution parameters. *Journal of Educational Statistics*, 13, 227-241.
- Cammili, G., Yamamoto, K., & Wang, M.-M. (1993). Scale shrinkage in vertical equating. *Applied Psychological Measurement*, 17, 379-388.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Hoover, H. D. (1984). The most appropriate scores for measuring educational development in the elementary schools: GE's. *Educational Measurement: Issues Practice*, 3(4), 8-14.
- Kim, S. (2008). EQUIRT: *A computer program for test equating with unidimensional IRT models*. Author.
- Kolen, M. J. (2006). Scaling and norming. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 155-186). Wesport, CT: American Council on Education and Praeger.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.
- Omar, M. H. (1998, April). Item parameter invariance assumption and its implications on vertical scaling of multilevel achievement test data. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Seltezr, M. H., Frank, K. A., & Bryk, A. S. (1994). The metric matters: The sensitivities of conclusions about growth in student achievement to choice of metric. *Educational Evaluation & Policy Analysis*, 16(1), 41-49.
- Williams, V. S. L., Pommerich, M., & Thissen, D. (1998). A comparison of developmental scales based on Thurstone methods and item response theory. *Journal of Educational Measurement*, 35, 93-107.
- Yen, W. M., & Burket, G. R. (1997). Comparison of item response theory and Thurstone methods of vertical scaling. *Journal of Educational Measurement*, 34, 293-313.

- Yen, W. M. (1985). Increasing item complexity: A possible cause of scale shrinkage for unidimensional item response theory. *Psychometrika*, 50, 399-410.
- Yen, W. M. (1986). The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement*, 23, 299-325.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). *BLOG-MG 3.0* [computer program]. Lincolnwood, IL: Scientific Software International.

· 논문접수 : 2011-09-01/ 수정본접수 : 2011-10-10/ 게재승인 : 2011-10-25

ABSTRACT

Estimating Changes of Elementary Students' Achievement in Mathematics During the School Year

Sun Kim(Teacher, Daejeon Seokgyo Elementary School)

The purposes of this study were twofold. First, this study was to examine the extent to which elementary students' mathematic achievement test scores were changed during one school year. Second, this study was to examine whether student's test scores across the sub-areas of the mathematics achievement test were differentially changed.

To assess students' achievement changes, four mathematic test forms (Form A, Form B, Form C, and Form D) for the 5th graders were developed and equated with the IRT true-score equating method. Using the Bookmark and the modified Angoff standard setting methods, two cut scores of the two methods were set on Form A to divide students into Excellent, Average, and Basic groups. Using the four equated forms, 150 students of the 5th grade in D city in Korea were tested four times during one school year.

The main results from this study were as follows. First, the means of the test scores tended to increase during the school year but the increase rate appeared to decrease. Second, the Excellent and Average groups showed meaningful increase in means over the school year but the scores of the Basic group increased in mean by time 2 of testing, after then they rather decreased. Third, the variances of student's test scores were larger at the end of the school year than in the beginning and the variance changes during the school year. Fourth, the means of sub-area scores generally increased over the school year. Finally, the means of the Excellent and Average groups on the sub-areas increased, whereas the means of the Basic group did not increase in most sub-areas.

Key Words : *growth, achievement test, equating, sub-area scores*