

미국의 NAEP 수학 평가들에 관한 고찰

조 영 미

(한국교육과정평가원 연구원)

《요약》

이 논문에서는 우리나라 국가수준의 학업성취도 평가와 유사한 목적을 지닌 미국 NAEP의 수학 평가들에 관하여 고찰하였다. 최근 한국교육과정평가원에서는 국가수준의 학업성취도 평가를 통하여 좀더 유의미한 분석 결과를 보고하기 위하여 평가 시행의 각 단계를 발전시키려는 노력을 다각도로 벌이고 있다. 이러한 노력의 일환으로, 이 논문에서는 평가 시행의 첫 단계인 '수학 평가를 개발'에 포커스를 두고 현재의 평가들에서 내실을 기해야 할 점과 향후 발전방향을 모색하고자 하였다. 이를 위해 30여년의 역사를 지닌 미국의 NAEP 수학 평가들 중에서 특별히 1990, 1996, 2005년을 중심으로 하여 개정 배경, 주요 특징과 요소별 특징으로 내용 영역과 행동 영역에 해당하는 부분을 중점적으로 살펴보았다. 이러한 고찰을 통하여 수학 평가들과 관련된 시사점을 이끌어낼 수 있었는데, 먼저 평가들의 요소별 시사점으로, 행동 영역, 문항 출제 비율, 성취기준 진술 방식 등에 관한 심도 있는 연구가 추후 필요함을 지적하였으며, 다음으로, 일반적인 시사점으로는, 다양한 사람들의 참여, 수학교육 연구의 적극적 반영 노력, 평가를 개발자의 디자인 마인드 필요 등 일곱 가지를 제시하였다.

주제어 : NAEP, 국가수준의 학업성취도, 수학 평가를

I. 서론

한국교육과정평가원(이하 평가원)에서는 교육인적자원부와 16개 시·도 교육청의 지원 아래 매년 '국가수준 학업성취도 평가'(이하 성취도 평가)를 실시하고 있다. 현재 성취도 평가는 전국의 학생들 중에서 초중고 학년급 별로 학생들을 표집하여 국어, 사회, 수학, 과학, 영어 과목에 대해 평가를 하고 그 결과를 분석하며, 이를 통하여 국가수준의 교육과정에서 규정하고 있는 교과별 교육목표에 대한 학생들의 학업성취도를 파악하고, 학업성취도의 증가기 추이를 분석하여 우리나라 학생들의 학력 추이 변화를 파악하는 것을 주요 목적으로 하

고 있다. 이는 1998년 기초연구를 시작으로, 올해로 6년째 접어들었으며, 수학의 경우 5번째 평가를 실시하게 된다.

일반적으로 학업성취도에 대해서는 많은 사람들이 관심을 가지고 개인적으로, 또는 소규모 집단을 꾸려 연구하고 있다. 이에 대해 평가원에서 실시하는 성취도 평가는 국가적 수준에서 수행하는 것으로 국내에서는 평가 대상이나 예산 등에서 타의 추종을 불허할 정도의 대규모 연구이며, 초중등교육법 제9조 제1항에 ‘교육인적자원부는 학교에 재학 중인 학생의 학업성취도를 측정하기 위한 평가를 실시할 수 있다’라는 법조문이 있을 정도로 국가적으로 중요하게 여겨지는 연구이다.

성취도 평가의 시행 단계는 대략 ‘평가틀 개발 → 평가도구 및 설문지 개발 → 예비검사의 시행과 그 결과 분석을 통한 본검사 평가도구 확정 → 평가[본검사] 시행 → 채점 및 결과 분석 → 보고서 작성’으로 이루어진다. 매년 실시되고 있는 이 연구에서는 우리나라 학업성취도와 관련하여 좀더 유의미한 정보들을 보고서에 담기 위하여 다각도로 노력하고 있다. 예컨대, 2003년 연구의 [평가도구 개발] 단계에서는 평가를 시행하는 5개 교과 중에서 일부 교과를 집중교과, 나머지 교과를 기본교과로 정하여, 집중교과의 경우 결과에 대한 심층 분석이 가능하도록 평가도구에 사용되는 문항의 수를 이전보다 큰 폭으로 늘여 개발하였으며, 학생들의 오류나 사고를 좀더 잘 드러낼 수 있도록 수행평가 문항을 개발하였다. 또한 [채점 및 결과 분석] 단계에서는 학생들의 성취수준을 우수, 보통, 기초, 기초미달로 구분하는 방법으로 수정된 Angoff 방법을 처음 사용하여 현장 교사들의 경험과 감각을 토대로 한 실제적인 성취수준을 설정하려고 하였다. 이와 같은 노력들은 성취도 평가 시행의 각 단계들을 좀더 내실 있게 하여 우리나라 학생들의 학업성취도에 관한 타당하면서 유의미한 정보를 이끌어내어 보고하고자 하는 것이다.

이 논문에서는, 이러한 노력의 하나로, 평가 시행의 첫 단계인 [(수학) 평가틀 개발]에 초점을 맞추고자 한다. 수학 평가틀에는 내용 영역, 행동 영역, 성취기준, 성취수준, 문항 유형, 영역 별 문항 출제 비율 등이 속하며, 이에 준하여 평가도구가 개발되고 이 평가도구를 통하여 나온 결과를 분석한다는 점을 볼 때, 성취도 평가에서 평가틀은 매우 중요하다.

국내 성취도 평가의 경우 99년 수학 평가틀에 대한 기본 안이 마련된 이후로 그 내용 중 일부 내용이 삭제되거나 새로운 내용이 첨가되는 등의 변화를 거치면서 현재의 수학 평가틀로 되었다. 예컨대 2003년 성취도 평가에서는 내용 영역을 수와 연산, 측정, 도형, 문자와 식, 확률과 통계, 규칙성과 함수로 설정하였으며, 행동 영역에 대해서는 고려하지 않았으며, 문항 유형으로는 선택형과 수행형 문항을 사용하였다. 사실 현재 이 수학 평가틀로도 학생들의 학업성취도에 관한 기본적인 정보, 이를테면 성취수준 별 비율이나 성별, 지역별 차이 등을 충분히 파악할 수 있다. 그런데, 교육이 국가적으로 차지하는 비중과 그 중 특히 학력 관리에 대한 국가적인 관심을 감안할 때, 우리나라 전체 학생들의 학력의 경향성을 파악할 수

있는 유일한 연구인 이 성취도 평가는 앞으로도 지속적으로 이루어질 것으로 보이며, 오히려 그에 대한 관심이 증폭될 가능성이 다분하다. 따라서 성취도 평가에서 어느 정도 기능을 수행하고 있다고 해서 현재의 평가들에 만족하고 안주하기 보다는 보완할 부분은 없는지 점검하여 내실을 기하고, 더 나아가서 향후 방향에 대해서도 심사숙고할 필요가 있다.

이를 위하여 이 논문에서는 미국에서 국가적인 수준에서 진행되는 National Assessment of Educational Progress(이하 NAEP)의 수학과 평가들의 1990년대 이후의 변천 과정을 살펴보았다. NAEP의 목적은 국내 성취도 평가와 거의 유사하며 그 역사가 30여년이 넘었다. 그러한 NAEP의 수학 평가들의 변천과정을 짚어보면서 NAEP 수학 평가에 국내 성취도 평가를 비추어 보았으며, 그로부터 국내 성취도 평가의 수학 평가들에서 내실을 기해야 할 점이나 발전 방향에 관한 시사점을 얻을 수 있었다. 특히 수학 평가들의 구성요소에서 좀더 심사숙고해야 할 점은 무엇이며, 국가 수준에서 진행되는 수학 평가를 개발 과정은 어떠한가 하는가에 관한 아이디어를 얻을 수 있었다.

Ⅱ. NAEP 수학 평가들 : 1990, 1996, 2005

1. NAEP 개관

미국에서는 1969년부터 NAEP를 통하여 주기적으로 학생들의 학업성취도와 그 변화 추이를 평가하고 그 결과를 ‘National Report Card’라는 제하의 보고서로 내놓고 있다. 평가 대상 과목은 예술, 공민, 지리, 수학, 읽기, 과학, 국사 등이며, 대체로 2-3년을 주기로 두서너 과목을 번갈아가면서 평가를 실시하고 있다. 평가 대상 학년은 4, 8, 12학년이며, 학생 수는 점점 확대되어 2003년에 실시된 평가의 경우 4학년에서는 190,000여명, 8학년에서는 153,000여명이 참여하였다. NAEP에 관여하는 기관도 다수로, 한 예로 1996년 NAEP 평가 기술보고서에 따르면, NCES(National Center for Education Statistics), NAGB(National Assessment Governing Board), CCSSO(Council of Chief State School Officers), ETS(Educational Testing Service), NCS(National Computer System), Westat, INC 등이 참여하였다.

NAEP은 크게 국가 수준의 평가(National NAEP), 주 수준의 평가(State NAEP), 추이 평가(Trend NAEP)로 이루어진다. 처음 NAEP이 실시될 때에는 국가 수준의 평가만이 실시되다가 1988년 국회에서 NAEP에서 주 수준의 평가를 실시할 수 있도록 하는 법을 통과시킨 후로

주 사이 비교를 목적으로 한 주 수준의 평가가 실시되었다. 국가 수준의 평가와 주 수준의 평가는 사용하는 평가 문항은 동일하며, 평가 대상이 다르다는 점에서 차이가 있다. 추이 평가는 학생들의 학력이 시간이 흐름에 따라 어떠한 변화를 보이고 있는가를 알아보는 것으로, 일정 수의 공통 문항을 몇 년에 걸쳐 반복적으로 사용하여 평가한다.

2. NAEP 수학 평가들의 변천

1973년 NAEP에서 최초로 수학 평가를 실시한 후로 30여 년 동안 모두 10번의 수학 평가가 실시되었다. 그동안 평균적으로 3년에 한 번꼴로 수학 평가가 실시된 셈이다. 향후 평가 계획에 따르면 2005년에 수학 평가가 실시되며 이후로는 2년마다 수학 평가를 실시하도록 되어 있다.

1990년 이후로 NAEP 수학 평가들은 크게 3번 바뀌었다. 1990년 평가들이 바뀌어 이 틀로 1992년 평가까지 치루었으며, 1996년에 다시 바뀐 틀은 2000년, 2003년 평가에서 사용되었으며, 내년 2005년 평가에서는 새로운 평가들이 적용된다. 이 장에서는 1990, 1996, 2005년 수학 평가들의 개정 내용을 살펴보았으며, 특히 개정 내용 중에서 개정의 배경, 주요 특징, 요소별 특징에 관하여 살펴보았다. 특히 요소별 특징에서는 내용 영역과 행동 영역(수학적 능력, 수학적 힘 등)을 중심으로 살펴보았다.

가. 1990 수학 평가들

1) 배경

1988년 국회에서는 NAEP에서 주 수준의 평가를 실시하여 주 사이에 학생들의 학업성취도를 비교할 수 있도록 하는 관련 법률을 통과시켰다. 그전까지는 국가 수준에서, 즉 주와 무관하게 전국 학생들 중 표집하여 평가하고 그 결과를 보고하여 왔기 때문에, 새롭게 주 별 간 비교를 목적으로 한 주 수준의 평가를 실시하기 위해서는 NAEP 평가 방식을 대폭 개정하여야 했다. 예컨대, 미국에서는 국가에서 정한 단일한 교육과정이 없기 때문에, 주 수준의 NAEP 평가를 위해서는 교과별로 평가할 내용 영역을 구체적으로 정하는 일이 시급하였다. 한편, 국회에서는 1988년 NAGB를 신설하여 NAEP 정책 전반을 책임지도록 하였다. 당시 이 기구는 정책가, 법률가, 학교장, 교육학자 등 각계각층의 대표 26명으로 구성되었으며, 현재까지도 NAEP 관련 정책을 수행하고 있다.

NAGB는 CCSO의 주도 아래 정책입안자, 변호사, 시민 등 18명으로 이루어진 국가평가기획단(National Assessment Planning Project: NAPP)을 구성하여 주 수준 평가의 목적과 조건에 관한 정책적인 ‘합의문’을 채택하고, 다음으로, 다양한 주에서 온 교사, 학교장, 수학교육전문

가들로 구성된 수학 내용체계 위원회(Mathematics Objectives Committee : MOC)를 구성하여, 앞서 언급한 합의문을 바탕으로, 평가를 위한 내용체계를 비롯한 수학과 평가들을 본격적으로 개발하도록 하였다.

2) 주요 특징

당시 MOC에서는 수학 평가들로 5개의 내용 영역 차원과 3개의 수학적 능력 차원으로 이루어진 격자 틀(matrix view)을 확정하였다(<표 II-1>). 이 표에 따르면, 예를 들어, (가)의 문항을 출제한다는 것은 수와 연산에 속하면서 ‘동시에’ 개념적 이해에 해당하는 문항을 출제함을 뜻한다. 이 틀대로 문항들이 개발된다면 상당한 내용 영역과 수학적 능력을 평가할 수 있는, 균형감 있는 평가지가 개발되고 그러한 평가지로 학생들이 치른 시험의 결과는 학생들의 수학 학력의 면면을 파악하는 데 포괄적이면서 유익한 정보를 줄 것으로 본 것 같다.

〈표 II-1〉 1990 수학 평가틀 : 격자들

수학적 능력	내용 영역				
	수와 연산	측정	기하	자료 분석, 통계, 확률	대수와 함수
개념적 이해	(가)				
절차적 지식					
문제 해결					

3) 요소별 특징

가) 내용 영역

1990 수학 평가틀에서는 ‘수와 연산’, ‘측정’, ‘기하’, ‘자료 분석, 통계, 확률’, ‘대수와 함수’ 등으로 구분하였으며, 각 영역별 문항 출제 비율은 다음 표와 같다.

〈표 II-2〉 1990 내용 영역별 문항 출제 비율

	4학년	8학년	12학년
수와 연산	45	30	25
측정	20	15	15
기하	15	20	20
자료 분석, 통계, 확률	10	15	15
대수와 함수	10	20	25

(수치는 백분율)

내용 영역 측면에서 이전의 평가와 1990 평가의 차이는, 문항 출제 비율을 정할 때 기하나 대수, 함수와 같은 내용 영역을 강조하고 대신 수와 연산 영역을 과거와는 달리 덜 강조

하였다는 점을 들 수 있다(NAGB, 1996). 즉, 4학년에서 45%를 차지한 수와 연산 영역의 비중은 단계가 높아짐에 따라 점점 낮아지고, 그 대신 기하와 대수, 또는 함수 영역의 비중이 높아지도록 하였던 것이다. 이는 당시 수학교육에서 수학은 단순 알고리즘의 습득 이상이라는, 당시로서는 진보적인 생각이 반영된 것으로 보인다.

다음으로, NAEP 수학 평가들에서 한 내용 영역에서 평가 요소가 진술된 ‘방식’을 유심히 살펴볼 필요가 있다. <표Ⅱ-3>에 제시된 것과 같은 1990년 방식은, 이후 내용에서 살펴보겠지만, 1996, 2005년 평가들에서 바뀐다. 1990년의 평가 요소들을 짚막하게 지식 중심으로 진술하는 방식은 모든 내용 영역에 공통적이다. 이 방식에서, 가르치고자 하는 지식을 세부 요소들로 분해하고 각각의 요소들의 학습을 통하여 지식을 가르치고자 한 행동주의의 모습을 엿볼 수 있다.

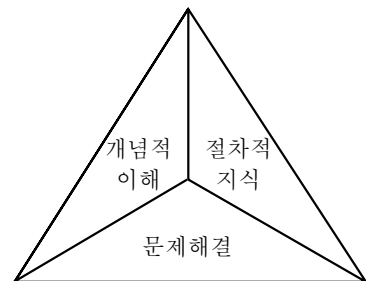
〈표 Ⅱ-3〉 1990 평가들의 평가요소 진술 방식의 예

주제 : 수와 연산	단계		
	4	8	12
1. 세기, 모으기와 자리값 관련 지우기			
a. 자연수 자리값	•	•	•
b. 자연수 어렵하기	•	•	•
c. 소수의 자리값	•	•	•
d. 소수 어렵하기	•	•	•
e. 양의 순서(자리값과 관련된 추정)	•	•	•
f. 기호		•	•

[참고] •: 해당 학년에서 그 평가 요소의 문제가 출제됨을 의미

나) 수학적 능력

NAEP은 1990년 이전 평가들에서 5가지로 구분되었던 과정 영역 차원을, 1990 평가들에서는 간소화하여 개념적 이해, 절차적 지식, 문제해결의 3가지 요소로 이루어진 ‘수학적 능력’ 차원으로 바꾸었다. 1989년 발간된 NCTM의 Curriculum and Evaluation Standards for School Mathematics(이하 Standards)에서는 평가 기준으로 ‘수학적 힘, 문제해결, 의사소통, 추론, 수학적 개념, 수학적 절차’ 등을 들고 있다. 이 중에서 ‘문제해결, 수학적 개념, 수학적 절차’가 수학적 능력 차원의 요소와 유사함을 알 수 있는데, 이로부터 Standards가 미친 영향을 확인할 수 있다.



[그림 Ⅱ-1] 수학적 능력의 요소 사이의 관계

[그림Ⅱ-1]의 의미는 다음과 같다. 매 단계에서 문제해결은 개념적 지식과 절차적 이해의

상호작용을 내포하고, 한 단계에서 복잡한 문제해결로 간주된 것이라도 다른 단계에서는 개념적 이해나 절차적 지식으로 간주될 수 있다. 이는 NAEP 수학 평가들에서 수학적 능력의 요소인 문제해결, 개념적 이해, 절차적 지식은 그 위치가 절대적으로 정해지기 보다는 학년급이나 상황에 따라 달라질 수 있는 상대적인 것임을 말해준다(NAEP, 1998).

각 수학적 능력의 요소별 능력 특성은 <표 II-4>와 같으며, 문항 출제 비율은 개념적 이해 40%, 절차적 지식 30%, 문제해결 30%로, 모든 단계에 똑같이 적용하였다.

<표 II-4> 수학적 능력의 요소 별 능력 특성

수학적 능력	하위 능력
개념적 이해	예나 반례를 알아내거나, 이름 붙이거나, 만들어 내기
	개념을 표현하기 위해 모델, 다이어그램, 상징 등을 사용하기
	원리를 알아내고 적용하기
	사실과 정의를 알고 적용하기 (이하 생략)
절차적 지식	적당한 절차를 올바르게 선택하여 적용하기
	절차의 적용이 타당함을 정당화하기
문제해결	문제를 알고 형식화하기
	자료의 충분함과 일관됨을 이해하기
	전략, 자료, 모델, 관련된 수학 지식을 사용하기
	절차를 만들어내거나 확장하거나 수정하기 (이하 생략)

나. 1996 수학 평가들

1) 배경

1990 NAEP 수학 평가가 시행된 이듬해인 1991년 하반기에 NAGB는 College Board에 1994년에 실시될 수학 평가를 위한 평가틀과 문항 개발 내역을 위한 지침을 마련토록 하였다. 이는 당시 NCTM에서 발간한 Standards와 Professional Standards for Teaching Mathematics (1991)에 드러난 수학교육개혁의 동향과, NAEP 수학 평가를 실질적으로 연계하기 위함이었다(NAGB, 1996). 그런데 막상 이 평가틀에 근거한 평가는 예산 문제로 1994년에 실시되지 못하고 1996년에 실시되었다.

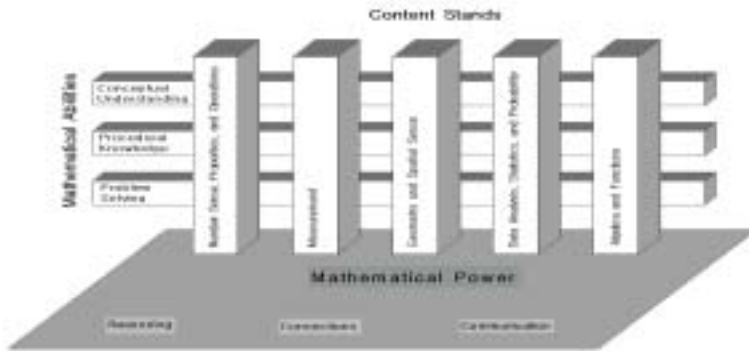
NAGB로부터 수학 평가틀 개정 권한을 위임 받은 College Board는 정책입안자, 사업가 등 12명의 대표자로 구성된 운영위원회를 조직하여 이 프로젝트의 방향과 범위를 정하고, 다음으로 7명의 수학교육자들로 기획위원회를 조직하여 수학 평가 틀의 초안을 마련하였다. 이때 수학교육개혁 방향과 1990년부터 NAEP에서 새로이 추진한 주 수준의 평가에서의 개혁 방향, NAEP의 평가와 학교수학 교육 사이의 관련성 등을 중점적으로 고려하였다.

2) 주요 특징

NCTM의 Standards는 NAEP 1996 수학 평가들에 적지 않은 영향을 끼쳤다. 1990 수학 평가들에도 이 저서가 영향을 미치기는 하였지만 당시는 이 책의 초고를 참조한 상태였으며, 이 책이 온전하게 영향을 미친 것은 이 때부터라고 보아야 할 것이다. 앞서 지적한 대로, Standards에서는 평가규준으로 ‘수학적 힘, 문제해결, 의사소통, 추론, 수학적 개념, 수학적 절차’를 제시하였으며, 이 중 수학적 개념, 수학적 절차, 문제해결 등은 1990 평가들에서 수학적 능력의 요소로 사용되었다. 1996 평가들에서는 평가규준의 하나인 수학적 힘을 새로운 차원으로 설정하였으며, 이 차원의 하위 요소로, 역시 평가규준인 추론, 연결성, 의사소통을 설정하였다. 이렇게 해서 1996 평가들은 수학 내용 영역, 수학적 능력, 수학적 힘의 세 가지 차원으로 이루어졌다.

1990 수학 평가들 중에서 주된 개정 대상은 5가지 수학 내용 영역과 3가지 수학적 능력으로 이루어진 ‘격자들’이었다. 문항을 실제로 개발할 때 격자들이 안고 있는 문제점이나 한계들이 하나하나 드러나면서 이를 개정하게 된 것으로 보인다. 예컨대, 학생들의 수학 학력의 최대한 많은 측면을 평가할 수 있는 균형감 있고 이상적인 평가지를 구성할 수 있는 틀로 기대되었지만, 막상 문항의 개발 단계에서 격자들의 각 셀에 대응되도록 문항을 개발하기가 쉽지 않았던 것이다. 개발된 문항을 수학적 능력을 숙지한 전문가에게 그 하위 요소별로 분류하도록 하였을 때 그들조차도 개발 의도와는 다르게 분류할 정도로, 문항을 수학적 능력의 요소들 중 어느 하나로 구분하는 작업이 명쾌하지 않았다. 격자들의 의도(<표Ⅱ-1> 참조)에 따라, 내용 영역과 수학적 능력을 ‘동시에’ 고려하여 문항을 개발하는 작업은 상당히 까다로워 문항 개발자들이 활발하게 아이디어를 내는 데 방해가 되었던 것으로 보인다. 한편 당시 수학교육계에서는 낱낱의 지식이나 능력보다는 한 영역에서의 지식을 다른 영역과 연결시키거나 통합할 수 있는 능력이 점점 중요시 되고 있었다. 격자들은 여러 내용 영역과 여러 수학적 능력을 연결시키고 통합적으로 보기 보다는 분리하여 개별적으로 다루도록 유도하는 경향이 있어, 수학교육계의 동향에 오히려 역행하는 한계를 또한 드러낸 것으로 보인다.

수학적 힘 차원의 출현, 격자들이 안고 있는 문제점과 한계를 고려하여 1996 평가들에서는 엄격한 격자들에서 다음 [그림Ⅱ-2]와 같이 바뀌었다. 이 그림에서 나타내고자 하는 포인트는, 문항을 개발할 때 내용 영역을 우선적으로 고려한 다음에, 수학적 능력과 수학적 힘은 부차적으로 고려한다는 의미를 표현한 것으로 볼 수 있으며, 이에 대해서는 요소별 특징에서 좀더 다루게 된다.



[그림 II-2] 1996 수학 평가틀

3) 요소별 특징

가) 내용 영역

1996 평가틀에서 내용 영역은 ‘수 감각, 특성 연산’, ‘측정’, ‘기하와 공간감각’, ‘자료 분석, 통계, 확률’, ‘대수와 함수’ 등으로 구분되었으며, 이전 평가틀에서 달라진 점은, ‘수와 연산’이 ‘수 감각, 특성, 연산’으로, ‘기하’가 ‘기하와 공간감각’으로 조정된 것이다. NCTM의 Standards에서는, 수에서 ‘수 감각’, 기하에서 ‘공간감각’에 관한 관심이 수업 내용과 방법 측면에서 증가하고 있음을 지적한 바 있는데(구광조·오병승·류희찬, 1992), 내용 영역을 조정할 때 Standards가 미친 영향을 알 수 있는 대목이다.

<표 II-5> 1996 내용 영역별 문항 비율

내용 영역	4단계	8단계	12단계
수 감각, 특성, 연산(최소/최대)	40/70	25/60	20/50
측정	20	15	15
기하와 공간 감각	15	20	20
자료 분석, 통계, 확률	10	15	20
대수와 함수	15	25	25

내용 영역 별 문항 출제 비율은 <표 II-5>와 같으며, 수치는 각 단계 별로 내용 영역에서 출제되어야 할 문항의 최소 비율이다. 당시에는 한 문항이 반드시 한 내용 영역에만 속하지는 않는다는 점을 감안하여 한 문항이 여러 내용 영역에 걸치는 것을 허용한 것으로 보인다. 이는 수학교육계에서 강조되던 연결성이나 통합성의 아이디어와도 일맥상통하는 측면이 있다. 여러 내용 영역들을 연결하여 푸는 문항이나 수학과 다른 학문 사이의 연결성을 평가하

는 문항들이 필요하며, 이 문항들은 사실상 여러 내용 영역에 속할 수 있다. 따라서 한 문항이 동시에 여러 내용 영역에 속할 수 있는 것이다. 한편, 문항의 최소 비율을 제시하였는데, 이는 어느 한 영역이 소홀히 다루어지지 않도록 하기 위해 취한 조치인 것으로 보인다.

한편 수 감각, 특성, 연산 영역에 한해서는 최소 비율과 함께 최대 비율을 정해 두었음을 알 수 있다. 1990년 수학과 평가들에서 내용 영역 차원의 특징으로 수와 연산 영역이 덜 강조되고 기하, 대수, 함수 등이 점점 비중 있게 다루어지는 경향이 있음을 지적한 바 있다. 수 감각, 특성, 연산 영역에만 최대 비율을 정해 두고 나머지 영역에는 아예 최대 비율을 정하지 않음으로써, 이전의 이러한 경향성을 유지하고자 한 것으로 보인다.

내용 영역에서 평가요소의 진술 방식은 다음 <표Ⅱ-6>과 같은 모습으로 바뀌었다. <표Ⅱ-3>과 <표Ⅱ-4>를 참조할 때, 내용 영역의 핵심인 지식과 수학적 능력의 요소가 합쳐진 형태로 진술한 것으로 보인다.

〈표Ⅱ-6〉 1996 평가들의 평가요소 진술 방식의 예

수 감각, 특성, 연산	단계		
	4	8	12
1. 세기, 모으기, 자리값을 연결 짓기			
a. 자연수와 소수를 모델화하고 기술하기 위해 자리값을 사용하기	•	•	•
b. 의미있는 맥락에서 기호를 사용하기	△	•	•

나) 수학적 힘-수학적 능력

수학적 힘은 Standards에서 평가규준의 하나로, 한마디로 수학적 지식을 통합적으로 사용할 수 있는 능력 전반을 뜻한다. 앞서 제시한 [그림Ⅱ-2]에서 보듯이, 수학적 힘의 추론, 의사소통, 연결성은 각각의 내용 영역과 수학적 능력의 기초를 이룬다. 이러한 수학적 힘을 평가한다는 것은 수학적 상황에서 ‘추론’할 수 있는 능력, 수학적 맥락에서 이끌어낸 결론을 ‘의사소통’할 수 있는 능력, 한 상황의 수학적 측면을 관련된 수학적 지식이나 정보와 ‘연결’시키는 능력 등을 평가하는 것으로 보았다.

1996 NAEP 수학 평가들의 특기할 만한 점은 수학적 힘이나 수학적 능력을 보고 범주로 사용하지 않고 다만 문항을 개발할 때에만 사용하였다는 점이다. 예컨대, 기하나 대수와 같은 내용 영역에 대해서는 각 내용 영역 별로 평가 결과에서 나온 학생들의 특징을 분석하여 보고서에 담지만, 수학적 힘이나 그 하위 요소들에 따라 학생들의 특징을 분석하는 작업은 하지 않았으며, 다만 초기에 문항을 개발할 때 평가지에 모든 수학의 내용적인 측면과 능력적인 측면을 골고루 갖춘, 이른바 균형감 있는 평가를 개발할 때에만 사용하도록 하였던 것이다. [그림Ⅱ-2]에서 수학적 능력을 후면에, 수학적 힘을 바닥에 위치시켜 표현한 것은 이러한 점을 나타내고자 한 것으로 보인다.

다) 문항군(Item Families)

1996 평가들에서 새롭게 등장한 문항군은, 학생들의 수학 학력의 폭과 깊이를 재기 위한 것으로, 학생들의 이해에 대한 종적 또는 횡적 특성을 보여 줄 수 있도록 구안된 평가 문항들의 모음이었다. 문항군은 특정 내용 영역에 대한 수학적 지식과, 여러 내용 영역을 걸쳐 개념, 원리, 절차 등을 다룰 수 있는 수학적 능력의 ‘깊이’를 재는 데 목적이 있었다.

다. 2005 수학 평가들

1) 배경

2005 NAEP 평가들에 따르면, 1990년대 초기만 하더라도 장애를 가진 학생이나 영어에 능숙하지 못한 학생들과 같이 특별한 배려를 필요로 하는 ‘특수 학생’의 경우 규정상 NAEP 평가에서 제외되었다. 그런데 이는 바람직하지 않다는 지적이 사회 각계각층에서 지속적으로 제기되면서 이전부터 특수 학생들을 포함시킬 수 있도록 절차와 시행 과정을 부분적으로 개선해 오고 있었다. 예컨대, 1996년 평가에서는 처음으로 특수 학생의 평가를 위한 다양한 편의 시설이나 장치들이 제공되었다. 예를 들어 소아마비인 학생이 자신의 실력을 제대로 발휘하도록 대신 필기해 줄 사람을 제공해 주며, 시험을 치루는 데 보다 많은 시간을 사용할 수 있도록 배려하는 것이다. 2000년 평가에서는 평가를 위한 편의 시설이나 장치를 제공 받은 학생들과 그렇지 않은 학생들을 대상으로 한 수학 평가 결과가 보고 되었다. 이와 같이 특수 학생들을 배려하는 방향으로 NAEP이 발전되면서 문항을 개발하거나 평가를 설계하는 과정이 좀더 면밀히 분석될 필요가 생겼으며, 2005 평가들은 이러한 필요성이 구체화되는 단계에서 기획된 것이다.

새로운 평가들을 개발하기에 앞서 NAGB에서는 공청회를 열어 새로운 평가들의 기본 목적을 위한 의견들을 모았으며, 2000년 9월에 CCSSO에 위탁하여 1996년에 사용된 평가들을 개선하도록 하였다. CCSSO는 25명으로 구성된 운영위원회와 19명으로 구성된 기획위원회를 구성하였다. 위원회의 구성원들이 합의를 이끌어내는 과정에서 참조한 내용들로는, 수학교육에서 현재 논의 되고 있는 내용, 주 수준의 수학과 교육 내용과 평가들, NCTM의 새로운 평가들, TIMSS의 결과물, Achieve Project에서 발간한 보고서, National Research Council of the National Academy of Sciences의 Adding it up이라는 최근 보고서 등이다. 한편 대학 교수, 주 평가전문가, 사설 연구 기관의 측정 전문가 등으로 기술자문 협의단(Technical Advisory Panel) 구성하여 새롭게 만들어진 평가들이 기존의 NAEP의 틀과 크게 상충되지 않는지, 실효성은 있는지 여부를 검토하도록 하였다.

2) 주요 특징

새로운 평가들에서는 1990, 1996 평가들에서 사용되던 수학적 능력이나 수학적 힘은 찾아볼 수 없고, 대신 ‘문항의 수학적 복잡도(mathematical complexity of items)’라는 항목이 그 자리를 채우고 있다. 이는 2005 수학과 평가들에서 가장 주목할 만한 변화로, 이후에 자세히 살펴볼 것이다.

3) 요소별 특징

가) 내용 영역

이전 평가들에서 바뀐 것은 먼저 1996년 수학 평가들에서 ‘수 감각, 특성, 연산’으로 명명되던 영역에서 ‘수 특성과 연산’으로 조정되었다. 또한 2005 평가들의 설명에 따르면, 계산 능력과 같은 경우, 수와 연산 영역에서 중요한 요소이기는 하지만, 직사각형의 둘레를 계산하기, 건물의 높이 구하기, 직각삼각형의 빗변의 길이 구하기 등과 같이 측정이나 기하에서도 다루어질 수 있기 때문에 별도의 내용 영역으로 설정해 놓지 않아도 평가 가능한 것으로 보았다고 한다.

〈표 II-7〉 2005 내용 영역별 문항 비율

내용 영역	4단계	8단계	12단계
수 특성과 연산	40	20	10
측정	20	15	30
기하	15	20	
자료 분석과 확률	10	15	25
대수	15	30	35

1996년 평가들과 비교해 볼 때, 8단계에서는 ‘대수와 함수’ 영역의 명칭이 ‘대수’로 바뀌고 문항 출제 비율은 25%에서 30%로 증가하고, ‘수 특성과 연산’ 영역의 문항 비율이 25%에서 20%로 감소하였다. NAEP 평가들의 설명에 따르면, 8단계에서는 유리수의 특성과 연산을 사용하거나 수 이론에 근거한 좀더 상위 단계의 학습을 하면서 새로운 수 개념이 생겨나는데, 이러한 수 개념의 상당 부분은 다른 내용 영역, 이를테면 측정, 자료 분석, 확률 등에서도 학습할 수 있고, 또한 이 단계에서는 비형식적인 대수 개념과 기하가 강조되어야 하기 때문에 대수는 약간 문항 비율이 올라가고 수 특성과 연산은 그에 대응하는 만큼 문항 비율이 감소되었던 것이다(NAGB, 2005).

12단계의 경우, 이전 평가들과 비교하여 수 특성과 연산 영역의 문항 비율은 20%에서 10%로 줄고, 대수 영역은 25%에서 35%로 늘어났으며, 자료 분석과 확률 영역의 문항 출제 비율이 20%에서 25%로 늘어났다. 4, 8단계와 비교해 보면 수 특성과 연산 영역의 비율이 급

격히 줄고, 대수, 자료 분석, 통계 영역이 대폭 늘었음을 확인할 수 있다. 2000 NAEP 평가 결과에 따르면, 12학년 학생들 중에서 79%가 대수의 2년 과정, 기하의 1년 과정 상당의 수업을 이수하는 등 고등학교에서는 점점 더 많은 학생들이 좀더 높은 수준의 수학 과정을 이수하고 있다고 한다. NAEP에서는 수학의 내용 영역을 전반적으로 평가할 필요가 있기 때문에 12학년의 문항 비율을 별도로 조정하였으며, 그 결과 기하/측정과 대수, 자료 분석과 확률에 강조점을 두었다고 한다. 한편 12학년에서는 측정과 기하를 하나로 묶어 총 네 개의 내용 영역으로 구분하였다. 이는 12학년에서 다루는 측정 주제가 기하적인 성격을 가지고 있다는 점을 고려한 것으로, 이 두 내용 영역을 구분한다는 것은 부자연스러우며 불필요하다고 보았기 때문이다.

각 내용 영역 내에서 평가요소를 기술하는 방식은 다음 <표Ⅱ-8>과 같이 바뀌었다.

<표Ⅱ-8> 2005 평가들의 평가요소 진술 방식의 예

수 감각		
4단계	8단계	12단계
a) 자리값과 자연수의 실제 자리값을 알기	a) 모델화하기 위해 자리값을 사용하여 정수와 소수를 기술하기	
b) 10진법, 수직선, 2차원 모델과 같은, 여러 가지 모델을 사용하여 수를 나타내기	b) 수직선과 다이어그램을 사용하여 유리수나 수치적 관계를 모델화하거나 기술하기	
c) -생략-		
d) 자연수를 쓰거나 다시 쓰기(예, 10: $5+5$, $12-2$, 2×5)	d) 유리수를 쓰거나 다시 쓰기	d) 실수를 쓰기, 다시 쓰기, 표현하기, 비교하기(예, 수직선, 모델, 다이어그램 등을 사용하여 수치적 관계, π , $\sqrt{2}$)

1990년 지식 중심의 나열식에서 1996년 지식과 수학적 능력이 결합된 형태의 진술로 바뀌었으며, 이번에는 관련된 평가 요소들을 단계별로 연결짓고, 단계에 따라 그 깊이나 폭이 달라지는 것이 드러나도록 진술하고 있다. 이는 1996년 평가들에서 새롭게 등장한 ‘문항군’의 아이디어, 즉 학생들의 종적 특성의 변화를 알아볼 수 있도록 학년급 간에 관련된 문항을 출제하는 아이디어가 평가요소 진술 방식에 반영된 것으로 보인다.

나) 문항의 수학적 복잡도

1996 수학과 평가들에서는 수학적 힘과 수학적 능력 차원이 사용되었으며, 이들은 보고 범주로 사용되지 않고 문항 개발 단계에서 수학을 하는 다양한 능력을 평가할 수 있는 문항들을 골고루 개발하기 위한 틀로 사용되었다. 그런데, 2005 평가들에서는 이 두 차원 대신에 ‘문항의 수학적 복잡도(이하 복잡도)’ 개념이 새롭게 등장하였다. 이 개념은 문항이 학생에게 요구하는 수학적 복잡함의 정도를 기술한 것이다. 이 평가들에서 제시한 다음의 <예1>과 <예2>를 보면, 전자가 문제를 푸는 학생에게 요구하는 것은 사실, 용어, 특성 등을 회상해 내는 것으로 복잡도가 낮으며, 후자의 경우는 간단한 논쟁을 해석하는 것으로 복잡도가 중간 정도인 것이다.

< 예1 >

다음 중 직사각형의 성질이 아닌 것은?

- A. 대변의 길이가 같다.
- B. 대변이 서로 평행하다.
- C. 모든 각의 크기가 같다.
- D. 모든 변의 길이가 같다.
- E. 두 대각선의 길이가 같다.

< 예2 >

순이 : “나는 6하고 다른 수를 곱해서 6보다 작은 수를 만들 수 있어.”

철수 : “그건 말도 안돼. 6하고 다른 수를 곱하면 그 값은 항상 6이거나 6보다 크단 말이야.”

누구의 말이 옳은가? 그렇게 생각한 이유를 말하여라.

평가들에 복잡도 개념이 등장하게 된 배경에는 평가들은 바로 ‘평가문항’의 개발을 위한 틀이라는 인식이 크게 작용한 것으로 보인다. 평가들을 통해 최종적으로 개발하려는 것은 평가에 적절한 문항이다. 문항은 저마다 난이도, 다루고 있는 수학적 내용의 깊이, 요구하는 수학적 능력 등이 다를 수 있으며, 이러한 다름에 기초하여 문항들을 구분할 수 있을 것이며, 이를 복잡도가 다른 것으로 구분하는 것이다.

이전의 평가들에서 수학적 능력이나 수학적 힘을 평가들의 한 축으로 고려할 때에는, 문항의 특성보다는 학생들의 사고를 대상으로 하였다고 볼 수 있다. 먼저 학생들의 사고를 개념적 이해, 절차적 지식, 문제 해결로 구분하고 그 후에 각각을 평가할 수 있는 문항을 개발하려고 하였다. 이에 대해 2005 평가들에서는 복잡도를 통해 문항의 특징을 먼저 고려하고자 한 것이다. 2005 평가들의 설명에 따르면, 그렇다고 해서 수학적 힘이나 수학적 능력의 차원의 아이디어가 완전히 배제된 것은 아니며, 그러한 아이디어들이 그대로 녹아 복잡도라는 새로운 차원으로 변형되고 개선했던 것이다.

복잡도에는 수학적 힘이나 수학적 능력의 하위 요소들이 포함되며 그 하위 요소들의 정도나 깊이에 따라 복잡도의 수준이 달라져, 높은 복잡도, 중간 복잡도, 낮은 복잡도로 구분되며, 각각의 특징은 다음 <표Ⅱ-9>와 같다.

〈표 II-9〉 2005 평가들의 문항의 수학적 복잡도

복잡도의 수준	특징
낮은 복잡도	<ul style="list-style-type: none"> - 사실, 용어, 특성 등을 회상하기 - 개념의 예를 인식하기 - 합, 차, 곱, 몫 등을 계산하기 - 동치인 표현을 인식하기 - 정해진 절차를 밟기 (이하 생략)
중간 복잡도	<ul style="list-style-type: none"> - 한 가지 방식 이상으로 수학적으로 상황을 표현하기 - 상황과 의도에 따라 다양한 표현을 고르고 사용하기 - 도형이나 명제들을 비교하기 - 풀이 과정에서 단계의 정당성을 제공하기 - 시각적 표현을 해석하기 - 패턴을 확장하기 (이하 생략)
높은 복잡도	<ul style="list-style-type: none"> - 상이한 표현이 상이한 목적을 위해 어떻게 사용될 수 있는가를 기술하기 - 다단계와 다양한 결정점을 지닌 절차를 수행하기 - 여러 가지 절차와 개념들 사이의 유사성과 차이점을 분석하기 - 패턴을 일반화하기 - 주어진 상황에 원래의 문제를 일반화하기 - 참신한 문제를 풀기 (이하 생략)

문항 출제 비율은 각 문항의 복잡도와 시간을 고려하여 정하였다. 낮은 복잡도 문항을 푸는 데에는 시간이 조금 걸리고, 높은 복잡도 문항은 시간이 많이 걸릴 것이다. 이런 점을 감안하여, 이상적인 문항 출제 비율로, 총점 중 절반은 중간 복잡도인 문항으로 하고, 나머지 절반은 낮은 복잡도와 높은 복잡도의 문항이 같은 비율이 되도록 하며 이 수치는 모든 학년급에 똑같이 적용하였다. 이와 같이 문항 출제 비율을 문제 풀이는 드는 시간에 따라 정한 것은 이전과는 사뭇 다른 점이다.

Ⅲ. 국내 성취도 평가 수학 평가들에 대한 시사점

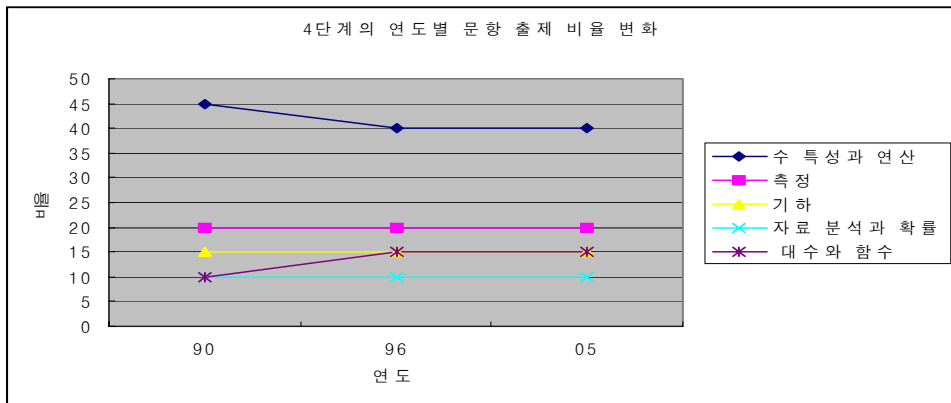
1973년부터 시작된 NAEP의 수학 평가는 30여년이 넘는 역사를 지녔으며, 국내 성취도 평가에서 수학 평가는 이제 5년째에 접어들었다. NAEP과 국내 성취도 평가의 목적은 유사하기 때문에, 이 목적을 달성할 수 있도록 정립되어온 NAEP 수학 평가들의 변천에서 역사가 짧은 국내 성취도 평가의 수학 평가들과 관련된 시사점을 추출해 낼 수 있었다. 이 장에서는 이를 두 가지 측면에서 살펴보고자 하며, 먼저 미시적으로 수학 평가들의 요소와 관련된 시사점을 살펴보고, 다음으로 수학 평가를 개발과 관련된 일반적인 시사점을 살펴보았다.

1. 수학 평가들의 요소별 시사점

앞에서 살펴본 바에 따르면, NAEP 수학 평가들의 ‘주요 특징’의 변화는 90년 내용 영역과 수학적 능력의 격자틀, 96년 수학적 힘을 바탕으로 한 내용 영역과 수학적 능력의 성긴틀, 05년 수학적 힘이나 수학적 능력을 발전시킨 문항의 수학적 복잡도 개념 등장으로 요약되며, 이러한 변화의 이면에는 한편으로는 수학교육과정의 변화 또는 개혁 방향을 반영하려는 것과, 다른 한편으로는 평가문항 개발이 좀더 성취도 평가에 부합되도록 평가들을 개정하려는 의도가 있었다. 이 과정에서 흥미로운 사실은, 이러한 세 번에 걸친 변화의 포커스는 수학적 능력, 수학적 힘, 그리고 이 두 가지가 ‘승화’되어 만들어진 문항의 수학적 복잡도와 같이, 내용 영역과 대구를 이루는 행동 영역이었다는 점이다.

국내 성취도 평가의 수학 평가들의 경우, 내용 영역은 그동안 상당히 일관성 있게 적용되었다. 이는 국가적으로 정해 놓은 국가교육과정에 근거하여 성취도 평가의 내용 영역의 틀을 마련하였는데 그동안 국가교육과정의 수학과 틀에 큰 변동이 없었기 때문이라고 볼 수 있다. 반면에 내용 영역과 대구를 이루는 행동 영역의 경우는 사정이 조금 다르다. 국내 성취도 평가의 경우, 99년 기초 연구에서 그러한 측면에 대해 구체적으로 정의하고, 내용 영역과 [인지적] 행동영역으로 이루어진 격자틀에 맞추어 평가도구를 개발하는 아이디어를 제시한 바 있다. 그 후로 행동 영역은 어떤 해에는 평가도구 개발에서 소극적으로, 즉 이미 개발된 문항들이 어떤 행동 영역에 속하는지를 기록하는 정도로 사용되었으며, 또다른 해에는 아예 사용되지 않기도 하였다. 행동 영역을 국내 성취도 평가의 수학 평가들에서 어떻게 할 것인지에 대한 심도 있는 연구가 필요하다고 생각된다.

NAEP 수학 평가들에서는 90, 96, 05년에 따라 각 단계별로 내용 영역의 비중이 달라졌다 ([그림Ⅲ-1] 참조). 4단계의 경우, 수와 연산의 비중이 절대적으로 크지만, 점차 감소하는 추세를 보이고 있으며, 자료 분석, 통계, 확률 영역은 절대적으로 낮은 비중이지만 점차 증가하는 추세를 보였다. 8단계의 경우, 수와 연산이 감소하고, 대수와 함수가 증가하였음을 알 수 있다. 12단계의 경우, 절대적으로 비중이 적었던 수와 연산이 큰 폭으로 감소하는 추세를 보였으며, 대수와 함수, 기하, 확률 통계 등은 증가하였다. 이와 같이 NAEP 수학 평가에서는 내용 영역에서 강조하려는 바가 달라짐에 따라 문항 출제 비율이 달라졌으며, 그러면서도 그러한 변화에서 경향을 읽을 수 있었다.



(그림III-1) NAEP 수학 평가들에서 출제 문항 비율의 변화 : 예시

지금까지 국내 성취도 평가의 수학 평가에서 내용 영역별 문항 출제 비율은 신경을 크게 쓰는 부분은 아니었다. 다만 평가영역의 비중을 고려하고 모든 평가영역에서 골고루 출제될 수 있도록 평가문항을 개발한다는 정도의 지침을 암묵적으로 따르는 정도였다고 볼 수 있다. 또한 매년 평가에 출제되는 문항 수는 학년급 별로 대략 35~50문항 정도로 적은 편이기 때문에 백분율은 그다지 의미가 없다고 볼 수도 있다. 그런데, 성취도 평가에서 좀더 타당하고 의미 있는 분석 결과를 이끌어내려면 학년급에 따라 각 평가 영역에 어느 정도 비중을 두어야 하는지에 관하여 경향성 정도 수준에서라도 정하는 것이 필요하다고 본다. 특히 안정적인 추이분석을 위해서는 그러한 작업이 더욱 필요할 것으로 생각된다.

NAEP에서 평가들의 평가요소의 진술 방식이 바뀌는 과정을 <표II-3>, <표II-6>, <표II-8>을 통하여 살펴보았다. 최근 2005년의 경우에는 동일군에 속할 수 있는 지식이 학년급에 따라 그 폭과 깊이가 달라지는 것을 드러낼 수 있도록 구성한 것으로 보인다. 우리나라에서 평가요소의 진술 방식에 대응되는 것은 ‘성취기준, 평가기준’이라고 볼 수 있다. 성취기준은 ‘교과 교육의 목표에 비추어 학생들이 알아야 할 것과 할 수 있는 것의 범위와 깊이를 구체적으로 제시한 것’이며, 평가기준은 ‘교과별 평가활동에서 실질적인 기준 역할을 할 수 있도록 학생들이 성취한 정도를 몇 개의 수준(예 : 우수, 보통, 기초)으로 구체화한 것’이다. 다음은 성취기준과 평가기준의 한 예로, 각 학년급과 평가영역 별로 출제하고자 하는 성취기준과 평가기준을 선정하여 빗금 친 부분에 해당하는 문항을 개발하며, 이 때 특히 평가기준의 상중하를 난이도의 상중하와 일치시키도록 하였다. 현재 성취도 평가에서는 초중고 학년급 간의 관련성까지 나타나도록 성취기준, 평가기준을 제시하고 있지 않다. NAEP의 평가들의 평가요소 진술 방식에서 2005년의 아이디어는 국내 성취도 평가에서 성취기준을 재조정할 때 적극적으로 고려해 볼 만하다고 여겨진다. 학교에서 가르치는 수학 지식 중에는 학년급

에 따라 폭과 깊이가 달라지는 경우들이 있으므로, 그 경우에는 그러한 연관을 포착할 수 있도록 문항을 내어 학생들이 종적으로 어떠한 변화를 나타내는지를 분석하는 것은 교수·학습 측면에서 의미 있는 작업일 것이다.

〈표Ⅲ-1〉 국내 성취도 평가의 수학 성취기준 평가기준 예시

규칙성과 함수	17. 두 수량 사이의 비와 비율의 의미를 이해한다.	상	생활장면에서 두 수량 사이의 비의 값을 분수와 소수로 나타낼 수 있다.
		중	두 수량 사이의 관계를 비로 나타낼 수 있다.
		하	두 수량 사이의 관계를 나타낸 비에서 기준량과 비교하는 양을 구별할 수 있다.

2. 수학 평가틀 개발과 관련된 일반적 시사점

이 절에서는 지금까지 살펴본 NAEP 수학 평가틀의 변천 과정에서 얻을 수 있는 성취도 평가의 수학 평가틀 개발에 관련된 일반적 시사점들을 정리하였다.

첫째, NAEP 수학과 평가틀의 개발이 다양한 사람들이 참여한 가운데 그들 사이의 합의를 바탕으로 하여 이루어진다는 점을 확인할 수 있었다. 예컨대, NAGB의 구성원은 정책가, 법률가, 학교장, 교육학자 등 각계각층의 인사 26명으로 구성되었다. 평가틀 개발 작업을 위임 받은 CCSO나 College Board에서 구성한 NAPP나 운영위원회는 다양한 직업군의 사람들로 적게는 12명, 많게는 25명으로 이루어졌다. 이들의 논의를 통하여 NAEP이 나아가야 할 방향과 범위가 정해졌다. 또한 MOC나 기획위원회에서는 교사, 학교장, 수학교육전문가 등을 중심으로 하여 적게는 7명, 많게는 19명이 참여하여 본격적으로 수학 평가틀을 개발하였다. 어떤 일을 ‘국가 수준’에서 수행한다고 할 때에는, 그 일의 규모가 전국적이라는 의미도 있겠지만, 동시에 각계각층의 의견을 최대한 많이 수렴한다는 의미로도 보아야 할 것이다.

둘째, 평가틀을 개정할 때, 이와 같이 다양한 사람들의 의견을 수렴하면서 동시에 수학교육의 연구 결과나 주장들을 적극적으로 반영한다는 것을 알 수 있었다. 1990, 1996 평가틀을 개정할 때 NCTM에서 내 놓은 Standards를 적극적으로 수용한 점을 들 수 있다. 1996 평가틀의 내용 영역에서 ‘수 감각’과 ‘공간 감각’이나, 1996, 2005 평가틀에서 주요 논의 대상이었던 수학적 힘, 수학적 능력, 추론, 의사소통, 연결성 등은 Standards에서 새롭게 주장된 내용들이었다.

셋째, NAEP 수학 평가틀과 관련된 보고서 내용은 개정 배경이나 특징을 파악하고 기타 궁금증을 해소할 수 있을 정도로 상당히 상세하게 기술되어 있었다. 지금까지 NAEP 수학 평가가 진행되어 온 배경, 새로운 수학 평가틀에서 주요 이슈가 되고 있는 사항, 주안점 등에 관하여 상술하고 있는 것이다. 이렇게 상술된 보고서를 보면서 평가문항의 개발 과정을 충분히 이해할 수 있다. 국가 수준에서 수행되는 성취도 평가가 제대로 진행되기 위해서는

관련자들의 이해와 양해가 필수적이다. 평가들에 관해 상술해 줌으로써 그러한 효과를 거두고 있다고 보여진다.

넷째, 앞서 말한 대로 상술한 보고서를 통해 한편으로 평가들에서 약간씩 변동이 있음을 알 수 있지만, 다른 한편으로는 주요 뼈대들이 일정하게 유지되고 있어 일관성을 찾을 수 있었다. 우선 평가들에서 다루는 내용으로, 배경, 내용 영역, 행동 영역(수학적 힘, 수학적 능력, 문항의 수학적 복잡도), 문항 유형, 문항 비율 등이 매 보고서에서 중심이 된다. 내용 영역에서도 약간의 변동은 있지만, 5가지 영역 구분이 유지되고 있으며, 문항 비율과 같은 경우, 비율이 달라지기는 하지만, 그 경향성에서는 일관성을 볼 수 있다. 국가적 수준에서 벌이는 성취도 평가는 일종의 제도(制度)라고 볼 수 있으며, 일관성은 제도가 갖추어야 할 중요한 조건이다. 제도에 일관성이 결여될 경우, 연구자 스스로가 연구에 의미를 부여하기 힘들어지며, 그러한 제도에서 나온 결과에 대한 신빙성을 확보하기 힘들어질 수 있다. 이런 점을 감안할 때, 평가들의 주요 뼈대들의 일관성을 유지하는 것은 중요하다고 보여진다.

다섯째, 주요 뼈대들은 일관성 있게 유지하면서도 또한 매 평가들마다 변동되는 내용이 있었다. 이렇게 변동되는 내용에 대해 그 근거나 정황을 상술하고 있는데, 이를 통하여 NAEP의 평가들이 정교해지고, 세련되어지는 방향으로 발전되고 있다는 것을 확인할 수 있었다. 이를테면, 평가요소를 진술하는 방식은, 지식 중심의 짙막한 진술(1990)에서 지식과 수학적 능력이 결합된 형태의 진술(1996)으로, 이는 다시 수학적 지식과 능력을 깊이와 폭을 알 수 있도록 학년급에 따라 진술(2005)하는 방식으로 변화였다. 또 다른 예로, 수학 내용 영역과 수학적 능력으로 이루어진 격자들에서, 수학 내용 영역과 수학적 능력, 수학적 힘으로 이루어진 성긴틀로 바뀌었고, 이를 다시 수학 내용 영역과 문항의 수학적 복잡도로 바뀌었다. 그 과정에서 보여준 설명들을 통해 수학 평가들이, 성취도 평가에 적절하고 균형감을 평가문항을 출제할 수 있도록 발전하고 있음을 확인할 수 있었다. 이렇게 평가들이 발전할 수 있는 것은, 앞서 두 번째에서 언급한 수학교육의 연구를 적극적으로 수용하려고 하는 노력과, 또한 평가 과정 자체를 되짚어 보는 절차를 통하여 가능하였을 것으로 생각된다.

여섯째, 수학 성취도 평가들을 개발하는 것은 일종의 디자인하는 것이라는 마인드를 가질 필요가 있다고 본다. 디자인은 주어진 어떤 목적을 달성하기 위하여 여러 조형요소 가운데서 의도적으로 선택하여 그것을 합리적으로 구성하여 유기적인 통일을 얻기 위한 창조활동이다. 성취도 평가에서는 학생들의 수학 학력을 있는 그대로 드러내는 것이 가장 기본적으로 해야 할 일로, 그러기 위해서 정교한 평가들의 개발이 필수적이다. 예컨대, NAEP의 경우 수학 학력을 1990년 평가에서는 내용 영역과 수학적 능력으로, 1996 평가에서는 내용 영역과 수학적 능력-수학적 힘으로 구분하였다가, 2005 평가에서는 내용 영역과 문항의 수학적 복잡도로 구분하는 것으로 바뀌어 왔다. 문항군 개념과 같은 경우도 학생들의 능력을 심층분석할 수 있을 정도로 드러내기 위해 등장한 아이디어였다. 2005 평가들에서 시간을 기준으로

문항 출제 비율을 정하는 것 역시 마찬가지로 맥락에서 볼 수 있다. 이러한 변화 속에 관통하고 있는 원칙은 ‘균형감’ 있는 평가도구 또는 평가지를 개발하는 것이라고 볼 수 있다. 정교한 평가틀이란, 달리 말하면 학력의 특징을 균형감 있게 드러낼 수 있는 평가지를 만들 수 있는 틀이다. 성취도 평가라는 평가 상황에서 학생들의 학력을 제대로 드러내는 방법을 찾는 실험과 검증 작업을 반복하는 과정에서 평가틀의 그러한 변화가 있었던 것이다. 이런 사례들로부터, 평가틀을 개발하는 연구자들에게는 디자인 개념이 필요할 것으로 생각된다. 수학 평가틀을 개발한다는 것은 학생들의 수학 학력을 드러낸다는 목적을 달성하기 위해 내용영역, 행동영역, 문항 비율, 문항 유형 등을 합리적으로 구성하는 활동으로, 평가틀을 어떻게 디자인하는가에 따라 그 결과가 다를 수 있다는 점을 인식하고 섬세하게 기획하려는 자세가 필요하다는 것이다.

마지막으로, 성취도 평가의 수학 평가틀을 개발할 때 우리나라 교육에 대한 이해와 향후 발전 방향 등을 고려하여 넓은 안목으로 접근해야 한다는 점을 지적할 수 있을 것이다. 미국에서는 1990 평가에서 여러 주의 성취도를 상호 비교하기 위하여 주 수준의 평가를 시행할 수 있도록 관련 법률을 통과시켰으며, 2005 평가에서 장애아나 영어에 능숙하지 못한 특수 학생들을 배려하는 것에 중점을 두었다. 이렇게 국가적으로 추구하는 교육의 모습에 변화가 생기게 되면서, 그에 따라 수학의 평가틀에 관한 재검토와 개정 작업이 이루어지게 되었다. 개인적으로 또는 소규모로 진행되는 성취도 연구, 특히 수학에 관한 성취도 연구에서는 반드시 국가적으로 추구하는 교육의 모습에 관심을 기울여야 할 강제성은 없다. 하지만 국가 수준에서 수행되는 수학 성취도 평가에서는 사정이 다를 수 있다고 보여진다. 수학 평가틀이므로 ‘수학’ 또는 ‘수학교육’에 당연히 무게중심을 두어야겠지만, 국가적으로 우리나라 교육의 현재의 쟁점과 향후 발전 방향에 대해 항상 예의주시할 필요가 있다.

IV. 맺음말

평가원에서 수행하는 국가수준의 학업성취도 평가 연구에서는 우리나라 학생들의 학력에 관한 좀더 타당하고 유의미한 정보들을 보고하기 위하여 평가 시행의 각 단계에 내실을 기하는 노력을 다각도로 벌이고 있다. 이러한 노력의 연장선으로 이 논문은 평가 시행의 첫 단계인 [(수학) 평가틀 개발]에 초점을 두고 수학 평가틀에 좀더 내실을 기할 수 있는 아이디어들을 NAEP 수학 평가틀에서 얻고자 하였다.

NAEP은 시간적으로 30여년이 넘게 시행되어온 연구로, 그동안 NAEP의 수학 평가틀에는 학교장, 교사, 교수 등 수학교육관련 종사자나 그 밖의 여러 계층의 의견이 반영되어 왔다.

학생들의 학업 성취 정도를 진단하고 학력 추이 변화를 파악한다는 점에서 NAEP과 국내 성취도 평가의 목적은 매우 유사하기 때문에 NAEP 수학 평가들의 변천 과정으로부터 국내 성취도 평가의 수학 평가들에서 앞으로 특별히 보완해야 할 점이 무엇이며, 평가들 개발은 어떤 식으로 이루어져야 하는지에 관한 시사점을 이끌어낼 수 있었다.

이러한 시사점 중에서 수학 평가들의 구성요소와 관련한 시사점으로는 첫째, 내용 영역과 대비되는 행동 영역—NAEP의 수학적 능력, 수학적 힘, 수학적 복잡도와 관련됨—을 평가들에서 어떻게 다룰 것인지에 대한 체계적이고 심도 있는 논의가 필요할 것으로 보인다. 둘째, 출제 문항 수와 관련하여, 학생들의 학력의 추이 분석이 좀더 안정적으로 이루어지고, 평가 영역 별로 교수·학습이나 교육과정 측면에서 깊이 있는 분석을 할 수 있도록 적절한 문항 수나 문항 출제 비율을 정할 필요가 있는데, 이를 위한 기초 연구가 필요한 것으로 보인다. 마지막으로, 결과 분석을 지금보다 심도 있게 할 수 있도록 성취기준·평가기준을 초·중고 학년급 간에 내용 영역의 요소들 사이의 연관성이 드러날 수 있도록 개발하는 것도 유익할 것으로 생각된다.

성취도 수학 평가와 관련하여 다음과 같은 일반적인 시사점을 얻을 수 있었다. 먼저 각계 각층의 다양한 사람들의 참여 속에 성취도 평가들에 관한 합의를 이끌어낸다는 점이다. 둘째, 이와 같이 다양한 사람들의 의견을 수렴하면서 동시에 수학교육의 선도적인 연구 결과들을 적극적으로 평가들에 반영한다는 것을 알 수 있었다. 한편 이와 같이 다양한 사람들의 의견을 수렴할 수 있는 것은 NAEP를 주관하는 NAGB에 국회에서 부여한 법적 권한이 있기 때문일 것이다. 셋째, 매 개정 때마다 정황과 근거를 상세하게 담은 평가들 보고서를 발간함으로써 관련자들의 이해를 도모하는 효과를 거둔다는 점이다. 넷째, 이러한 보고서 내용에서 개정 때마다 평가들이 약간씩 변동하였다는 점을 알 수 있지만, 그 속에서 나름의 일관성을 찾을 수 있었다. 성취도 평가 연구를 제도라는 관점에서 볼 때 이는 제 삼자로 하여금 그 결과에 대해 신뢰할 수 있는 기반을 제공해 준다고 볼 수 있다. 다섯째, 평가들이 바뀌면서 변동되는 부분이 있는데, 그러한 변동이 단순한 ‘바뀜’이 아니라 ‘발전’이 되어야 한다는 것이다. 이를 위해서는 이전 평가들에 대한 반성작업이 수반되어야 할 것이다. 여섯째, 평가들을 개정하는 것은 디자인하는 것이라는 마인드를 가질 필요가 있다. 학생들의 학력의 면면을 제대로, 있는 그대로 드러내기 위해 어떠한 틀을 사용할 것인지에 대해 실험 정신과 창조적인 역량을 발휘할 필요가 있다. 마지막으로, 수학 평가들이기 때문에 수학에 무게중심을 두어야겠지만 국가에서 지향하는 교육의 방향과 같이 보다 넓은 안목에서 수학 평가들을 바라보아야 한다는 점이다.

교육과정에서 다룬 내용을 학생들이 어느 정도 학습하였는가를 평가하는 작업이 제대로 이루어져 구체적인 분석 자료들이 축적되어 있다면, 이러한 자료들은 교육과정 개정 작업에서 매우 큰 도움이 될 것이다. 학생들의 학력이 떨어지고 있는가, 아니면 그 반대인가를 제

대로 평가하여 파악하고 있다면, 정부나 일선 학교에서는 이에 대응한 적절한 조치를 취할 수 있을 것이다. 이들이 가능하기 위해서는 국가수준의 학업성취도 평가의 각 단계가 내실 있게 진행되어야 하며, 그 첫 단계인 평가들의 내실을 다지는 작업은 그러한 측면에서 매우 중요하다.

국내 성취도 평가의 수학 평가들에 관한 시사점을 몇 가지 이끌어내기는 하였지만, 이 논문에서는 미국의 NAEP 수학 평가들의 변천 과정을 체계적으로 정리하는 데 초점을 두었다. 평가에 관한 연구가 경험적으로나 이론적으로나 발달의 여지가 많은 국내 수학교육계에서는 ‘평가들’ 자체에 대한 논의도 그다지 활발하지 않은 형편으로, 이 논문의 내용이 이후 평가들에 관련한 연구를 수행할 때 기초자료로써 충분히 가치가 있을 것으로 생각하였기 때문이다. 한편 NAEP 수학 평가들의 세 번에 걸친 개정 내용을 정리하는 데 치중하다 보니, 비판적으로 고찰하는 데 부족한 면이 없지 않으며, 국내 성취도 수학 평가들과 충분히 연계시키지 못하였다. 이 논문의 내용을 기초로 이에 대해서는 별도의 연구가 진행되어야 할 것이다.

참고 문헌

- 구광조 · 오병승 · 류희찬(1992). *수학교육과정과 평가의 새로운 방향*. 서울 : 경문사.
- 김명숙 · 황혜정 · 최승현 · 이명희 · 강운선 · 박선미 · 김재춘 · 박정 · 설현수(1999). **국가수준 교육성취도 평가 연구 II : 사회 · 수학 영역 예비 문항 개발 및 현장 적용 연구**. 한국교육과정평가원. 연구보고 RRE 99-9-1.
- 김명숙 · 노국향 · 박정 · 부재울 · 양길석(1998). **국가수준 교육성취도 평가 방안 연구**. 한국 교육과정평가원. RRE 98-8.
- 나귀수 · 한경혜 · 황혜정(2001). **2001년도 국가수준 교육성취도 평가 연구 : 수학**. 한국교육 과정평가원. 연구보고 RRE 2001-5-4.
- 나귀수 · 이봉주 · 한경혜(2002). **2002년 국가수준 학업성취도 평가 연구(I) : 수학**. 한국교육 과정평가원. 연구보고 RRE 2002-1-4.
- 이명희 · 백순근 · 황혜정 · 양길석 · 설현수 · 박정 · 문무경 · 이종원(2000). **2000년도 국가수준 교육성취도 평가연구(I) : 사회 · 수학 본검사 실시 및 결과 분석**. 한국교육과정평가 원. 연구보고 RRE 2000-9.
- Kenney, P. A. Silver, E. A. (1997). *Results from the Sixth Mathematics Assessment of the National Assessment of Educational Progress*. NCTM.
- Silver, E. A, Kenney, P. A. (2000). *Results from the Seventh Mathematics Assessment of the National Assessment of Educational Progress*. NCTM.
- NAEP(National Assessment of Educational Progress), (1988). *Mathematics Objectives*. 1990 Assessment.
- NAGB(National Assessment Governing Board), (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress*. [On-line] available <http://www.nagb.org>.
- NAGB(National Assessment Governing Board), (2001). *Mathematics Framework for the 2005 National Assessment of Educational Progress*. [On-line] available <http://www.nagb.org>.
- NAGB(National Assessment Governing Board), (1996). *Mathematics Framework for the 1996 National Assessment of Educational Progress*. NAEP Mathematics Consensus Project.
- NCES(National Center for Education Statistics), (1994). *Overview of NAEP Assessment Frameworks*.

• 논문접수 : 2004년 4월 15일 / 수정본 접수 : 2004년 5월 19일 / 게재 승인 : 2004년 6월 7일

ABSTRACT

A Study on Mathematics Frameworks of National Assessment Educational Progress in the United States

Young-Mi Cho

(Research Fellow, Korea Institute of Curriculum & Evaluation)

The aims of National Assessment of Educational Achievement(NAEA) are to affirm the accountability of school education and to scientifically manage and elevate the quality of education at the national level.

NAEA is conducted as following procedures : (1) Developing assessment framework, (2) construction assessment instruments, (3) Conducting a field test, (4) Revising or modifying assessment items based on the field test analysis, (5) Implementing the main assessment, (6) Scoring, (7) Analyzing data and presenting detailed substantive results.

Lately KICE makes every efforts to improve each above procedure for reporting more reliable and meaningful results of students' scholastic achievements. In this paper we focused the first procedure, developing [mathematics] assessment framework, and thought about how to improve it.

For this purpose, we examined the 1990, 1996, 2005 National Assessment of Educational Progress(NAEP) mathematics frameworks. NAEP has been implemented for more than 30 years and has the assessment aims which are similar to NAEA. So we could elicit suggestions for the mathematics framework of NAEA.

We suggested the necessity of the deep research about how to deal with mathematical ability, item distributions, and objectives in NAEA. And we recommended that for developing mathematics framework, we shall collect opinions of varied persons, make an effort for reflect results of mathematics educational research, and need to have a sense of design.

Key Words : NAEP, National Assessment of Educational Achievement, mathematics framework