

기초학력 진단평가의 성취수준 설정에 관한 연구: 초등학교 3학년 ‘기초수학’ 사례를 중심으로¹⁾

김희경²⁾ · 조성민³⁾

교육격차를 완화하여 국가 경쟁력을 높이고 평생학습 사회에서 개인의 장래 성장과 발전의 기틀을 마련하기 위하여 모든 학생들의 기초학력을 보장하려는 움직임이 다양하게 전개되고 있다. 본 논문에서는 학생들의 기초학력을 결정짓는 성취수준 설정의 의미와 방법을 알아보고, 2013년 초등학교 3학년을 대상으로 실시되는 기초학력 진단평가의 ‘기초수학’ 과목에 대한 성취수준과 분할점수 설정 방법을 살펴보고자 한다. 학교현장에서 활용할 수 있는 성취수준 설정 방법과 적용 사례의 제시는 준거참조평가에서 성취수준 설정 시 고려해야 하는 것에 대한 시사점을 제공할 수 있으리라 기대된다.

주요용어: 성취수준 설정, 분할점수, 초3 기초학력 진단평가, Ebel방법

I. 서론

세계 여러 나라들은 각국의 교육 책무성을 점검하고 국가 경쟁력을 강화하기 위하여 국가 수준에서 학생들의 기초학력을 보장하기 위한 노력을 다각도로 기울이고 있다. 21세기 지식기반 정보화 사회를 살아가기 위해 요구되는 지식의 내용과 수준에 대한 논의와 더불어, 평생에 걸쳐 이를 배우고 익힐 수 있는 기틀을 마련하는 일이야말로 현대사회를 살아가기 위해 필수적이라고 할 수 있기 때문이다. 따라서 많은 나라에서 최소한의 기초학력이 형성되는 시기라 할 수 있는 초등학교에서의 기초학력을 보장함으로써 국가의 미래 뿐 아니라 개인의 성장과 발전, 행복을 담보하고자 한다.

기초학력은 무엇을, 어떤 수준으로 할 것인가에 따라 다르게 해석되어, 기초학습기능, 기초학습능력 등과 혼용된 채 학자들 사이에서 명확한 합의에 이르지 못하고 있다(양명희, 2006). 이는 ‘기초’에 대한 개념 정의 방식이나 ‘학력’이 규정하는 내용, ‘학력’에 대한 사회의 요구 수준에 따라 의미가 달라지기 때문이다. 예를 들어, 기초학력을 학교교육을 통해 길러야 할 ‘학력’의 밑바탕이 되는 ‘기초’로 파악하여, 그 ‘기초’를 무엇으로 보느냐에 따라 기초학력에 대한

1) 본 논문은 한국교육과정평가원 2012년 수탁과제 ‘온라인 3R's 기초학력 진단 평가도구 개발’ 연구 결과의 일부를 수정·보완한 것임.

2) 한국교육과정평가원(heekyoung@kice.re.kr)

3) 한국교육과정평가원(csminy@kice.re.kr), 교신저자

견해가 구분될 수 있다. 즉, 기초적·도구적 기능을 수행하는 국어와 수학 교과에서 형성되는 ‘언어와 수에 관한 학력’으로 보는 견해, 국어와 수학 교과에 한정하지 않고 각 교과의 토대가 되는 기초적 지식이나 기능으로 보는 견해, 후(後) 교육 단계의 학력 형성에 기초가 되는 전(前) 단계의 학력으로 보는 견해, 국민 교육 제도로서의 의무교육 단계에서 형성되는 최소한의 학력으로 보는 견해, 자립한 국민으로 사회생활을 영위하기 위한 최소한의 필수불가결한 능력으로 보는 견해로 구분된다(김명화 외, 2011). 본 논문에서는 기초학력을 읽기, 쓰기, 셈하기와 같은 능력으로, 어떤 교육을 받는데 기초적으로 필요한 학습 능력으로 정의한 연구 결과(서울대학교교육연구소, 1994)를 바탕으로, 다른 교과를 학습하는 데도 꼭 필요하며 나아가 사회생활을 하는 데 기본적으로 요구되는 능력으로 규정하였다. 이는 ‘기초학력 향상 지원 기본계획(교육과학기술부, 2010)’에서 제시한 ‘기초학력 미달 학생’의 개념을 사용하여 ‘기초학력’을 개념화한 것으로, 학교 학습에 강조점을 두어 기초학습 능력 및 교육과정에서 요구하는 최저 수준의 기본 학습 능력을 의미한다.

기초학력을 중시하는 세계적인 흐름에 따라 우리나라는 국가수준 학업성취도 평가, 초등학교 3학년 기초학력 진단평가(이하 초3 기초학력 진단평가), 교과학력 진단평가 등을 통해 초등학교에서의 기초학력을 평가하고 부족한 부분에 대한 보정 교육을 실시하고 있다. 초등학교는 공교육이 시작되는 시기로, 기초학력을 제대로 형성하지 않으면 학습에 대한 흥미가 감소하고 학습 결손의 누적이 두드러지기 시작하는 시기이다(정혜영, 2010). 특히 초등학교 3학년이 마무리되는 시기까지 기초학력이 형성되지 못한 학생은 이후 학교 학습에 대한 흥미를 잃게 되고 학교에 적응하지 못하여 성공적인 사회생활을 하는 데에도 어려움을 겪을 가능성이 높다(고정화, 2008). 따라서 2001년 교육부의 ‘국가인적자원개발기본계획(2001.12.)’에 따라 2002년부터 초3 기초학력 진단평가가 실시되었으나(이인제 외, 2004), 2010년부터는 시·도교육청이 주관하는 ‘교과학습 진단평가’와 통합되어 시행되고 있다(정혜영, 2010). 현재 초3 기초학력 진단평가는 초등학교 3학년에 진입하는 수준에서 교육과정에 근거한 기초학력의 도달 여부를 판단하고 부족한 부분을 보정하려는 목적에 따라 실시되고 있으며, 성취기준에 대한 학생들의 도달 정도를 기초학력 ‘도달’과 기초학력 ‘미도달’의 두 가지 수준으로 구분하여 그 결과를 제공한다.

일반적으로 성취기준의 도달 여부를 판단하는 성취수준을 설정하는 것은 매우 어려운 일이다. 이를 위해 최소 능력을 규정하거나, 최소 능력을 구성하는 내용 요소들을 설명하기도 하고, 응시 집단 내의 상대적인 비율로서 성취수준을 결정하기도 한다. 또 최소 능력을 바탕으로 하는 방법과 응시 집단 내의 비율로 결정하는 방법을 통합하여 정답률에 따라 성취수준 설정을 위한 비율을 바꾸기도 한다. 이외에 이미 검증된 결과와의 비교를 통해 성취수준을 결정하기도 한다(Ebel, 1972). 그러나 어떤 방법을 활용하는 경우에도 이를 결정하는 과정에는 여러 가지 영향을 미치는 변인이 존재하게 된다(장은아, 2007). 따라서 결정된 성취수준 설정의 타당성을 인정받기 위해서는 설정 과정이 적절했는지, 설정 과정에 참여한 패널의 판단은 일관성이 있었는지 등을 설명할 수 있어야 한다(Kane, 1994). 본 논문에서는 성취수준 설정 방법의 적용 사례로서, 초3 기초학력 진단평가의 기초수학에 대한 성취수준과 분할점수(cut-off scores)를 결정하는 과정을 살펴보고자 한다. 이는 성취수준과 분할점수 설정의 의미와 방법에 대한 이해를 돕고, 향후 학교 수준에서 학생들의 성취수준을 결정할 때 고려해야 하는 것들에 대한 시사점을 제공할 수 있으리라 기대된다.

II. 이론적 배경

1. 성취기준

초3 기초학력 진단평가는 기초학력 수준을 국가수준의 공통된 기준으로 진단하고, 그 결과에 따라 적절한 보정교육을 실시함으로써 모든 학생의 기초학력을 보장하려는 목적을 가진다(김희경 외, 2013). 따라서 초3 기초학력 진단평가는 기초학력의 도달 여부를 교육과정의 교육 목표 및 내용에 근거하여 판단한다는 점에서 학생이 무엇을 얼마만큼 알고 있는가에 관심을 두는 준거참조평가(criterion-referenced assessment)로 실시하는 것이 타당하다.

준거참조평가란 준거(criterion)에 비추어 학생을 판단하는 평가로서, 사전에 결정된 어떤 성취기준 또는 교육목표를 달성하였는지 혹은 달성하지 못하였는지에 초점을 둔다(한국교육평가학회, 2004). 따라서 준거참조평가에서는 학생이 '무엇을' '얼마나' 성취했는가에 관심을 갖게 되고, '무엇을'에 해당하는 성취기준 혹은 교육목표와 '얼마나'에 해당하는 성취수준이 중요한 요소가 된다.

성취기준은 '교수·학습 활동의 실질적인 기준으로서 각 교과목에서 가르치고 배워야 할 내용(지식, 기능, 태도)과 그 내용에 대한 학습을 통해 학생들이 성취해야 할(또는 보여주어야 할) 능력 및 특성을 명료하게 진술한 것'(김정호 외, 1999; 13)으로 2013년 초3 기초학력 진단평가에서는 2013년 초등학교 3학년 학생에게 적용된 2007 개정 수학과 교육과정의 내용 체계를 바탕으로 성취기준을 구체화하였다.

한편 2007 개정 수학과 교육과정에 대한 성취기준과 평가기준 개발 연구(박소영 외, 2007; 오은순 외, 2007)에서는 각 영역에서 학생들이 성취한 정도를 몇 개의 수준(상/중/하)로 나누고 각 수준에서 기대되는 성취 정도를 구체화함으로써 교과별 평가 시 실질적인 기준이 되는 것을 평가기준이라고 정의하였다. 따라서 평가기준은 '성취기준을 어느 정도로 달성했는가'를 기준으로 상/중/하로 분류하며 성취기준을 달성한 수준을 '중', 달성하지 못한 수준을 '하'로 정의한다. '100이하의 자연수'의 교육과정 내용 중 하나인 '10이하의 수 범위에서 두 수로 분해하고, 두 수를 하나의 수로 합성할 수 있다'의 성취기준과 평가기준은 <표 1>과 같다.

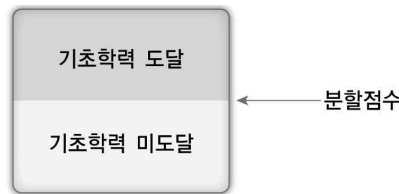
<표 1> 교육과정 내용에 따른 성취기준, 평가기준의 예(오은순 외, 2007)

내용	성취기준	평가기준		
		상	중	하
10이하의 수 범위에서 두 수로 분해하고, 두 수를 하나의 수로 합성할 수 있다.	9이하의 수 범위에서 하나의 수를 두 수로 가르거나 두 수를 하나의 수로 모을 수 있다.	9까지의 수를 다양하게 두 수로 가르거나 모을 수 있다.	9이하의 수를 두 수로 가르거나, 합이 9이하인 두 수를 하나의 수로 모을 수 있다.	9개 이하의 구체물을 두 묶음으로 나누거나 합이 9개 이하인 두 묶음의 구체물을 하나로 모은 것을 보고 수로 나타낼 수 있다.
	10을 두 수로 가르거나 두 수를 하나의 수로 모을 수 있다.	10을 다양하게 두 수로 가르거나 모을 수 있다.	10을 두 수로 가르거나 두 수를 모아 10을 만들 수 있다.	10개의 구체물을 두 묶음으로 나누거나 합이 10개인 두 묶음의 구체물을 하나로 모은 것을 보고 수로 나타낼 수 있다.

‘중’ 수준은 성취기준을 달성한 것으로, ‘하’ 수준은 달성하지 못한 것으로 구분된다는 점에
서, 평가기준은 기초학력의 도달여부를 결정짓는 수준 설정에 참고자료가 될 수 있다. 예를
들어, 9개 이하의 구체물을 두 묶음으로 나누거나 합이 9개 이하인 두 묶음의 구체물을 하나
로 모은 것을 보고 수로 나타낼 수는 있으나, 구체물을 이용하지 않은 채 수를 가르거나 모으
지 못하는 학생은 ‘9이하의 수 범위에서 하나의 수를 두 수로 가르거나 두 수를 하나의 수로
모을 수 있다’는 성취기준에 도달하지 못한 것으로 판단할 수 있다.

2. 성취수준 설정

성취수준 설정(standard-setting)은 학생들의 성취수준을 구분하기 위하여 분할점수(cut-off
scores)를 결정하는 작업을 의미하며(김희경 외, 2013), 초3 기초학력 진단평가의 경우 학생들
의 성취수준을 기초학력 도달과 미도달이라는 두 수준으로 구분하기 때문에 [그림 1]과 같이
하나의 분할점수(합격점)가 필요하다.



[그림 1] 분할점수의 개념(출처: 김희경 외, 2013, p35)

Livingston & Zikey(1982)는 성취수준 설정이란 ‘존재하는 정답’을 찾아내는 것이 아니라,
전문가들이 합의하여 어느 정도의 수준이면 최소기준을 도달했다고 판단할 수 있을지(what
score is enough to be qualified?)를 ‘의사결정’하는 작업이라고 정의하였다. 따라서 성취수준
설정을 위한 최적의 방법은 맥락에 따라 다를 수 있으며, 지금까지 개발된 성취수준 설정 방
법으로는 Nedelsky방법, Angoff방법, Jaeger방법, Ebel방법, Bookmark방법 등이 있다(성태제,
2010). 본 절에서는 현재 가장 많이 활용되고 있는 Angoff방법, Bookmark방법과 2013년 초3
기초학력 진단평가에서 활용된 Ebel방법에 대하여 살펴보고자 한다.

1) Angoff방법

Angoff방법은 1971년 Angoff가 처음 제안한 방법으로 성취수준 설정에 참여한 교사나 교과
내용 전문가가 문항을 분석한 후, ‘도달’에 속하는 사람들 중에서 경계선에 있는 최소능력자의
문항정답률을 예상하여 성취수준을 설정하는 방법이다. Angoff방법은 절차가 간단하고 선다
형이 아닌 서답형 문항에도 활용할 수 있다는 점에서 널리 활용되고 있다(성태제, 2010). 그러
나 Angoff방법은 그대로 사용하는 경우보다 원래 방식을 맥락에 따라 적절하게 조정하여 적
용하는 경우가 많다(김경희 외, 2012). 변형된 Angoff방법(modified Angoff method)에서는 한
번의 평정 결과로 성취수준을 설정하는 것이 아니라, 여러 번의 라운드를 통해 Angoff방법을
반복 적용함으로써 정답률이 과소 혹은 과대 추정되는 것을 예방한다. 이때 각각의 라운드가

끝난 후에는 다른 평가자들의 성취수준 설정에 대한 정보를 제공받고 이를 토대로 성취수준 설정에 참여한 전문가들 사이에 토의를 진행한다. 이와 같이 변형된 Angoff방법은 우리나라의 국가수준 학업성취도 평가, 미국 의사 자격증 시험(The US Medical Licensing Examination) 등에서 성취수준 설정 방법으로 활용되고 있다. 이러한 Angoff방법은 다른 성취수준 설정 방법에 비해 이해하기 쉽다는 장점으로 인해 1990년대 초반 가장 널리 활용되었으나 모든 문항에 대해 예상 정답률을 추정해야 하는 번거로움이 문제점으로 지적되고 있다.

2) Bookmark방법

Bookmark방법은 1996년 NAEP(National Assessment of Educational Progress)에서 성취수준 설정을 위한 방법으로 처음 소개된 후, 미국의 28개 주 이상에서 주단위 성취도 평가의 성취수준 설정 시 사용되고 있다(ACT, 2005). Bookmark방법은 선다형 문항과 서답형 문항으로 구성된 검사에서도 사용할 수 있고, 성취수준 설정에 참여한 교사 혹은 교과 내용 전문가가 평가 전문가의 도움을 통해 비교적 단순하게 수준을 설정할 수 있어 하나의 시험에서 우수학력, 보통학력, 기초학력, 기초학력 미달 등의 여러 수준을 구분하는 분할점수를 설정하는 데 용이하다(장은아, 2007).

Bookmark방법은 평가 전문가가 문항반응이론(Item Response Theory)에 의해 문항난이도를 추정하고 난이도에 따라 문항을 배열하여 문항집(ordered item booklet)을 만든 후 교사 혹은 교과 내용 전문가에게 제공하는 것부터 시작된다. 교사 또는 교과 내용 전문가는 소그룹, 중그룹 등에서 토의한 후 쉬운 문항부터 검토하며 해당 수준의 최소능력자가 응답하지 못할 것이라고 판단되기 시작하는 문항에 북마크(bookmark)를 표시한다. Bookmark방법은 다른 방법에 비해 비교적 간단하고 소요 시간도 적게 들지만, 문항집을 제작해야 하고 문항의 수가 적거나, 사례 수가 적은 경우 문항반응이론에 의거한 문항난이도 추정이 정확하지 못할 수도 있다는 한계가 제기되고 있다(주연희, 2013).

3) Ebel방법

문항의 내용 분석을 통해 성취수준을 설정하는 Ebel방법은 문항의 난이도(difficulty)와 적절성(relevance), 2가지 측면으로 문항을 분류하고 분류된 문항군에 하나의 예상 정답률을 부여한다. 즉, 패널이 문항의 난이도를 기준으로 쉬운 문항, 보통인 문항, 어려운 문항과 문항의 적절성을 기준으로 필수적인 문항, 중요한 문항, 수용할 만한 문항, 문제가 있는 문항으로 구분하여 다음 표를 완성한다.

<표 2> Ebel방법에 의한 분류

적절성 난이도	필수적인 문항	중요한 문항	수용할 만한 문항	문제가 있는 문항
쉬운 문항				
보통인 문항				
어려운 문항				

이와 같은 분류표를 이용하여 문항을 분류하면, 각 셀에는 유사한 문항들이 묶여지고 유사한 문항군에 대하여 예상 정답률을 추정하게 되므로 개별 문항에 대해 정답률을 추정하는 Angoff방법에 비해 간단하면서도 Angoff방법을 적용했을 때의 결과와 상관관계는 매우 높게 나타난다(성태제, 1991).

최근에는 다양한 방법으로 문항 분류의 기준을 설정하는 변형된 Ebel방법(modified Ebel method)이 도입되어 성취수준 설정 방법의 폭을 넓히고 있다.

Ⅲ. 연구 방법

1. 성취수준 설정 참여자

성취수준 설정의 참여자는 평가 내용에 대해 잘 알고, 실제로 평가에 응시할 학생들의 학습 능력과 수준을 잘 파악할 뿐 아니라, 측정 및 평가에 대한 이론적 지식을 겸비한 사람이 적합하다(홍소영, 2011). 따라서 본 연구에서는 초등학교 교사 중 2학년 또는 3학년을 지도한 경험이 있는 교사를 수준설정 참여자로 선정하여 패널을 구성하였다. 초등학교 정교사 자격을 취득하기 위해서는 ‘교육평가’에 관한 교직이론 수업을 이수해야 하고, 초등학교 2학년을 담당하는 경우에는 초등학교 2학년 학생들의 변화, 발달 과정에 대한 이해와 경험적 지식이, 초등학교 3학년을 담당하는 경우에는 초등학교 3학년에 진입한 수준의 학생들에 대한 이해와 경험적 지식이 있으리라 기대하기 때문이다.

<표 3> 수준 설정 참여자의 교직경력

구분 \ 경력	패널 1	패널 2	패널 3	패널 4	패널 5	패널 6
교직 경력	20	8	13	15	8	3
초등학교 2학년 담임 경력	4	1	0	1	0	1
초등학교 3학년 담임 경력	1	1	3	1	2	0

2. 연구 자료

본 연구에서는 성취수준 설정을 위하여 전국의 15개 학교 65개 학급을 표집하여 2013년 초 3 기초학력 진단평가의 ‘기초수학’ 검사를 시행하였다. 표집된 학교의 학생 중 결시 등을 제외하고 ‘기초수학’ 검사를 시행한 학생 수는 1,483명이었다. 성취수준 설정을 위해 시행된 2013년 초3 기초학력 진단평가 기초수학의 유형 별 문항 수는 <표 4>와 같다.

<표 4> 기초학력 진단평가 기초수학 평가틀

내용영역	문항 유형		
	선다형	서답형	합계
수와 연산	12	1	13
도형	4		4
측정	6		6
확률과 통계		1	1
규칙성과 문제해결	1	1	2
문항수	23	3	

서답형은 각각 2문항씩 하위 문항을 포함한다. 서답형의 하위 문항에는 선다형에 이어 문항 번호를 설정하여 하위 문항별로 서로 다른 문항 번호를 부여하였다. 따라서 선다형 23문항과 서답형 6문항, 총 29문항이 성취수준 설정에 활용되었다.

3. 연구 절차

예상 정답률을 작성하기에 앞서 기초학력에 도달한 학생의 특성과 도달하지 못한 학생의 특성을 확인하여 기초학력에 도달한 최소능력자의 특성을 개념화하였다. 기초학력의 개념에 따라 성취수준을 설정하는 경우에도 기초학력의 도달 기대 수준, 즉 최소한의 능력이 어느 정도인가를 밝혀야만 기초학력의 개념에 대하여 충분한 의사소통이 가능하기 때문이다(임천택, 2009). 따라서 패널로 참석한 교사들은 수학과 교육과정을 바탕으로 초등학교 2학년까지 마친 학생에게 요구되는 기초학력의 수준과 최소한의 기초학력을 가진 학생이 나타낼 특성을 구체화하고 성취수준 설정의 근거로 활용하였다.

성취수준과 최소능력에 대하여 논의한 후, 예상 정답률 작성을 위하여 성취수준 설정에 활용될 2013년 초3 기초학력 진단평가 ‘기초수학’의 문항을 검토하였다. 이후 Ebel방법을 변형하여 성취수준을 설정하였다. 본 연구에서는 변형된 Ebel방법을 성취수준 설정 방법으로 채택하였는데, 그 이유는 다음과 같이 두 가지로 요약할 수 있다.

첫째, 전국의 100개 고등학교에서 변형된 Angoff방법과 변형된 Ebel방법을 적용하여 교사들로 하여금 학생들의 성취 결과를 A/B/C/D/E 등급으로 구분하여 성취수준을 설정하도록 한 후, 이 과정에 참여했던 423명의 교사를 대상으로 설문 조사를 하였다. 그 결과, 52%는 Ebel방법을 선호, 40%는 Angoff방법을 선호, 나머지는 뚜렷이 선호하는 방법이 없는 것으로 나타났다(한국교육과정평가원, 2012. 8). 특히 Ebel방법을 선호한다고 응답한 교사들은 난이도와 내용에 따라 문항을 범주화하고 나면 예상 정답률을 추정하기가 쉽고, 문항별로 범주화해서 예상 정답률을 추정하므로 시간적으로도 절약이 되는 것을 그 이유로 제시하였다.

둘째, Ebel방법은 문항을 범주화하여 유사한 문항끼리 묶기 때문에, 같이 묶여진 유사 문항들에 대한 정답률을 한꺼번에 추정할 수 있다. 변형된 Ebel방법에서는 다양한 방법으로 문항을 분류할 수 있는데, 학교현장에서 문항 출제 시 작성하는 이원목적분류표를 활용하는 방법을 고려할 수 있다. 즉, 이원목적분류표에 제시되는 난이도에 따라 쉬움/보통/어려움으로 분류하고, 행동 영역(지식/이해/적용 등) 또는 내용영역으로 문항을 분류하면 유사한 문항군으로

쉽게 묶여진다. 성취평가제의 도입에 따라 학교에서 문항 출제 시 이원목적분류표의 중요성이 강조되므로(박선화 외, 2012) 이원목적분류표를 활용한 Ebel방법은 학교에서 성취수준 설정 시 활용이 용이할 것이라는 판단에 따라 본 연구에서는 변형된 Ebel방법을 성취수준 설정의 방법으로 채택하였다.

본 연구에서는 Ebel방법을 적용하기 위해 문항의 난이도(쉬움/보통/어려움)와 내용영역(수와 연산, 도형, 확률과 통계, 측정, 규칙성과 문제해결)에 따라 문항을 분류하기로 합의하고, 유사 문항군의 정답률을 개별적으로 작성하였다. 이후 실제 표집에서 나타났던 정답률과 조별 정답률, 최댓값, 최솟값 등을 바탕으로 조별로 회의를 진행하여 자신의 정답률 추정값을 수정할지에 대한 여부를 결정하고, 예상 정답률을 독립적으로 다시 작성하였다. 이와 같은 방법으로 3라운드에 걸쳐 예상 정답률을 작성한 후 3라운드에서의 평균값을 산출하여 최종 분할점수를 결정하였다.



[그림 2] 성취수준 설정 절차

IV. 연구 결과

1. 성취수준 및 최소능력 설정

2013년에 초등학교 3학년인 학생에게 적용되는 학교 교육과정은 2007 개정에 따른 수학과 교육과정의 적용을 받고 있다. 따라서 2007 개정 수학과 교육과정에 대한 성취기준과 평가기준 개발 연구(박소영 외, 2007; 오은순 외, 2007)의 결과를 바탕으로 성취수준과 최소능력을 설정하였다. 성취기준을 달성하여 기초학력에 도달했을 경우에는 평가기준의 ‘상’ 또는 ‘중’에 해당하고, 성취기준에 도달하지 못한 경우에는 평가기준의 ‘하’에 해당한다.



[그림 3] 기초학력 도달 여부와 평가기준

따라서 2013년 초3 기초학력 진단평가 기초수학 문항의 성취기준별로 평가기준의 ‘중’/‘하’의

내용을 비교하여 기초학력에 도달한 최소능력의 의미를 구체화하였다. 예를 들어 초등학교 2학년 수학과 교육과정에는 측정 영역의 '시각과 시간' 영역에 '1시간은 60분임을 알고, 시간을 '시간', '분'으로 말할 수 있다'와 '1시간, 1일, 1주일, 1개월, 1년 사이의 관계를 이해한다'는 성취기준이 제시되어 있고, 2007 개정 수학과 교육과정에 대한 성취기준과 평가기준 개발 연구(박소영 외, 2007; 오은순 외, 2007)에서 제시한 평가기준은 <표 5>와 같다.

<표 5> '1분, 1시간, 1일, 1주일, 1개월, 1년 사이의 관계를 이해한다'와 관련된 평가기준

	상	중	하
1시간은 60분임을 알고, 시간을 '시간', '분'으로 말할 수 있다.	1시간은 60분임을 알고, '몇 시간 몇 분'을 '몇 분'으로, '몇 분'을 '몇 시간 몇 분'으로 능숙하게 말할 수 있다.	1시간이 60분임을 알고, '몇 시간 몇 분'을 '몇 분'으로, '몇 분'을 '몇 시간 몇 분'으로 말할 수 있다.	1시간이 60분임을 안다.
1시간, 1일, 1주일, 1개월, 1년 사이의 관계를 이해한다.	1시간, 1일, 1주일, 1개월, 1년 사이의 관계를 이해하고, 시간을 계산할 수 있다.	1시간, 1일, 1주일, 1개월, 1년 사이의 관계를 말할 수 있다.	하루가 24시간, 1주일이 7일, 1개월은 30일(또는 31일), 1년은 12개월임을 안다.

2013년 초3 기초학력 진단평가에서는 이 두 성취기준을 결합하여 '1분, 1시간, 1일, 1주일, 1개월, 1년 사이의 관계를 이해한다'라는 성취기준을 설정하여 문항을 개발하였으므로 그에 대한 '중' 수준의 평가기준은 '1분, 1일, 1주일, 1개월, 1년 사이의 관계를 말할 수 있다'라고 할 수 있다. 따라서 1시간이 60분, 1일이 24시간, 1주일이 7일, 1개월이 30일(또는 31일), 1년이 12개월임을 알지 못하는 학생은 기초학력에 도달하지 못한 것이고, 이 성취기준에 대한 최소능력자는 1분, 1일, 1주일, 1개월, 1년 사이의 관계를 말할 수 있는 학생이라고 구체화하였다.

2. 성취수준 설정 라운드별 결과

성취수준 설정에 참여한 패널은 3명씩 조를 편성하여 조별 협의를 진행하였다. 정답률을 작성한 후에는 다른 조의 문항 분류 내역, 예상 정답률 평균, 중앙값, 최댓값, 최솟값 등의 정보를 제공하고, 조별로 논의를 진행하였다.

1) 성취수준 설정 1라운드

Ebel방법에 의한 성취수준 설정의 시작은 문항을 유사 문항군으로 분류하는 것이다. 본 연구에서는 난이도에 따라 문항을 쉬움/보통/어려움으로, 내용영역에 따라 수와 연산, 도형, 확률과 통계, 측정, 규칙성과 문제해결로 분류하였다. <표 6>을 살펴보면 '수와 연산'에 해당하면서 난이도가 '쉬운' 문항으로 1조는 1, 3, 25번 문항을 분류하였고, 2조는 1, 3, 8, 9, 17, 25번을 분류한 것을 알 수 있다. 패널들은 이렇게 분류된 유사 문항군에 대한 정답률을 추정하는데, 추정오류를 고려하여 0, 5, ..., 90, 95, 100%와 같이 5의 배수의 형식으로 문항의 예상 정답률을 개별 작성하였다.

<표 6> 성취수준 설정을 위한 1라운드 결과

내용영역	난이도	문항번호						예상정답률 (단위: %)				
								패널 1	패널 2	패널 3	1조평균	
								패널 4	패널 5	패널 6	2조평균	
수와 연산	쉬움	1	3	25				80	60	85	75.00	
		1	3	8	9	17	25	80	80	80	80.00	
	보통	8	9	13	17	21	24	75	55	80	70.00	
		6	13	15	16	19	21	70	75	70	71.67	
어려움	6	11	15	16	19		50	40	70	53.33		
	11	24						45	50	50	48.33	
도형	쉬움	5	12						80	75	80	78.33
		12					80	80	75	78.33		
	보통	2	7						60	65	70	65.00
		5	7						70	70	65	68.33
어려움	2						50	65	55	56.67		
측정	쉬움	4	10						95	90	95	93.33
		4	10						80	80	80	80.00
	보통	14	22						80	85	85	83.33
		14	22						75	70	70	71.67
어려움	20	23						70	80	75	75.00	
	20	23						50	50	40	46.67	
확률과 통계	쉬움	28						80	75	70	75.00	
	보통	28	29						75	65	80	73.33
		29					75	70	65	70.00		
어려움												
규칙성과 문제해결	쉬움	26						80	80	70	76.67	
		26					60	65	80	68.33		
	보통	27					70	65	65	66.67		
		18	27						50	60	70	60.00
어려움	18						30	30	30	30.00		

<표 6>에 따르면 1라운드를 진행한 결과, ‘측정’ 영역에서는 1조와 2조의 문항 구분이 일치하였다. 그러나 ‘수와 연산’ 영역은 14문항 중에서 6문항, ‘도형’ 영역은 4문항 중에서 2문항, ‘확률과 통계’ 영역은 2문항 중에서 1문항, ‘규칙성과 문제해결’ 영역은 4문항 중에서 1문항만 난이도에 따른 문항 구분이 일치하였다.

1조와 2조에서 공통적으로 ‘어려움’으로 분류한 문항은 실생활 맥락에서 연산의 개념을 묻거나, 시간을 나타내는 단위 사이의 관계를 묻는 문항이다. 문항의 난이도에 대한 패널의 분류는 초등학교 저학년 학생들은 시간을 분, 또는 분을 시간으로 환산하는데 어려움을 보이고, 학습 상황 보다 생활 맥락에 대한 문항의 정답률이 낮다는 연구 결과(양명희, 2006; 조영미, 2006)와 일치한다.

또한 초등학교 1학년과 2학년의 교육 내용 중 많은 부분을 차지하는 ‘수와 연산’ 영역의 문

항은 주로 '쉬움'과 '보통'에 배치되었는데, 1조의 경우 '확률과 통계', '규칙성과 문제해결' 영역에서는 '쉬움'에 문항을 하나도 분류하지 않았다. 2002년 10월 실시된 초3 기초학력 진단평가 평가를 분석한 결과에서도 초3 학생들은 수의 개념을 잘 이해하고 사칙 연산에 따른 계산은 능숙한 반면, 길이의 어렵이나 실생활에서 단위를 계산하고 전환하는 데는 어렵하다고 제시하였다(양명희, 2006).

한편 '확률과 통계' 영역과 '규칙성과 문제해결' 영역은 교육과정에서 다른 영역에 비해 상대적으로 내용이 빈약하고, 수리력이 기초학력으로 개념화됨에 따라 2012년까지 기초학력 진단평가에서는 제외되었다. 따라서 패널들은 이 두 영역을 상대적으로 어렵다고 인식하였는데, 특히 1조는 '규칙성과 문제해결' 영역에서 '쉬움'에 문항을 분류하지 않았고, 2조는 '어려움'에 해당하는 문항의 예상 정답률을 30%로 제시하였다.

난이도에 따른 문항 분류가 조별 협의를 통하여 결정되었음에도 불구하고 '쉬움/보통/어려움'이라는 난이도에 대한 패널의 인식은 조별로 상당히 다르게 나타났다. 1조에 속하는 패널 2가 제시한 '쉬움'에 해당하는 문항의 예상 정답률은 2조에서 '보통'에 해당하는 문항에 대하여 제시한 예상 정답률의 평균보다 낮게 나타났다.

1라운드 결과, 패널들은 기초학력 도달 여부를 결정하는 성취수준을 전반적으로 높게 설정하였음을 알 수 있다. 고정화(2008)에 따르면, 기초학력 미도달인 학생은 숫자의 중간에 0이 포함되어 있는 경우 숫자로 나타낸 수를 읽는 것에 어려워하고, 도형에 관한 명제의 타당성을 판단하는 내용에서도 어려움을 겪는다. 그러나 본 연구에 참여한 패널은 6명 중 4명이 십의 자리에 0이 포함된 숫자를 읽는 문항에 대하여 80% 이상의 정답률을 예상하였고, 3명이 길이의 단위를 묻는 문항에 대하여 90% 이상의 정답률을 예상하였다. 이는 문항에 대한 응답자들 기초학력에 도달한 최소능력자가 아닌 기초학력에 도달한 일반적(평균적) 학생들로 전제하여 문항 정답률을 예상한 것이라는 추측을 가능하게 한다. 따라서 기초학력에 도달하는 학생들 중에서 최소능력자의 의미를 재논의하고, 그 특징을 좀 더 구체화하여 강조해야 할 필요성이 제기되었다.

2) 성취수준 설정 2라운드

기초학력 도달 여부의 판단을 위한 수준설정 2라운드에 앞서 패널에게 1라운드에서 다른 패널이 작성한 예상 정답률과 자신의 예상 정답률을 비교, 분석하는 시간이 제공되었다. 이때 각 조에는 다른 조에서 난이도에 따라 문항을 어떻게 구분하였는지, 각 난이도에 대한 조별 예상 정답률의 평균은 얼마인지 등이 제공된다. 또한 실제 학생들을 표집해서 검사를 치렀을 때 나타난 문항별 정답률이 제공되어 2라운드를 시작하기에 앞서 조별로 유사 문항군을 재분류할 기회를 갖는다.

그러나 1차 회의 결과, 수와 연산 영역에 해당하는 문항 중 1개만 문항의 난이도가 변경되었다. 즉, 2조가 1라운드에서 '쉬움'으로 분류하였던 '9번' 문항을 '보통'으로 변경하였을 뿐, 1라운드에서 결정된 문항별 난이도는 조정되지 않았다.

난이도가 변경된 9번 문항은 '받아내림이 없는 세 자리 수의 뺄셈'에 관한 것으로, 2학년에서 다루어지는 내용이다. '받아내림'은 뺄셈에서 많은 학생들이 오류를 보이는 주제이며, 뺄셈 지도에서 가장 주의를 두어야 하는 부분이다(최진숙, 유현주, 2006; 김수미, 2012). 따라서 1라운드에서는 세 자리 수를 다루고 있다고 하더라도 받아내림이 없기 때문에 학생들에게는 '쉬

운' 문항에 해당한다는 의견에 따라 이 문항의 난이도가 분류되었다. 그러나 세 자리 수의 범 위에서의 덧셈과 뺄셈은 2학년 2학기에 처음으로 다루어진다는 점에서 3학년 진입 단계에 있 는, 기초학력에 도달한 최소능력을 가진 학생에게 '쉬운' 문항이 아닐 수도 있다는 논란이 제 기되었다. 이에 따라 1조와 마찬가지로 2조에서도 9번 문항의 난이도를 '보통'으로 조정하였 다.

<표 7> 수준 설정을 위한 2라운드 결과

내용영역	난이도	문항번호	예상정답률 (단위: %)			
			패널 1	패널 2	패널 3	1조평균
			패널 4	패널 5	패널 6	2조평균
수와 연산	쉬움	1 3 25	75	60	81	72.00
		1 3 8 17 25	75	76	75	75.33
	보통	8 9 13 17 21 24	70	55	70	65.00
		6 9 13 15 16 19 21	68	68	68	68.00
어려움	6 11 15 16 19	50	40	63	51.00	
	11 24	40	41	40	40.33	
도형	쉬움	5 12	75	70	74	73.00
		12	75	80	73	76.00
	보통	2 7	70	65	70	68.33
		5 7	66	68	63	65.67
어려움	2	45	55	53	51.00	
측정	쉬움	4 10	80	75	90	81.67
		4 10	80	82	80	80.67
	보통	14 22	70	70	81	73.67
		14 22	70	76	70	72.00
어려움	20 23	68	65	68	67.00	
	20 23	50	41	40	43.67	
확률과 통계	쉬움	28	68	61	65	64.67
	보통	28 29	60	60	74	64.67
		29	60	52	50	54.00
어려움						
규칙성과 문제해결	쉬움	26	70	71	60	67.00
		26	60	63	70	64.33
	보통	27	60	61	50	57.00
		18 27	50	57	59	55.33
어려움	18	25	15	15	18.33	

한편 2007 개정 수학과 교육과정에 대한 성취기준과 평가기준 개발 연구(박소영 외, 2007) 에서 '받아내림이 없는 세 자리 수의 뺄셈'과 관련하여 제시한 평가기준은 <표 8>과 같다.

<표 8> '받아내림이 없는 세 자리 수의 뺄셈'에 대한 평가기준

	상	중	하
받아내림이 없는 세 자리 수끼리의 뺄셈을 할 수 있다.	수식이나 문장으로 된 문제를 보고 받아들임이 없는 세 자리 수끼리의 뺄셈을 형식화하여 능숙하게 계산할 수 있다.	그림으로 제시된 문제 상황을 보고 받아들임이 없는 세 자리 수 끼리의 뺄셈의 원리를 이해하여 계산할 수 있다.	수 모형을 이용하여 받아들임이 없는 세 자리 수끼리의 뺄셈을 할 수 있다.

평가기준에 따르면 수 모형을 이용하여 받아들임이 없는 세 자리 수끼리의 뺄셈은 할 수 있지만, 수 모형이 없이 수식, 그림, 문장 등으로만 제시되는 경우에는 받아들임이 없는 세 자리 수끼리의 뺄셈을 할 수 없다면 기초학력에 도달하지 못한 학생으로 판단할 수 있다. 또한 수식이나 문장을 보고 세 자리 수끼리의 뺄셈을 형식화하여 계산하지는 못하지만, 그림을 이용하여 문제가 제시된 경우에는 받아들임이 없는 세 자리 수끼리의 뺄셈을 할 수 있다면 기초학력에 도달한 최소능력을 가진 학생으로 볼 수 있다.

2라운드에서 난이도가 조정되어 변경된 문항은 적었으나, 대부분의 문항에 대한 예상 정답률이 수정되었다. 패널 4가 '도형' 영역의 '보통' 수준, 패널 5가 '측정' 영역의 '쉬움' 수준에 대하여 예상 정답률을 소폭으로 올렸을 뿐, 그 외의 문항에 대해서는 예상 정답률이 모두 낮아졌다.

3) 성취수준 설정 3라운드

2라운드에서 문항을 11개의 문항군으로 분류한 1조는 각 문항군에 대한 난이도를 판정한 결과의 약 72.7%를 변경하였고, 14개의 문항군으로 문항을 분류한 2조는 난이도를 판정한 결과의 약 83.3%를 조정하였다. 그러나 3라운드에서 1조는 66.7%, 2조는 42.9%의 예상 정답률을 조정하였다.

1조와 2조의 예상 정답률에 대한 평균의 차이가 가장 컸던 문항군은 '규칙성과 문제해결' 영역에서 난이도가 '어려움'에 해당하는 문항들이었다. 1조는 난이도를 '어려움'으로 분류하면서 52.20%로 정답률을 예상하였고, 2조는 16.67%로 예상하여 두 조의 예상 정답률의 평균 차이가 35.53%p였다. 반면에 '도형' 영역에서 '쉬움'으로 분류된 문항에 대한 예상 정답률의 평균 차이는 2%p였다.

3라운드에서 2조는 '9번' 문항의 난이도를 '보통'에서 '쉬움'으로 다시 조정하였다. 3라운드에서는 표집 시행 결과도 제공되었는데, '9번' 문항의 정답률은 96.4%였다. 90% 미만의 정답률이 나온 문항들이 있었음을 고려할 때, 96.4%라는 결과는 초등학교 3학년에 진입한 학생에게는 쉬운 문항이라고 판단한 것이다. 그러나 표집 시행이 초등학교 3학년을 대상으로 11월에 실시되었고, 기본적인 연산에 관한 오류는 학년이 올라갈수록 많이 교정된다는 연구 결과(김수미, 2012)를 고려하여 1조에서는 '보통' 수준으로 난이도를 유지하였다.

<표 9> 수준 설정을 위한 3라운드 결과

내용영역	난이도	문항번호						예상정답률 (단위: %)			
								패널 1	패널 2	패널 3	1조평균
		패널 4	패널 5	패널 6	2조평균						
수와 연산	쉬움	1	3	25				70	60	79	69.67
		1	3	8	9	17	25	74	76	75	75.00
	보통	8	9	13	17	21	24	65	55	70	63.33
		6	13	15	16	19	21	66	68	68	67.33
어려움	6	11	15	16	19		50	40	59	49.67	
	11	24						40	41	40	40.33
도형	쉬움	5	12				70	70	70	70.00	
		12					74	71	71	72.00	
	보통	2	7				65	65	68	66.00	
		5	7				65	64	61	63.33	
어려움	2						35	43	51	43.00	
측정	쉬움	4	10				70	75	79	74.67	
		4	10				80	85	80	81.67	
	보통	14	22				65	65	75	68.33	
		14	22				70	78	70	72.67	
어려움	20	23				60	60	68	62.67		
	20	23				50	45	40	45.00		
확률과 통계	쉬움	28						65	61	65	63.67
	보통	28	29				58	60	69	62.33	
		29					58	52	50	53.33	
어려움											
규칙성과 문제해결	쉬움	26						68	70	60	66.00
		26					60	60	65	61.67	
	보통	27					60	60	50	56.67	
		18	27				50	50	56	52.00	
어려움	18					20	15	15	16.67		

3. 분할점수 산출

성취수준 설정은 학생들의 성취수준을 구분할 수 있는 분할점수(cutt-off scores)를 결정하는 작업을 의미한다. 초3 기초학력 진단평가의 경우 학생들의 성취수준을 기초학력 ‘도달’과 ‘미도달’의 두 수준으로 구분할 하나의 분할점수가 필요하다. 또한 기초학력 진단평가가 기초학력 도달 여부를 판정하고 그에 따라 보정교육을 실시하기 위하여 실시되고 있는 점을 고려할 때, 각 세부 내용영역별 기초학력의 도달 여부를 구분해주는 것도 필요하다. 따라서 본 연구에서는 기초수학이라는 과목에 대한 분할점수를 제시하는 것은 물론, 내용영역별로 분할점수를 제시하여 보정교육을 위한 정보를 제공할 수 있도록 하였다.

<표 10>에서는 3라운드에 걸쳐 진행된 결과를 바탕으로 각 내용영역별로 분할점수의 평균

기초학력 진단평가의 성취수준 설정에 관한 연구 : 초등학교 3학년 ‘기초수학’ 사례를 중심으로

을 구하여 제시하였다. 각 라운드에서 패널들의 유사 문항군에 대한 예상 정답률을 통해 분할 점수를 산출하는 데, 문항 범주별로 ‘문항 배점 합’과 ‘성취수준별 최소 능력자의 예상 정답률’을 곱하고, 모든 범주에 대해 합산하여 개별 교사의 성취수준별 분할점수를 산출한다. 이를 공식으로 제시하면 다음과 같다.

$$\text{분할점수} = \sum (\text{범주별 문항 정답률} \times \text{범주별 문항 배점 합})$$

라운드가 진행될수록 각 조에서 제시된 평균 점수는 낮아지고, 그 차이도 줄어드는 경향이 나타났다. ‘수와 연산’ 영역을 제외한 네 영역에서 1조의 평균이 2조보다 높게 나타났다.

<표 10> 내용영역별 조별 예상 정답률의 평균

내용영역	1라운드		2라운드		3라운드	
	1조	2조	1조	2조	1조	2조
수와 연산	30.20	32.45	28.50	29.96	27.72	30.00
도형	8.60	8.15	8.48	7.75	8.16	7.25
측정	15.10	11.90	13.34	11.78	12.34	11.96
확률과 통계	7.33	7.20	6.47	5.83	6.23	5.75
규칙성과 문제해결	8.82	8.73	8.20	7.32	7.76	7.20
합계	70.05	68.43	64.99	62.64	62.21	62.16

본 연구에서는 이와 같은 결과를 바탕으로 각 내용영역별 기초학력 도달 여부를 결정하는 분할점수를 선정하였다. 그 결과, ‘도형’ 영역과 ‘규칙성과 문제해결’ 영역을 제외한 영역에서는 모두 분할점수가 영역별 배점의 60%보다 높게 선정되었고, 분할점수의 합계는 63점이었다.

<표 11> 하위 영역별 분할점수

내용 영역	배점	분할점수
수와 연산	46	30
도형	12	7
측정	18	12
확률과 통계	10	6
규칙성과 문제해결	14	8
합계	100	63

그러나 <표 10>에서 3라운드의 결과를 반올림하지 않은 채 산출한 합계를 고려하여 2013년 초3 기초학력 진단평가 기초수학의 분할점수는 ‘62점’으로 결정하였다.

V. 결론 및 제언

국가수준에서 학생들의 기초학력 도달 여부를 판단하고 그에 따른 보정교육을 목적으로 실시하는 초3 기초학력 진단평가는 초등학교 3학년에 진입하는 수준에서 기초학력에 도달한 최소능력자를 기준으로 분할점수를 설정한다. 본 연구에서는 초등학교 2학년 또는 3학년을 지도한 경험이 있는 교사를 대상으로 '변형된 Ebel방법'을 적용한 수준설정을 진행하여, 2013년 초3 기초학력 진단평가 기초수학의 분할점수를 설정하였다.

변형된 Ebel방법은 Angoff방법에 비해 사용 방법이 간단하고 시간도 절약되어 최근 교사들이 선호도를 보이고 있는 성취수준 설정 방법이다. 성취수준 설정은 어떤 방법을, 어떤 절차에 걸쳐 진행되었으며 과정에 영향을 미친 변인이 무엇인지를 명시하는 것이 중요하다(장은아, 2007). 따라서 본 연구에서는 2013년 초3 기초학력 진단평가 기초수학에서 성취수준 설정 과정을 자세히 설명하고, 각 라운드별 결과와 활용된 자료들에 대하여 설명하였다.

본 연구 결과, 다음과 같은 몇 가지 시사점을 얻을 수 있었다. 첫째, 성취수준 설정을 진행하는 과정에서 조별 회의를 통해 각 문항의 난이도 분류와 예상 정답률을 조정할 수 있었음에도 불구하고, 패널들은 조별 회의를 통해 문항의 난이도 분류 수정에는 소극적이었다. 성취수준 설정의 초기 단계에서 문항을 검토하며 조별로 합의한 난이도에 대한 생각이 쉽게 변하지 않음을 알 수 있다. 일반적으로 패널은 자신이 지도하는 학생들의 수준에 근거하여 문항에 대한 난이도를 결정하는 경향이 있다(홍소영, 2011). 즉, 표집시행 결과에서 나타난 실제 학생들의 문항 정답률이 제시되어도 개인적 경험을 바탕으로 하여 문항의 난이도를 판단하고 그 수치를 수정함에 있어 소극적이었다. 따라서 교사가 개인적인 경험 이외에 문항의 난이도에 대하여 객관적인 판단을 할 수 있는 방안이 좀 더 적극적으로 모색되어야 할 것이다.

둘째, 문항의 난이도를 '쉬움/보통/어려움'으로 구분할 때, 각 수준에 대한 인식의 차이가 크게 나타났다. 주언희(2013)는 성취수준 진술의 중요성을 강조하였는데, 성취수준을 어떻게 기술하고 기술된 성취수준을 각 패널이 어떻게 이해하느냐에 따라 분할점수는 달라질 수 있기 때문이다. 따라서 기초학력에 도달한 최소능력자에 대한 충분한 논의를 통해 최소능력자의 개념을 구체화하고 성취수준 진술을 명료하게 하는 방안이 모색되어야 할 것이다.

성취평가제에서는 90%이상의 성취율을 나타내는 성취도 'A'에 대하여 '내용영역에 대한 지식습득과 이해가 매우 우수한 수준이며 새로운 상황에 일반화할 수 있음'이라고 제시하고 있다(교육과학기술부, 2011). 그러나 학교현장에서는 보다 구체화된 성취수준이 설정되어야 하므로, 단위학교별로 같은 학년을 담당하는 교사 간 협의에 의해서 결정될 가능성이 크다. 특히 2012년을 기준으로 전국 국공립 초등학교의 평균 학급수가 20.6학급인 점을 고려할 때(한국교육개발원, 2012), 6명의 패널을 통해 성취수준 설정을 진행한 본 연구는 그 의미가 크다고 할 수 있다. 따라서 성취수준 설정의 초기 단계에서 형성된 문항에 대한 난이도, 최소능력자에 대한 생각이 성취수준 설정이 진행됨에 따라 변하지 않았다는 연구 결과를 고려하여 단위 학교에서도 성취수준 설정 시에는 초기 단계부터 충분한 논의와 합의가 전제되어야 할 것이다. 또한 패널의 특성은 성취기준 설정 결과에 많은 영향을 주게 되므로 본 연구에 참여한 패널의 규모 뿐 아니라 그 특성과 결과와의 관련성을 학교 수준에서 면밀하게 비교하여 적용하는 방안을 모색해야 할 것이다.

최근 교육평가의 동향은 규준참조평가(norm-referenced assessment)에서 준거참조평가(criterion-referenced assessment)로 전환되고 있다. 이전에는 학생과 학생 간 비교를 통하여

학생 사이의 경쟁을 부추겼던 반면, 이제는 학생과 교육목표, 학생과 성취기준 간의 비교를 통해 학생 자신이 배워야 할 것, 목표로 하고 있는 것을 꾸준히 성취하고, 성장하고 있는지를 판단하고, 경쟁이 아닌 상생·협력하는 학교문화를 조성하고자 한다. 이러한 평가 동향은 국제적인 것으로, 해외에서도 더 이상 '결과지향평가'(assessment of learning)가 아닌 '학습지향평가'(assessment for learning)를 강조하고 있다. 따라서 학생의 추후 학습과 성장을 위해 도움이 되는 평가로서 학생의 현재 위치, 세부 내용영역에서의 강약점을 제시해줄 수 있도록 준거 참조평가 결과의 활용을 극대화하기 위해서는 '성취수준 설정'이 신뢰롭고 타당하게 이루어져야 하며, 학교현장에서 교사가 쉽게 활용할 수 있는 성취수준 설정 방법이 보급되어야 할 것이다.

본 연구에서 성취수준 설정 방법으로 소개한 '변형된 Ebel방법'은 문항을 유사 문항군으로 분류하여 성취수준 설정을 진행하므로 분류 방법을 잘 변형하면, 총점 측면에서의 분할점수, 각 세부 내용영역에서의 분할점수를 동시에 산출할 수 있기 때문에 효율성이 크다. 즉, 본 연구에서의 경우와 같이 문항 분류에 있어서 세부 내용영역(수와 연산, 도형, 확률과 통계, 측정, 규칙성과 문제해결)으로 분류하게 되면 기초수학 총점에서의 분할점수와 함께 각 세부 영역에서의 분할점수를 성취수준 설정의 결과로서 제시할 수 있게 된다. 따라서 학생들에게 기초수학 전반에 대한 수준과 함께 각 세부 영역에서 더 노력해야 하는 부분을 피드백 해줄 수 있다.

본 연구에서와 같이 학교 현장에서 교사가 직접 활용할 수 있는 성취수준 설정 방법 및 적용 사례를 구체적으로 제시하고, 이에 대한 효율적이고 간편한 변형 방법 등이 지속적으로 소개된다면, 학교 수준에서의 성취수준 설정 방법이 안정화되는 것은 물론 학생 평가에 있어 상대평가에 의한 서로 간의 경쟁 유도가 아닌 학습 목표를 위한 내면적인 동기를 부여해주는 교육 환경을 조성하는 데 기여할 수 있을 것이라 기대된다.

참고 문헌

- 고정화(2008). 초등학교 3학년 수학 기초학력 미도달 학생의 특징 분석. *수학교육학연구* 18(3), 283-308.
- 교육과학기술부(2010). 2010년도 기초학력 향상 지원 계획. *교육과학기술부(교육정보기획과 2010.5.)*
- 교육과학기술부(2011). 창의·인성 교육 강화를 위한 「중등학교 학사관리 선진화 방안」 발표 -고교 석차 9등급제 평가를 성취평가제로 전환- *교육과학기술부 보도자료(2011.12.13.)*
- 김경희, 박은아, 송미영, 상경아, 김수진, 김희경, 신진아, 서지영, 이채희, 김지영, 김기연, 최숙기(2012). *중등학교의 성취평가제 운영 방안 연구*. 한국교육과정평가원 연구보고 CRE 2012-8.
- 김명화, 김도남, 권점례, 김완수, 황인우, 강태훈(2011). 3R's 기초학습부진 선별 도구 개발 연구. *한국교육과정평가원 연구보고 CRE 2011-10.*
- 김수미(2012). 학년 상승에 따른 초등학생들의 자연수 사칙계산 오답유형 및 오답률 추이와 그에 따른 교수학적 시사점. *한국초등수학교육학회지* 16(1), 125-143.
- 김정호, 이화진, 채선희, 김진숙, 박태호, 이재승, 황혜정, 윤현진, 조미혜, 양종모, 이경언(1999).

- 제7차 교육과정에 따른 성취기준과 평가기준 개발 연구: 초등학교 1, 2학년. 한국교육과정평가원 연구보고 RRC 99-5.
- 김희경, 김도남, 박인용, 조성민, 최인봉, 한정아, 최숙기(2013). 온라인 3R's 기초학력 진단 평가도구 개발. 한국교육과정평가원 연구보고 CRE 2013-3.
- 박선화, 구영산, 김정훈, 김성혜, 김현미, 박은아, 서지영, 손민정, 양운정, 윤지훈, 이문복, 이상하, 이용백, 이채희, 진의남, 최혁준, 황인표(2012). 성취평가제 도입 운영을 위한 자료 개발 및 연수·홍보. 한국교육과정평가원 연구자료 ORM 2012-121
- 박소영, 오은순, 이재기, 박선화, 윤현진, 김인숙, 김평국(2007). 2007년 개정 교육과정에 따른 초등학교 2학년 절대평가기준 개발 연구. 한국교육과정평가원 연구보고 CRC 2007-7.
- 서울대학교교육연구소(1994). 교육학용어사전. 도서출판 하우.
- 성태제(1991). 목표지향 검사를 위한 준거설정 방법: 오답 추출능력에 의한 준거설정 방법의 타당성. 교육학연구 29(2), 147-164.
- 성태제(2010). 현대교육평가. 학지사.
- 양명희(2006). 우리나라 초3 아동들의 기초학력과 정의적 특성 탐색-기초학력 진단평가 결과를 중심으로. 교육발전연구, 22(1), 47-71.
- 오은순, 박소영, 이재기, 박선화, 윤현진, 김인숙, 김평국(2007). 2007년 개정 교육과정에 따른 초등학교 1학년 절대평가기준 개발 연구. 한국교육과정평가원 연구보고 CRC 2007-6.
- 이인재, 최석진, 이재기, 이봉주, 채선희, 김도남, 이규민, 강미현, 김혜숙, 김수정(2004). 2003년 초등학교 3학년 국가수준 기초학력 진단평가 연구-종합. 한국교육과정평가원 연구보고 CRE 2004-1-1.
- 임천택(2009). 쓰기 기초학력의 재개념화와 평가 틀의 개선 방향. 학습자중심교과교육연구, 9(1), 1-21.
- 장은아(2007). Angoff 방법과 북마크 방법을 활용한 한국어능력시험의 기준 설정에 관한 연구. 고려대학교대학원 박사학위논문.
- 정혜영(2010). 국가수준 초등 기초학력 학업성취도 평가의 운영 및 논점: 미국, 영국, 캐나다, 프랑스, 일본을 중심으로. 초등교육연구 23(4), 157-179.
- 조영미(2006). 수학 기초학력 부진아 지도를 위한 교과서 및 교사용 지도서의 개선 방안 탐색 -초3 국가수준 기초학력 진단평가 기초 수학 결과 분석-. 학교수학 8(1), 69-83.
- 주언희(2013). 기준설정을 위한 패널의 전문성 및 라운딩 규칙 분석-원점수방법, Angoff방법, Bookmark방법을 중심으로. 경희대학교 대학원 석사학위논문.
- 최진숙, 유현주(2006). 덧셈·뺄셈의 오류유형 분석 및 지도방안에 대한 연구-초등학교 3학년을 중심으로- 교과교육학연구 10(2), 303-327.
- 한국교육개발원(2012). 2012 교육통계분석자료집-유·초·중등교육통계편. 한국교육개발원 통계자료 SM 2012-13-1.
- 한국교육과정평가원(2012. 8.). 성취평가제 컨설팅 교수단 워크숍 자료.
- 한국교육평가학회(2004). 교육평가 용어사전. 학지사.
- 홍소영(2011). 초등학교 학업성취도평가의 기준설정을 위한 Angoff와 Bookmark 방법의 비교. 학습자중심교과교육연구 11(4), 495-523.
- ACT(2005). *Developing achievement levels on the 2005 National Assessment of Educational Progress in grade twelve mathematics*. Process Report. ACT, Inc.
- Ebel, R. L.(1972). *Essentials of educational measurement*. Prentice-Hall, Inc.

기초학력 진단평가의 성취수준 설정에 관한 연구 : 초등학교 3학년 '기초수학' 사례를 중심으로

- Kane, M(1994). Validating the performance standards associated with passing scores.
Review of Educational Research 64(3), 425-461.
- Livingston, S. A., & Zieky, M. J. (1982). *Passing score*. Educational Testing Service.

A Study on the Standard Setting for the Basic Skills Competency Test: Focusing on 3rd Grade Basic Mathematics

HeeKyoung Kim⁴⁾ · Seongmin Cho⁵⁾

As education is one of key factors to improve national competitiveness, quality education becomes a top priority in all countries around the world. In South Korea, The Basic Skills Competency Test is administered annually to select below-basic third graders and provide intervention programs. The Basic Skills Competency Test was designed to diagnose the minimal competency of third graders according to a national-level standards. The purpose of this study was to analyze standard-setting procedures and set cut-off scores. This study offered the empirical evidence that standard-setting is generally applicable in schools by the modified Ebel method. More future researches addressing appropriate and efficient standard-setting methods in real school situations are needed to be taken in.

Key Words : Standard-setting, Cut-off Score, Basic Skills Competency Test, Ebel Method

Received September 13, 2013

Revised September 23, 2013

Accepted September 26, 2013

4) Korea Institute for Curriculum and Evaluation(heekyoung@kice.re.kr)

5) Korea Institute for Curriculum and Evaluation(csminy@kice.re.kr), corresponding author