

수학과 교육과정의 질 관리 측면에서 『국가수준 학업성취도 평가 연구』 내실화 방안 탐색

조 영 미*

이 논문에서는 한국교육과정평가원에서 시행하고 있는 『국가수준 학업성취도 평가 연구』를 교육 프로그램 평가 연구로 규정하고, 그 연구의 주요 목적 중에 하나인 ‘교육과정의 질 관리’ 목적을 위해 평가 방법과 결과 분석 방법에서 개선되어야 할 점을 탐색하였다. 이를 위해 교육과정의 질 관리라는 측면에서 국가수준 성취도 평가 연구의 현황을 네 가지로 구분하였으며, 각각에 대해 면밀히 분석하고 그에 바탕을 둔 내실화 방안을 다음과 같이 제시하였다. 첫째, 수학 교육 내용에 관한 학생들의 이해 정도 평가와 관련하여, 성취수준의 의미를 교과 차원에서 좀더 구조화하고 명료화하여야 한다. 둘째, 내용 영역 간 학생들의 이해 정도 비교 평가와 관련하여, 타당성을 높이기 위하여 출제 문항 수를 현재보다 늘리는 방안을 강구하여야 한다. 셋째, 성취수준별 수학 학력 특징에 관한 정보 수집과 관련하여, 평가문항 개발 단계에서부터 각 성취수준별 특징이 드러나도록 문항을 개발하고, 또한 그 방안을 구체화 하여야 한다. 넷째, 수학 교육 내용 적정화에 활용 가능한 기초 자료 수집과 관련하여, 교육과정의 질 관리만을 목적으로 한 문항군을 별도로 개발하고, 수행평가로부터 유의미한 심층 분석 정보를 얻을 수 있도록, 표집 대상 학생 전체가 아니라 일부만을 대상으로 수행평가를 실시하여야 한다.

1. 서 론

최근 교육평가의 동향을 살펴보면, 학생 개인의 학업 성적 평가 연구뿐만이 아니라, 교육현상 전반에 걸쳐 교육의 질적 수준을 확인하고 개선하는 데 목적을 둔 교육 프로그램 평가 연구에 대한 관심이 고조되고 있다(임인재, 1995; 황정규, 1998).

Romberg(1992)에 따르면, 교육 평가의 역사는 19세기 이전, 19세기, 20세기 중반, 20세기 중반 이후로 구분할 수 있다. ‘19세기’에 Thorndike와 그의 제자들에 의해 학생들의 학력

평가 검사가 본격적으로 시작되었고, ‘20세기 중반’까지는 그러한 학력 평가 결과로 얻은 학생들에 관한 정보가 주로 각 학생에 대한 평가의 근거로만 사용되었으며, ‘20세기 중반 이후’에서야 그러한 정보가 교육 정책이나 프로그램에 대한 평가의 근거로 사용되기 시작하였다. 그러한 교육평가의 역사에 비추어 볼 때, 교육 정책이나 프로그램에 대한 평가의 역사는 이제 반세기 정도 된 셈이다. 대표적인 예로 1963년부터 구체적인 추진 계획을 마련해 지금까지 진행되고 있는 미국의 National Assessment of Educational Progress(이하 NAEP), 1978년부터 시행된 영국의 Assessment of Performance Unit

* 공주교육대학교, ymcho@gjue.ac.kr

(이하 APU), 국제적 학력 평가 연구로 분류할 수 있는, IEA의 TIMSS나 OECD의 PISA 등을 들 수 있다.

국내에서 현재 시행되고 있는 여러 평가 중에서 위와 유사한 맥락을 지닌 평가 연구로는, 교육인적자원부와 16개 시·도 교육청의 지원 아래 한국교육과정평가원(이하 평가원)이 주도하고 있는 '국가수준 학업성취도 평가 연구(이하 국가수준 성취도 평가)¹⁾를 꼽을 수 있다. 이 연구에 대한 구체적인 시행 방안은 평가원 설립 원년인 1998년에 마련되었다. 한편 이 연구 시행 이전에 우리나라 교육정책이나 프로그램에 대한 평가 연구가 없었던 것은 아니다. 그동안 교육평가 학계를 중심으로 교육의 질적 관리를 위한 정보, 교육정책의 과학화와 개선, 교수·학습 방법의 개선, 교육의 책무성 고양 등을 위해 실증적(實證的) 자료의 수집이 필요하며, 그러한 목적에 부합하는 전국 단위의 학력평가가 필요하다는 주장이 꾸준히 제기되었으며, 실제로 그러한 내용을 목적으로 한 평가 연구들이 평가원 설립 이전에 실행되었다²⁾. 그런데 그들은 간헐적으로 실시되고, 연구들 간의 연계성이 낮고, 체계적이지 못하여, 그로부터 얻은 분석 결과들의 활용도가 매우 낮다는 근본적인 문제점을 안고 있었다. 따라서 평가원에서 국가수준 성취

도 평가의 시행 방안을 마련할 때에는 그와 같은 '근본적인 문제점'을 해결하는데 초점을 두고 있었다.

1998년 제안되었던 국가수준 성취도 평가 연구 방안들은, 10여년 가까이 실질적으로 평가가 시행되고 그 결과가 분석되는 과정에서 적지 않은 변화를 겪었다. 대표적인 변화로는, 2003년 이후의 점수 체제의 변화, 즉 100점 만점의 원점수 체제에서 최고 점수 체제로의 변화를 꼽을 수 있다. 이는 학생들의 학력이 연도별로 어떻게 변화하는가를 알아보기 위한 것으로, 2003년 국가수준 학업성취도 평가를 기준 연도로 하여 척도화와 검사동등화를 위한 방법 및 소프트웨어를 개발하여 적용하였다. 이러한 평가 방법상의 변화를 피함으로써 '연도별 학력 추이 분석'이라는, 국가수준 성취도 평가의 여러 가지 목적³⁾ 중에서 한 가지가 상당부분 달성 가능하게 되었다(정구향, 2004). 이와 같은 예를 보더라도, 국가수준 성취도 평가에서 추구하는 목적을 달성하기 위해서는 평가 방법이나 결과 분석 방법에 있어 적절한 변화나 발전이 수반되어야 한다. 그런 점에서 볼 때, 국가수준 성취도 평가에서 사용하는 평가 방법이나 결과 분석 방법 자체는 연구 문제가 될 자격이 충분하다.

국가수준 성취도 평가와 관련된 연구들은 크

-
- 1) 국가수준 성취도 평가의 평가 과목은 국어, 사회, 수학, 과학, 영어이며, 평가 대상 학년은 초등학교 6학년, 중학교 3학년, 고등학교 1학년으로, 각 학년을 대상으로 1%-3%의 학생들을 표집하여 평가를 실시한다. 7 월경에 예비검사를 실시하고, 이 결과를 토대로 평가문항을 확정하여 10월 말경에 본검사를 실시한다.
 - 2) 1959년, 1963년도 학력검사(중앙교육연구소), · 1968년도 기초학력조사(중앙교육연구소) · 1972년도 전국수준 학력평가(한국행동과학연구소) · 1974년도 전국 학력평가 연구(한국교육개발원) · 1977년도 한국 초중학교 학생의 특성과 학업성취 수준(한국교육개발원) · 1979년도 학력검사(한국교육개발원) · 1980년도 국민학교 학력평가(한국행동과학연구소) · 1988년도 중학교 1학년, 고등학교 2학년에 대한 전국학력평가(국립교육평가원) · 1993년도 국민학교 4학년, 중학교 3학년, 고등학교 2학년을 중심으로 한 주지교과에 대한 학업성취도 평가 연구(국립교육평가원) 등 다수
 - 3) 국가수준 성취도 평가의 주요 목적은 다음과 같다. ① 초·중·고등학교 학생의 학업성취도를 체계적이고 과학적으로 진단하여 학업성취도의 변화 추이 파악 ② 교과 교육과정에서 규정하고 있는 교육목표에 비추어 학생이 목표를 어느 정도 도달하였는지 분석하고, 교육과정의 문제점과 정착 정도를 파악하여 교육과정 개선에 기초가 되는 참고 자료 ③ 문항 분석, 성취도와 배경변인과의 관련성 분석을 통해 교수·학습 방법 개선 및 장학정책 수립을 위한 기초 자료 산출

게 두 종류로 구분할 수 있다. 첫 번째 연구 유형은 평가 결과를 다루거나 그것을 활용한 것으로, 그 예로는, 황혜정(2001)의 『2000년도 국가수준의 중·고등학교 수학과 교육성취도 평가 연구』와 조영미(2005)의 『우리나라 초등학교 6학년 학생들의 수학 성취수준별 특징 탐색』을 들 수 있다. 두 번째 연구 유형은 평가의 목적이나 방법 등의 개선을 목적으로 한 것으로, 김명숙(2003)의 『국가수준 학업성취도 평가 연구의 주요 쟁점과 발전 방향』과 김홍원(1998)의 『사회과 교육의 질 관리를 위한 국가 수준에서의 사회과 학업성취도 평가 방안 개선 연구』를 들 수 있다. 이 논문은 국가수준 성취도 평가의 또 다른 주요 목적인 ‘교육과정의 질 관리’를 위해서는 어떤 점이 개선되어야 하는지를 다루고 있기 때문에 두 번째 연구 유형에 속한다.

국가수준 성취도 평가 연구는 평가 전공자들을 위주로 한 ‘총론’과 각 교과교육 전공자들을 위주로 한 ‘각론’의 이중 구조를 가지고 있다. 총론에서 평가 문항 개발 지침을 제공하고 전문적인 평가 기법을 적용하여 평가 결과를 산출한다면, 각론에서는 교과의 특성을 적극 반영하여 평가 문항을 개발하고 산출된 평가 결과를 분석·해석한다. 국가수준 성취도 평가에서 추구하는 여러 목적들을 달성하는 데 있어, 그것의 성격에 따라 총론과 각론이 기여하는 정도는 다를 수 있는데, 앞서 언급한 ‘연도별 학력 추이 변화 분석’ 목적에 대해서는 총론이 기여하는 바가 큰 것에 반해, ‘교육과정의 질 관리’라는 목적에 대해서는 각론의 역할이 좀 더 크다고 볼 수 있다. 특히 교육 내용과 관련해서는 교과교육 연구자의 전문적 안목이 절대적으로 필요하다. 이러한 문제의식에 따라 이 연구에서는 ‘수학과 교육과정의 질 관리’와 관련하여 국가수준 성취도 평가의 수학과 현황

과, 그것이 안고 있는 문제점은 무엇이며, 그 개선 방향과 관련하여 수학교육계에서 관심을 가져야 할 점이 무엇인지를 구체적으로 밝히고자 하였다.

최근 들어 부각되고 있는 교육 정책이나 교육 프로그램에 대한 평가 연구의 실행 과정을 들여다보면 수학교육계의 전문적 안목과 노하우를 필요로 하는 부분이 있기 때문에, 수학교육 평가에서도 교육 정책이나 프로그램 평가 방법이나 결과 분석 방법 등에 관심을 가질 필요가 있으며, 이 논문에서 그러한 점이 부각되기를 기대해 본다.

II. 교육과정의 질 관리와 국가수준 성취도 평가

국가수준 성취도 평가의 목적이나 활용 방안을 다룬 연구들은 한결같이 ‘교육과정의 질 관리’를 강조해 왔다. 대표적인 예로, 이종승(1987)은 ‘교육과정의 시행에 있어서 질적 관리를 위한 노력을 계속하지 않는다면, 초기에 성공적으로 적용되던 교육과정도 시간이 경과함에 따라 그 효율성이 저하되는 퇴화의 위험을 맞게 된다. 따라서 교육과정에서 추구하는 목표들이 효과적으로 달성되고 있는지를 주기적으로 평가하여, 어떤 중대한 퇴화현상이 발견되면 그 퇴화의 원인을 밝혀내고 개선책을 강구해야 한다. 우리는 전국적 학력평가를 통하여, 학교의 교육목표는 학생들의 신체적, 심리적 발달수준에 비추어 타당한가, 학습경험의 선정과 조직은 적절한가 등 교육과정의 타당성 여부를 평가할 수 있고, 그 속에 내포된 문제점을 발견하여 차후 교육과정을 개선하는데 참고자료로 이용할 수 있다’라고 언급한 바 있으며, 조난심(1998)은 ‘사실 우리나라는 그동안

제7차에 이르기까지 빈번하게 교육과정 개정이 이루어졌지만, 어느 시기에도 설정된 교육과정에 대해 실제로 학생들이 어느 정도로 성취했는지를 체계적으로 평가해 본 적이 없었다. ...이는 달리 말하여, 교육과정에 대한 국가수준의 질 관리 및 그와 관련된 학교의 책무성에 대한 평가가 체계적으로 이루어진 적이 없었다는 것이다. 이는 또 많은 교육 정책들이 체계적인 자료의 뒷받침없이 수립되고 시행되었음을 의미한다'라고 하였다.

이러한 경향은 미국의 NAEP 평가에서도 찾아볼 수 있다. Lindquist(2001)에 따르면, NAEP의 평가 결과가 Curriculum and Evaluation Standards for School Mathematics(1989)와 Principles and Standards for School Mathematics (2000)에 간접적으로, 또는 직접적으로 미친 영향과, 그와는 반대로, 두 교육과정(Standards)이 NAEP의 평가들, 평가 문항 등 평가 방법에 미친 영향을 역사적으로 검토하면서, '평가는 교육과정(Standards)을 진일보하도록 한다면, 또한 반대로, 교육과정 이면에 있는 이념은 평가로 하여금 진일보하도록 한다'고 압축적으로 평가와 교육과정의 관계를 표현하였다.

다음으로, 본문의 논의 내용을 이해하는데 도움이 될 정보, 특히 수학과 교육과정과 관련하여 국가수준 성취도 평가 연구의 주요 특징을 설명하면 다음과 같다. 먼저 국가수준 성취도 평가는 절대평가를 지향한다는 점에 주지할 필요가 있다. 즉, 평가 대상 학생들 사이에 서열을 매기는 데 초점이 있는 것이 아니라, 각 교과교육에서 추구하고 있는 교육목표에 학생들이 어느 정도 또는 어떻게 도달하였는가를 알아보는 데 초점이 있다는 것이다.

둘째, 절대평가를 지향하기 때문에 평가 문항을 개발할 때 교육과정에 제시된 교육 내용을 기초로 한다. 수학과 교육과정에 개략적으

로 진술되어 있는 교육 내용을 요목화 하여 성취기준과 평가기준을 마련하고, 이 기준에 근거하여 평가문항을 개발한다. 국가수준 성취도 평가 수학 검사지는 이와 같은 방식으로 개발된 평가문항들로 구성된다.

셋째, 국가수준 성취도 평가에서는 평가 결과의 하나로 성취수준별 비율을 산출한다. 성취수준은 우수학력, 보통학력, 기초학력, 기초학력미달로 구분되며, 이를 위해 우수학력과 보통학력 사이, 보통학력과 기초학력 사이, 기초학력과 기초학력미달 사이를 구분하는 분할점수를 산출한다. 앞서 언급한대로, 국가수준 성취도 평가는 절대평가를 지향하기 때문에 분할점수를 산출할 때 교육과정에서 제시한 교육목표가 주요 고려대상이 된다. 예컨대 보통학력은 교육목표를 어느 정도까지 도달한 상태로 볼 것인가를 고려하게 되는 것이다.

Ⅲ. 수학과 교육과정과 국가수준 성취도 평가의 연계 현황 및 개선 방향

1. 수학과 교육과정의 교육 내용에 관한 학생들의 이해 정도 평가

가. 현황

일반적으로 교육평가의 주요 기능은 교육목표 도달 정도를 평가하는 것이다. 국가 교육과정에 제시된 교육 내용을 학생들이 이해하고 있는 정도를 평가하고자 하는 국가수준 성취도 평가에서는 학생들의 학력을 양화(量化)함으로써 그러한 목적을 수행한다. 평가 결과의 하나로 산출하는 '성취수준'별 비율이 대표적인 예가 될 수 있다. 성취수준은 교육

을 통해 학생들이 실제 성취한 정도, 또는 교육 목표 도달 정도를 의미하는 것으로, 이 평가에서는 성취수준을 크게 우수학력, 보통학력, 기초학력, 기초학력미달로 구분하여 각각에 속하는 학생 비율이 얼마나 되는지를 추정한다.

성취수준별 비율을 결정하는 관건은 무엇보다도 성취수준의 의미이다. 그 의미를 어떻게 설정하는가에 따라 그 비율이 달라지기 때문이다. <표 III-1>에서 보듯이, 절대평가를 지향하는 국가수준 성취도 평가에서는 2001년까지는 '최소 필수 목표'가, 2002년부터는 '교육과정의 기본내용'이 성취수준을 가르는 핵심 기준으로 사용되고 있다.

한편 '최소 필수 목표'나 '교육과정의 기본내용' 등의 핵심 기준 자체가 해석하기에 따라서 달라질 수 있는 임의적인 성격을 지니고 있기 때문에, 성취수준의 비율을 산출하기 위해서는

별도의 방법이 필요하다. 그 방법으로 2001년까지는 <표 III-2>와 같이 '문항의 난이도'를 중심으로 한 방법이 사용되었다. 그런데 그동안 이 방법의 한계가 지적되고, 그 개선 방향으로 다음을 골자로 한 내용이 꾸준히 제기되었다. 즉, 문항 자체의 난이도뿐만 아니라 각 교과 의 내용 전문가들의 판단이 결합되어 성취수준을 설정하여야 하며, 일반적으로 절대평가에서 성취수준을 설정하는데 보편적이고 안정적인 방법으로 알려져 있는 것⁴⁾을 사용하여야 하며, 특히 국내 국가수준 성취도 평가와 유사한 목적을 지닌 미국의 교육향상 평가(NAEP), 영국의 교육과정 평가(National Curriculum Assessment)에서 성취 수준 설정을 위해 오랫동안 활용해 온 방법들을 고려해야 한다는 것이다. 그래서 2002년 평가에서부터는 <표 III-2>의 방법이 '수정된 앙고프 방법(modified Angoff Method)'으로 바뀌었다.

<표 III-1> 국가수준 성취도 평가에서 사용한 성취수준의 일반적 의미

성취수준	2002~현재	2000~2001
우수학력	교육과정의 기본내용을 대부분(80% 이상) 이해한 수준	해당 학년에서 보통의 학생들이 일반적으로 성취하기를 기대하는 것보다 높은 성취를 보이는 수준, 즉 보통학력에 해당하는 것을 성취함과 동시에 심화·발전된 내용을 추가적으로 성취한 수준을 의미함.
보통학력	교육과정의 기본내용을 상당부분(50%~80%) 이해한 수준	해당 학년에서 보통의 학생들이 일반적으로 성취하기를 기대하는 수준, 즉 기초학력에 해당하는 것을 포함하여 정상적인 교수·학습 활동을 통해 성취할 것이라고 기대하는 일반적인 내용을 추가적으로 성취한 수준을 의미함.
기초학력	교육과정의 기본내용을 부분적(20%~50%)으로 이해한 수준	해당 학년의 모든 학생들이 반드시 성취하기를 기대하는 최소필수 목표를 성취한 수준
기초학력 미달	기초학력에 도달하지 못한 수준	해당 학년의 모든 학생들이 반드시 성취하기를 기대하는 최소필수 목표를 성취하지 못한 수준

4) 절대평가에서 성취수준을 설정하는 방법으로는 판단전라포착 방법(JPC 방법), 총체적 프로파일 방법, 북마크 방법 등과 같이 여러 가지가 있다(양길석, 2000).

수정된 앙고프 방법은 다음과 같다. 먼저 교과내용에 전문적인 식견을 가지고 있고, 해당 학교급의 학생들을 지도해 본 경험이 있는 교사, 즉 교과내용 전문가들을 선정한다. 이때 전국의 대도시, 중소도시, 읍면지역별 안배를 하여 교사를 선정한다. 둘째, 교과내용 전문가들은 수학과와 각 성취수준의 의미에 대하여 3-4 시간 정도 허심탄회한 논의를 거쳐 비교적 통일된 개념을 갖도록 한다. 셋째, 이렇게 얻은 각 성취수준에 대한 개념을 가지고, 평가문항별로 각 성취수준의 최저 경계선에 있는 학생들 중에서 그 문항을 맞힐 학생의 비율, 즉 기대정답률을 추정해 보도록 한다. 넷째, 모든 평가문항에 대하여 참여한 모든 교사들이 각 성

취수준별로 기대정답률을 산출하고, 그 결과를 모아 통계 처리를 하여, 각 성취수준을 구분하기 위한 분할기준점수를 산출한다. 다섯째, 그러한 분할기준점수를 사용하여 성취수준별 정답률을 산출한다. 예를 들어, 2003년과 2004년 평가 분석에서 수정된 앙고프 방법을 적용한 결과로 얻은 성취수준별 학생 비율은 <표 III-3>과 같았다.

나. 개선 방향

<표 III-1>에 제시된 성취수준의 의미는 '일반적'인 것으로, 국가수준 성취도 평가의 평가대상인 5개 교과에 모두 적용된다. 수정된 앙고프 방법 적용 과정에서는 이 일반적 의미를

<표 III-2> 2001년 평가에 사용된 성취수준의 판별 준거 및 실제 점수 환산 방법

성취수준	성취 수준 판별 준거	실제 점수 환산 방법
우수학력	난이도가 '하', '중'인 문항 모두에서 옳은 답을 하고, '상' 수준 문항의 50% 이상에서 옳은 답을 한 경우	난이도가 '하'와 '중'인 모든 문항에 해당하는 점수의 총합에 '상' 문항에 해당하는 점수의 50% 이상을 얻은 경우
보통학력	난이도가 '하', '중'인 문항 모두에서 옳은 답을 하고, '상' 수준 문항의 50% 미만에서 옳은 답을 한 경우	난이도가 '하'와 '중'인 모든 문항에 해당하는 점수의 총합에 '상' 문항에 해당하는 점수의 50% 미만을 얻은 경우
기초학력	난이도가 '하'인 문항에서 모두 옳은 답을 하고, '중' 수준 문항의 50% 미만에서 옳은 답을 한 경우	난이도가 '하'인 모든 문항에 해당하는 점수의 총합에 '중' 문항에 해당하는 점수의 50% 미만을 얻은 경우
기초학력 미달	난이도가 '하'인 문항의 50% 미만에서 옳은 답을 한 경우	난이도가 '하'인 문항에 해당되는 점수의 50% 미만을 얻은 경우

<표 III-3> 2003년과 2004년 중학교 3학년의 수학 성취수준별 비율 (단위: %)

성취수준	2003년	2004년
우수학력	13.5	16.9
보통학력	33.1	36.4
기초학력	41.9	38.0
기초학력미달	11.5	8.7

각 교과에 적합하도록 해석하여 교과 차원에서 각 성취수준의 '구체적' 의미에 대하여 합의해 간다. 그러한 구체적 의미가 교과전문가의 각 성취수준에 대한 개념이 되며, 수정된 앙고프 방법에서 핵심이 된다. 이런 사정을 고려할 때, 수정된 앙고프 방법의 적용 결과인 <표 III-3>과 같은 수학과 성취수준별 비율의 해석에서 결정적인 역할을 하는 것은 수학과에서 부여한 '구체적' 의미임을 알 수 있다⁵⁾.

사실 전문적으로 이루어지는 평가 연구들의 면면을 보면, 각 교과 수준에서 '구체적' 의미가 명문화되어 있는 것을 볼 수 있다. 대표적으로 PISA와 NAEP의 연구를 들 수 있다. 먼저 PISA의 경우 학생들의 수학 성취도를 보다 구체적으로 파악하기 위하여 학생들의 수학적 소양 수준을 1수준에서 6수준까지의 성취수준으로 구분하고 각 수준의 구체적 의미를 제시하고 있으며, 그 중 1수준과 6수준에 해당하는 것만 예시적으로 제시하면 다음과 같다(이미경, 2004).

- 1수준 : 관련된 정보가 모두 제시되고 질문이 분명하게 정의된 익숙한 맥락에 관한 문제들에 답할 수 있는 수준이다. 명확한 상황에서 정보를 확인할 수 있으며 정확한 지침에 따라 전형적인 절차를 수

행할 수 있다. 분명하게 제시된 행동과 주어진 조건으로부터 직접적으로 따라 나오는 행동을 수행할 수 있다.

- 6수준 : 복잡한 문제 상황에 대한 자신들의 조사와 모델링에 근거해서 정보를 개념화하고 일반화하고 이용할 수 있다. 학생들은 다양한 자료들과 표현들을 열거할 수 있으며 다양한 자료들과 표현들을 유연하게 해석할 수 있다. 학생들은 고등의 수학적 사고와 추론을 할 수 있다. 학생들은 잘 알려져 있지 않은 상황에 접근하기 위한 새로운 접근 방법과 전략을 개발하기 위해 기호적이고 형식적인 수학적 조작과 관련성을 익숙하게 다룰 수 있으며, 이를 통해 획득한 통찰력과 이해를 잘 알려지지 않은 상황에 적용할 수 있다. 학생들은 자신들의 행동과 반성을 형식화하여 정확하게 의사소통할 수 있다.

미국의 NAEP에서는 4, 8, 12학년을 대상으로 평가를 실시하고 있으며, 각 학년별로 성취수준을 기초(basic), 숙달(proficient), 우수(advanced)로 구분하고 그 구체적 의미를 기술하고 있으며, 다음은 예시적으로 8학년의 것을 제시한 것이다.

- 기초 수준 : 다이어그램, 표, 그래프 등을 사용하여 문제를 올바르게 풀 수 있는

5) 실제로 2003년 국가수준 성취도 평가 분석 결과를 발표하였는데, 당시 언론에서는 '중고생 절반이 수업을 못 따라간다니'(동아일보, 04년 1월 12일 자)와 같이 매우 충격적으로 해석한 바 있다. <표 III-1>의 일반적 의미에서 기초학력을 '교육과정의 기본내용을 부분적으로 이해한 수준'으로 정의하였는데, 이에 더하여 기초학력과 기초학력미달에 속하는 학생들은 수업을 따라가지 못하는 학생으로 해석한 것으로 보인다. 그런데, 앙고프 방법을 적용할 때 교과내용 전문가들이 부여한 성취수준에 관한 여러 구체적 의미를 고려할 경우, 성취수준 비율에 대한 해석은 충분히 달라질 수 있다. 수학과에서는 성취수준의 의미를 여러 측면에서 분석하여 구체적 의미를 공유하고자 노력하였으며, 그 중 간단한 한 가지를 소개하면 다음과 같다. 수학 교육과정의 교육목표가 가장 잘 구현된 것은 교과서이므로, 이 교과서의 체제를 이용하여 각 성취수준을 해석해 볼 수 있다. 한 단원에는 여러 '예제'와, 그 예제에서 이해한 내용을 적용하는 '문제'들이 나온다. 단원의 끝에는 '연습문제'가 나오는데, 이 '연습문제'에서는 여러 문제와 함께 '응용문제'가 나온다. 이러한 교과서의 체제에 비추어, 기초학력 수준의 학생들은 '예제'를 풀 수 있는 정도, 보통학력 수준은 '예제' 후에 곧바로 이어지는 '문제'를 풀 수 있는 정도, 우수학력 수준은 한 단원의 끝에 나오는 연습문제의 '응용문제'를 풀 수 있는 정도로 해석하였다. 이러한 해석을 고려할 때, 교과서의 '예제'를 풀 수 있다고 본 기초학력 학생을 두고 '수업을 못 따라간다'고 해석한 것은 일반적 의미만 고려하고 실질적인 의미를 지닌 수학과 차원의 구체적 의미를 고려하지 않은 데서 비롯된 오해라고 할 수 있을 것이다.

수준이다. 전략과 기술적 수단을 적절히 선택하고 사용하여 NAEP에서 다루는 모든 내용 영역의 문제를 풀 수 있는 수준이다. 또한 이 수준의 학생들은 기본적인 대수적 개념과 비형식적인 기하 개념을 사용하여 문제를 해결할 수 있다.

- 숙달 수준 : 추측을 할 수 있고, 자신의 생각을 정당화할 수 있으며, 적절한 예를 제시할 수 있는 수준이다. 분수, 비율, 소수 사이의 연결성이나 대수나 함수와 같은 여러 수학적 개념들 간의 연결성을 이해할 수 있다. 또한 이 수준의 학생들은 실제적인 상황에서 문제해결을 할 수 있을 정도의 충분한 이해를 가지고 있다.
- 우수 수준 : 학생들이 모델을 발전시킬 수 있는 일반화를 위해 예와 반례를 만들어 낼 수 있는 수준이다. 어떤 답의 합리성 여부를 따지기 위해 수 감각과 기하학적 인식을 사용할 수 있다. 또한 이 수준의 학생들은 추상적 사고를 통하여 독자적으로 문제를 해결할 수 있으며, 자신이 내린 결론이 바탕을 두고 있는 추론 과정을 설명할 수 있다.

이와 같이 성취수준에 대한 구체적 의미가 갖는 실질적인 중요성에도 불구하고 현재 국가수준 성취도 평가에서는 아직까지 그 의미가 명료화되지 않은, 따라서 명문화되지 않은 상태라고 할 수 있다. 수학과는 물론 각 교과 차원에서 성취수준의 구체적 의미를 명료화하고 명문화하는 작업은 성격상 간단히 이루어질 수 없다. 그러한 작업이 가능하려면 몇 가지 선결조건이 필요하다. 우선 '수학 교육과정의 기본 내용'이라는 개념에 대한 합의가 전제되어야 하는데, 그러한 전제를 충족하기가 결코 간단치 않다. 둘째, 2002년과 2003년 국가수준 성취도 평가에서는 앙고프 방법의 적용 단계에 10-

16명의 교과내용 전문가들이 3-4시간에 걸쳐 논의하여 수학과 성취수준에 대한 구체적 의미를 마련하기는 하였지만, 그 의미가 '국가수준'의 의미가 되도록 하기 위해서는 좀 더 오랜 시간을 두고 많은 교과내용 전문가들이 참여한 가운데 이루어질 수 있는 논의가 필요하다. 셋째, 그러한 논의는 사변적이지 않고, 구체적인 정보에 터하여 진행되어야 하는데, 그러한 구체적 정보들을 밝히고 수집하는 작업이 선행되어야 한다.

요컨대, 현재 국가수준 성취도 평가에서 5개 교과에 모두 적용되는 <표 III-1>의 성취수준에 대한 일반적 의미를 원칙으로 유지하면서, 각 교과 수준에서는 성취수준의 구체적 의미를 명료화하여 명문화하여야 한다. 국가수준 성취도 평가에서 분석한 결과로부터 교육 내용에 대한 학생들의 이해 정도에 관한 정보를 얻고자 할 때, 교과 수준의 성취수준에 대한 구체적 의미는 반드시 참조하여야 할 사항이기 때문이다.

2. 수학 교육 내용 이해 정도 측면에서 내용 영역 간 비교 평가

가. 현황

수학과 교육과정과 관련지어 국가수준 성취도 평가에 기대할 만한 내용 중 하나는 내용영역 간 비교 정보라고 볼 수 있다. 수학과와 경우 국민공통기본교육과정을 수와 연산, 도형, 측정, 확률과 통계, 문자와 식, 규칙성과 함수의 6개 내용 영역으로 구분하고 있는데, 학생들의 이해에 있어 내용 영역 간 어떤 차이가 있는지를 파악하는 것이다. 이러한 정보는 수학과 교육과정의 내용 영역 간 수준이나 난이도 등을 조정할 때 긴요하게 사용될 수 있다.

실제로 2002년 평가에서는 다음 <표 III-4>와 같은 결과로부터 내용 영역별 학생들의 전반적인 이해 정도를 파악하고 비교한 적이 있다. 예컨대, 중학교 3학년의 경우 우수학력 비율은 도형 영역에서 14.9%로 가장 낮으며, 반대로 확률과 통계 영역에서 32.4%로 가장 높았다. 기초학력미달 비율은 규칙성과 함수 영역에서 15.4%로 가장 높았으며, 도형 영역에서 4.7%로 가장 낮았다. 이러한 결과로부터 성취수준별로 학생들이 어려워하는 내용 영역이 무엇인지에 관한 정보를 얻을 수 있다.

위와 같은 전반적인 이해뿐만 아니라 문항 차원에서도 학생들의 이해에 관한 내용 영역 간 비교를 할 수 있다. [그림 III-1]과 [그림 III-2]의 문항에서 다룬 주제들은 중학교 2학년 수

준의 문자와 식 영역과 규칙성과 함수 영역에서 각각 기본적으로 다루어지는 내용으로 볼 수 있다. 문자와 식 영역의 문항보다 규칙성과 함수 영역에서, 정답률이 약 10% 정도 낮으며, 성취수준별로는 우수학력보다는 보통학력과 기초학력에서 그 정답률이 현저히 낮아지는 것을 알 수 있다. 이는 학생들이 규칙성과 함수 영역을 어려워한다는 것과 특히 성취수준이 낮을수록 그 정도가 심하다는 것을 문항 수준에서 보여주는 것이다.

나. 개선 방향

국가수준 성취도 평가의 결과로부터 전반적으로, 그리고 문항 차원에서 내용 영역 간 학생들의 이해에 관한 비교 정보를 추정해 볼 수

<표 III-4> 2002년 중학교 3학년 성취도 평가의 내용 영역별 성취수준의 백분율

성취수준	수와 연산	문자와 식	규칙성과 함수	도형	측정	확률과 통계	전체
우수학력	21.9	24.6	23.5	14.9	31.6	32.4	18.6
보통학력	30.4	34.7	31.3	49.1	31.2	28.5	41.6
기초학력	35.7	28.1	29.8	31.3	26.8	30.0	32.5
기초학력미달	12.0	12.6	15.4	4.7	10.3	9.1	7.3

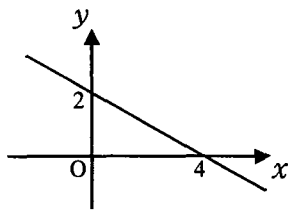
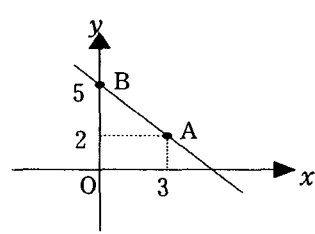
x 에 대한 이차방정식 $2(x-1)^2=8$ 을 풀면? ① $x=2$ ② $x=\pm 2$ ③ $x=-3$ 또는 $x=5$ ④ $x=3$ 또는 $x=-1$ ⑤ $x=-3$ 또는 $x=1$		이차방정식 $x^2-10x+25=3$ 을 풀면? ① $x=\pm\sqrt{3}$ ② $x=5\pm\sqrt{3}$ ③ $x=-5\pm\sqrt{3}$ ④ $x=5+\sqrt{3}$ ⑤ $x=5-\sqrt{3}$			
정답률	64.9%		정답률	55.1%	
수준별 정답률	기초	52.2%	수준별 정답률	기초	32.5%
	보통	89.0%		보통	67.6%
	우수	96.7%		우수	94.0%

[그림 III-1] 2003년 중학교 3학년의 문자와 식 영역 평가문항 예시

있음을 알아보았다. 그러한 추정이 좀 더 타당성을 갖기 위해서는 각 내용 영역별로 '대표성 있는' 문항들을 '충분히' 출제할 필요가 있다. 그런 점에서 볼 때, 국가수준 성취도 평가에서 내용 영역별로 '대표성 있는' 문항을 출제하고 있다는 점에 대해서는 의문의 여지가 없는 편이지만, '그 수가 충분인가'에 대해서는 재고의 여지가 많다. <표 III-5>에서 보듯이, 수학과에

서 내용 영역별 출제 문항 수는 많은 경우 10 문항, 적은 경우에는 5문항이다. 이 정도 수치의 문항으로는 내용 영역 간 비교가 타당성 있게 이루어졌다고 보기 힘들 것이다. 따라서 각 내용 영역별 출제 문항 수를 현재보다 훨씬 많이 늘려야 한다.

실제로 2002년 평가에 대해서는 <표 III-4>와 같이 내용 영역별 비율을 산출하였지만, 출제

<p>그래프가 오른쪽 그림과 같은 일차함수의 식은?</p> 		<p>다림과 같이 두 점 A, B를 지나는 직선을 그래프로 하는 일차함수의 식은?</p> 			
<p>① $y = -\frac{1}{2}x + 2$ ② $y = \frac{1}{2}x + 2$</p> <p>③ $y = \frac{1}{2}x + 4$ ④ $y = -2x + 2$</p> <p>⑤ $y = 2x + 2$</p>		<p>① $y = x + 5$ ② $y = -x + 5$</p> <p>③ $y = \frac{2}{3}x + 5$ ④ $y = -\frac{2}{3}x + 5$</p> <p>⑤ $y = -\frac{3}{2}x + 5$</p>			
정답률	41.1%	정답률	37.9%		
수준별 정답률	기초	16.0%	수준별 정답률	기초	15.6%
	보통	50.0%		보통	54.6%
	우수	92.7%		우수	91.7%

[그림 III-2] 2003년 중학교 3학년의 규칙성과 함수 영역 평가문항 예시

<표 III-5> 2003년 수학과 성취도 평가의 선택형 문항의 내용 영역별 구성

학년	연도	수와 연산	문자와 식	규칙성과 함수	도형	측정	확률과 통계	합계
초6	03	10	7	6	10	10	5	48
	02	9	5	5	7	9	5	40
중3	03	9	10	9	9	5	6	48
	02	8	8	6	8	5	5	40
고1	03	8	10	8	10	6	6	48
	02	7	8	8	7	5	5	40

되고 있는 문항 수로 그러한 결과를 산출하는 것은 무리라는 자체 비판이 제기되어 2003년 평가 이후로는 그러한 분석 결과를 산출하지 않고 있다. 사실 그동안 국가수준 성취도 평가에서 출제되는 평가 문항의 수를 늘려야 한다는 지적은 꾸준히 있어왔다(2001, 나귀수 외; 2002, 이봉주 외). 그런데 현행 평가 방식을 그대로 유지하면서 문항 수를 늘리게 되면, 학생들의 부담이 지나치게 크게 증가한다. 예컨대 지금까지 국가수준 성취도 수학과 평가에서는 대체로 학생들로 하여금 60분 정도의 시험 시간에 30~40문항을 풀도록 하였다. 문항 수가 늘어나면 그만큼 시험 시간을 늘려야 하며, 그렇게 될 경우 학생들의 시험에 대한 심리적 부담감이 증가되어 오히려 시험을 대충 볼 가능성이 커진다. 따라서 학생들의 심리적 부담감을 줄이면서 문항 수를 늘리는 방안을 마련하여야 한다.

출제 문항 수를 늘려 가급적 많은 교육 내용을 평가하면서도 각 학생이 시험에 임하는 시간을 줄여 학생들의 심리적 부담을 낮추는 방안으로, 행렬 표집(matrix sampling) 방법을 생각해 볼 수 있다. 행렬 표집에서는 먼저 필요한 대로 교육 내용에 대한 평가 문항들을 출제하고 나서 이 문항들을 몇 개의 묶음(cluster)들로 쪼개고, 각 학생들은 여러 묶음들 중에서 특정 묶음에 해당하는 문항들에 대해서만 응답을 하면 된다. 즉, 학생들은 문항 전체가 아니라 일

부 문항에 대해서만 답을 하기 때문에 시험 시간이 그만큼 줄어들게 된다. 문항 수가 너무 많아 모든 피험자에게 모든 문항에 대한 검사 실시가 어려운 경우 피험자 집단마다 다른 문항으로 구성된 검사를 치르게 하고, 일부 문항에 대해서만 공통된 문항의 검사를 보게 함으로써 서로 다른 검사를 치르는 학생들의 성취도 결과를 동등하게 비교하려는 것이다.

예를 들어, 미국의 NAEP에서는 <표 III-6>에서 보듯이 1996년에 시행된 평가에서 4학년, 8학년, 12학년 각각에서 수학 문항이 144개, 164개, 165개 출제되었다. 이렇게 많은 문항이 출제되기는 하지만 이 문항들을 블록(Block)화하여 시험지를 구성하기 때문에, 실제로 한 학생이 풀게 되는 시험 문항의 수는 20문항 이하이다(Allen, 1999). 행렬 표집 방식은 PISA에서도 사용하고 있다. 2003년에 실시된 PISA의 경우, 인지 검사 도구는 총 167개의 문항으로 구성되어 있으며, 행렬 표집을 이용하여 13종의 검사지를 각 학생에게 나누어 배정하였다(채선희 외, 2003)

NAEP이나 PISA와 마찬가지로 국가수준 성취도 평가는, 학생 개개인의 학업성취도가 아니라, 학생 집단 전체의 학업성취도를 추정하는 데 목적으로 둔 교육프로그램 평가이다. 따라서 그 목적을 달성할 수 있는 적절한 수단이 있다면, 한 학생이 모든 문항을 풀 필요는 없다. 그럼에도 불구하고, 암암리에 우리나라에서

<표 III-6> 1996년 미국 NAEP 수학과 내용 영역별 출제 문항 수

단계	수 감각, 특성, 연산	측정	기하와 공간감각	자료 해석, 통계, 확률	대수와 함수	합계
4단계	59	25	25	17	18	144
8단계	47	27	31	25	34	164
12단계	40	23	27	34	41	165
합계	146	75	83	76	93	473

는 정서적으로 여전히 ‘모든 학생들은 모두 똑같은 문항을 풀어야 하며, 분석은 그러한 결과에 기초하여야 한다’는 생각이 지배적인 것으로 보인다. 이러한 정서에 주의를 준 Romberg (1992)의 다음의 말에 귀를 기울일 필요가 있을 것이다.

어떤 사안에 대하여 의사결정을 내리고자 할 때 그 ‘사안’을 구체화할 필요가 있다. 그동안 치룬 시험들을 보면 한결같이 모든 학생들이 모든 문항에 답을 하도록 하고 있으며, 이는 관습이 되어버렸다. 목적이 무엇이든지 간에 상관없이 학생 개개인의 정보가 모든 수준에서 축적되었다. 이러한 관습은 사실 비용 면에서나 시간 면에서나 상당한 낭비를 낳는다. ...학교 수행에 관한 개략적인 정보를 얻고자 한다면, 이는 다른 방식으로 하여 얼마든지 효과적으로 원하는 바를 얻을 수 있다.

위의 인용문대로 우리나라에서는 평가 대상에 속하는 모든 학생들은 똑같이 모든 문항을 풀어야 하고, 또한 시험을 치룬 학생들에게는 평가 결과를 담은 성적표를 통하여 피드백을 주어야 한다는 것이 관습처럼 되어 있다. 현재 국가수준 성취도 평가 역시 그러한 관습의 연장선 위에서 시행되고 있다고 볼 수 있다. 그런데, Romberg의 지적대로, 평가 목적을 효율적으로 달성하기 위한 방식을 강구한다면, 때로는 학생들이 일부 문항만을 볼 수도 있으며, 때로는 성적표를 발송하지 않을 수도 있는 것이다. 국가수준 성취도 평가에서 평가방법, 시행, 결과 분석 내용, 결과 보고 내용 등 일련의 과정들이 그러한 목적을 ‘효율적으로’ 이루도록 설계되었는지를 심각하게 고민할 필요가 있다. 이 평가를 실시하는데 관계되는 집단이 많고, 표집대상이 크고, 적지 않은 예산이 소요되기 때문에, 비효율적인 방식을 택하였을 경우 그것이 미치는 파장이나 낭비는 ‘국가수준’에

서 상당하기 때문이다. 그런 점에서 볼 때, 국가수준 성취도 평가를 통해 교육과정의 질을 관리한다는 목적을 달성할 수 있도록, 모든 학생들이 똑같이 모든 문항을 풀어야 하고, 성적표를 반드시 발송해 주어야 한다는 등의 관습과 같은 편견에서 탈피하고, 대신 목적 달성에 효율적인 평가 기법을 강구할 필요가 있다.

3. 평가를 통해 수학 성취수준별 학력 특징에 관한 정보 수집

가. 현황

국가수준 성취도 평가에서는 평가 결과의 하나로, 문항별로 성취수준별 정답률을 산출하고 있다. 예컨대, [그림 III-3]의 첫 번째 문항의 경우 기초학력 수준에 속한 학생들 중에서 그 문항의 정답을 맞힌 학생 비율이 72.9%였음을 알 수 있다. 또한 성취수준의 특징을 잘 보여주는 문항, 즉 대표문항을 선정하는데, 그 기준은 대체로 해당하는 성취수준에서 정답률이 70% 이상인 문항으로 한다. 원래 성취수준별로 각 문항에 대해 숙달되었다고 말하는 경우는 고전점 사이론을 이용할 때 정답률이 50%~80%인 문항인데, 학생들이 추측하여 풀 수 있다는 점을 고려하여 70% 이상인 문항으로 한다.

이러한 자료들을 활용하면 <표 III-7>과 같이 교육 내용 별로 각 성취수준에 속한 학생들의 강점과 약점을 파악할 수 있다. 교육과정 개정의 추세를 보면, 제6차 교육과정 이후로 학생들의 수준에 적합한 교육과정의 운영이 주요 화두가 되어 오고 있다. 제7차 교육과정의 ‘수준별 교육과정의 운영’은 그 정점에 있다고 보아도 될 것이다. 학생들의 성취수준별 학력 특성에 관한 정보들을 체계적으로 축적하여 데이터베이스화하면, 이는 학생들의 수준을 고려하여 교육과정을 운영하거나 수준별 수업을 실행

하는 데 기초적인 자료로서 도움이 될 수 있다. 예컨대 수준을 상, 중, 하로 구분할 때, 각 수준별로 강점과 약점이 드러나 있을 경우 어떤 약점을 보완하여 좀 더 나은 수준으로 상승시킬 것인가와 같은 식의 접근이 가능할 수 있기 때문이다.

나. 개선 방향

학생들의 성취수준별 학력 특성에 관한 정보들을 체계적으로 축적하여 데이터베이스화하는 것의 가치나 유용성이 인정된다는 전제 아래,

좀더 풍부한 데이터베이스를 구축하기 위해서는 앞서 2절에서 언급한 행렬 표집과 같은 방식을 적극 도입하여 가급적 많은 교육 내용을 포괄할 수 있도록 평가 문항의 수를 늘려야 할 것이다.

또한 평가 결과에서 수학과 성취수준별 학력 특징을 잘 드러낼 수 있도록 수학과 평가문항 개발 단계에서부터 체계적인 접근을 할 필요가 있다. 예컨대 2003년 중학교 3학년 학생들을 대상으로 출제된 수학 문항들에 대한 성취수준별 정답률을 보면, 성취수준 중에서 특히 기초 학력 수준에서 정답률이 70% 이상인 문항이

다음 표는 동호네 학교의 5학년 반별 학생 수를 조사하여 나타낸 것입니다. 6학년 때 5개 반으로 나누어 각 반의 학생 수를 같게 하려고 합니다. 한 반의 학생 수를 몇 명씩 해야 하나요?

<5학년 반별 학생 수>

반	1	2	3	4	5
학생 수 (명)	36	38	37	39	35

기초미달	28.1%
기초학력	72.9%
보통학력	89.8%
우수학력	97.5%

① 35 명 ② 36 명 ③ 37 명 ④ 38 명 ⑤ 39 점

다음 도형은 정사각형 3개를 이어 붙인 모양입니다. 이 도형의 둘레의 길이는 몇 m입니까?

기초미달	18.4%
기초학력	41.6%
보통학력	82.9%
우수학력	98.1%

① 0.36 m ② 0.72 m ③ 0.84 m ④ 0.96 m ⑤ 1.2 m

다음 그림은 정사각형을 합동인 4개의 직사각형으로 나눈 것입니다. 한 직사각형의 둘레의 길이가 100 cm일 때, 정사각형의 둘레의 길이는 몇 cm입니까?

기초미달	9.2%
기초학력	10.5%
보통학력	32.0%
우수학력	81.2%

① 100 cm ② 120 cm ③ 160 cm ④ 200 cm ⑤ 400 cm

[그림 III-3] 초등학교 6학년의 성취수준별 대표 문항

없었다. 이는 기초학력 학생들이 그만큼 수학 교과에서 다루는 내용에 대한 이해 정도가 낮다는 것을 나타내는 것일 수도 있지만, 평가문항의 난이도가 높아 기초학력 수준에 속한 학생들이 실력 발휘를 할 수 있는 문항이 없었을 가능성도 있다. 2004년 평가에서는 평가 문항의 개발 단계에서 문항 개발진들에게 그러한 사정을 감안하여 문항을 개발하도록 하였으며, 그 결과 2004년 중학교 3학년 수학 문항 중에서 15% 정도 문항에 대해 기초학력 수준의 학생들이 70% 이상의 정답률을 얻었다.

국가수준 성취도 평가에서 평가 결과로 네 성취수준의 비율을 산출하고 있는데, 각 성취수준의 특징이 드러나도록 평가 문항을 개발하는 것은 국가수준 성취도 평가의 결과 활용의

폭을 넓힌다는 측면에서 바람직하다고 생각된다. 그런 점을 감안할 때, 평가문항을 개발할 때 ‘성취수준의 특징을 드러낼 수 있도록 평가문항을 개발한다’는 지침을 별도로 추가하고, 이에 관한 노하우를 마련해야 할 것이다.

4. 평가를 통해 수학 교육 내용 적정화에 활용 가능한 기초 자료 수집

가. 현황

우리나라에서 교육과정을 개정할 때, 연구·개발 모형에 따라, 즉 교육과정 전문가들이 기존 교육과정의 문제점을 분석하고 개선 방향을 탐색하기 위해 공동의 숙의를 거치는 과정을 통해 진행되기 시작한 것은 제4차 교육과정 개

<표 III-7> 초등학교 6학년 학생들의 성취수준별 강점과 약점

성취수준	강점	약점
우수 학력	<ul style="list-style-type: none"> ◎ 괄호가 없는 사칙혼합계산 문제를 해결할 수 있다. ◎ 복잡한 문제해결과정에서 나타난 오류를 찾아 이를 바르게 고칠 수 있다. ◎ 실생활 맥락에서 두 가지 이상의 개념에 대한 이해와 다소 복잡한 계산 능력이 요구되는 문제를 해결할 수 있다. 	
보통 학력	<ul style="list-style-type: none"> ◎ 괄호가 제시된 식의 사칙혼합계산을 할 수 있다. ◎ 간단한 전략을 적용하여 문제를 해결할 수 있다 	<ul style="list-style-type: none"> ◎ 괄호가 없는 식의 사칙혼합계산에는 취약하다. ◎ 문제해결 과정을 보고 틀린 곳을 찾을 수 있다. 그런데 오류를 수정하는 데에는 취약하다. ◎ 여러 수학적 개념을 활용한 실생활 문제를 해결하는 능력은 부족하다.
기초 학력	<ul style="list-style-type: none"> ◎ 1보다 작고 분모가 10의 거듭제곱인 분수를 소수로 바꾸는 것과 같이 간단한 수 개념을 이해할 수 있다. 그런데 수에 대한 개념적 이해와 약분과 같은 절차적 지식을 동시에 요구하는 문항에서는 취약하다. ◎ 도형의 그림이 주어진 상태에서 그 도형의 성질을 분석할 수 있다. 	<ul style="list-style-type: none"> ◎ 수에 대한 개념적 이해와 약분과 같은 절차적 지식을 동시에 요구하는 문항에서는 취약하다. ◎ 이차원 표현으로부터 삼차원 모양을 상상하여 문제를 해결하는 능력에서는 취약하다.

정 때부터이다. 그리고 제4차 교육과정 개정에서부터 꾸준히 대두된 이슈 중에 하나는 '교육 내용의 적정화'였으며, 교과별 교육 내용 적정화 담론은 주로 교과별 학습량 감축 및 난이도 조정이 주를 이루어 왔다(임재훈 외, 2004). 교육과정을 개정할 때, 이와 같은 교육 내용 적정화를 위해 참조한 근거 자료로는 크게 두 가지 유형이 있는데, 첫째, 외국의 교육과정과 비교 분석한 자료이며, 둘째, 각각의 교육 내용에 관한 학생들의 이해 정도와 상태를 평가하여 얻은 자료이다. 후자와 관련하여 제5차 교육과정 개정 시안 보고서에 주목해 보면, 당시 초·중학교 수학과 교육과정 운영 실태 조사의 일환으로 국민 학교 전 학년과 중학교 1·2학년의 학습 내용에 대한 이해도 조사를 실시한 것을 볼 수 있다. 이 '이해도 조사'에서는 총 194문항을 사용하여 학생들의 기초학력을 파악하고 그 결과를 교육과정 개정의 근거 자료로 활용하였다(성효석 외, 1986). 그러한 사례를 볼 때, 일종의 이해도 조사라고 할 수 있는 국가 수준 성취도 평가 연구로부터 얻은 자료들은 교육과정 개정에서 여전히 주요 이슈가 될 교육 내용 적정화에 유용하게 활용될 수 있다.

평가원에서는 제7차 수학과 교육 내용의 적정성을 평가하고 이를 근거로 수학과 교육 내용의 적정화 방안을 모색하고자 한 『수학과

교육 내용 적정성 분석 및 평가』 연구를 수행한 바 있다(임재훈 외, 2004). 이 연구에서는 각 교육 내용에 대해 현재 지적되고 있는 문제점은 무엇이며, 그와 관련하여 추후 어떤 정보나 연구가 필요한지를 다루고 있다. 예컨대, '자연수의 혼합계산'과 관련하여, 설문조사 결과 교사들이 가르치기 힘들다고 응답하였는데 그 원인이 무엇인지를 추정하고 있으며, 또한 싱가포르와 일본에서 자연수의 혼합계산을 가르치는 방식을 검토하면서, 다음과 같은 연구 문제를 제안한 바 있다.

이러한 점을 고려하면 자연수 혼합계산을 상급학년이나 학교급으로 이동하는 방안을 고려할 수 있으나, 이를 위해서는 이 내용이 학생들에게 심리적으로 얼마나 어려운 내용인지에 대한 연구가 필요하다. 이 영역에 관한 학생들의 성취 정도에 대한 광범위한 연구 조사가 필요하며, 학생들의 이해 수준도 낮은 것으로 드러나는 경우에는 ...

실제로 국가수준 성취도 평가에서 출제된 문항을 통하여 위에 제기된 연구 문제와 관련된 정보를 찾아볼 수 있다(그림 III-4). 그 정보는 세 가지로 요약할 수 있는데, 첫째, 다른 내용을 다룬 문항들에 비해 학생들의 정답률이 낮으며, 둘째, 자연수 혼합계산에서 괄호가 있는 경우보다 괄호가 없는 경우 정답률이 30% 정

03-A-3. 다음을 바르게 계산한 것은 어느 것입니까?				02-05. 다음을 계산하면 얼마입니까?			
$4 + 3 \times 12 - 8 \div 2$				$70 - 49 \div 7 + \{ 3 \times 7 - (4 + 2) \}$			
① 16		② 36		③ 38		④ 66	
④ 72		⑤ 80				⑤ 78	
전체 정답률(%)	성취수준별 정답률(%)			전체 정답률(%)	성취수준별 정답률(%)		
	기초	보통	우수		기초	보통	우수
46.3	18.5	47.0	86.5	72.6	30.2	69.9	93.9

[그림 III-4] 혼합 사칙 계산 문제

도 낮으며, 셋째, 상대적으로 우수학력보다는 보통학력과 기초학력에서 괄호가 없는 문항에서 정답률이 현저히 떨어진다는 것이다. 이러한 정보는 추후 교육과정 개정에서 교육 내용의 적정성을 검토하고 적정화 방안을 세울 때 기초 자료로 사용될 수 있을 것이다.

국가수준 성취도 평가 보고서에서는 ‘문항별 분석’, ‘문항 내용 분석’ 등의 항목에서 각각의 수학 교육 내용에 대한 학생들의 이해 상태를 체계적으로 정리하고 있다. 정리 방식에 있어 ‘2000년~2002년’과 ‘2003년과 2004년’ 사이에는 적지 않은 차이가 있다. 보고서 지면이 한정되어 있기 때문에 평가에 출제된 문항 중에서 일부 문항을 선택하여 보고서에 수록하는 데, ‘2000년~2002년’에는 주로 정답률이 기준이 되어, 정답률이 낮은 문항들을 선택하여 분석하였다. 이에 대해 ‘2003년과 2004년’에는 두 가지 측면에서 이전과는 다른 방식으로 문항을 분석하였다. 첫째, 교육 내용을 기준으로 삼아, 관련이 깊은 교육 내용별로 문항을 묶어 시사점을 분석하고자 하였다. 이를테면, 초등학교에서는 사칙 계산 능력의 평가가 주요한 관심사라고 보고, 그와 관련된 문항들을 한데 묶어 분석하였다(<표 III-8>). 이와 같은 분석이 가능했던 결정적인 여건으로, 2002년까지의 평가와 달리 2003년에는 문항 수가 20% 정도 늘었다

는 점을 들 수 있다. 다음으로, 2003년과 2004년 보고서에서 시도한 두 번째 문항 분석 방식은, 당해연도의 분석 결과만이 아니라, 이전 연도에 나온 분석 결과도 사용하였다는 점이다. 즉, 2003년 보고서에서는 2002년의 평가 결과를, 2004년 보고서에서는 2002년과 2003년의 평가 결과도 함께 사용하여 분석하였다. 이와 같은 분석 방식이 가능하였던 것은, 매년 새로 개발되는 검사지에 이전 연도에 사용한 문항 중 일부 문항을 공통 문항으로 다시 사용하고 이를 활용하여 ‘동등화’⁶⁾하는 등 소위 전문적인 평가 기법을 도입하였기 때문이다. 이렇게 분석 방식에 변화를 줌으로써 학생들이 특정 교육 내용을 어떻게 이해하고 있는지를 이전보다는 좀더 입체적으로 파악할 수 있었다.

학생들이 교육 내용을 어떻게 이해하고 있는가에 관해 심층적인 정보를 얻는 데에는 수행평가 문항이 유익하다는 것은 익히 알려진 사실이다. 2003년 국가수준 성취도 평가에서는 수행평가 문항을 출제하고 채점할 때 이전과는 차별되는 노력을 기울였는데, 이를테면 출제자들을 평소 수행평가에서 학생들의 반응에 대한 연구를 해본 경험이 있는 연구자들로 구성하였으며, 문항을 채점할 때에는 학생들의 오류를 반영한 채점코드를 정하여 학생들의 반응 유형을 분류하여 채점하였다. [그림 III-5], <표 III-9>의 예는 초등학교 6학

<표 III-8> 여러 가지 사칙 계산 관련 문제

문항 번호	문항 설명	정답률(%)
A-2	소수의 덧셈을 바르게 하기	76.1
A-3	간단한 수에 관한 사칙연산이 섞여 있는 식의 계산순서를 알고 계산하기	46.3
A-4	(소수)×(자연수)를 계산하여 도형의 둘레 구하기	72.3
A-6	생활 장면의 문제를 해결하기 위해 분모가 다른 분수의 덧셈과 뺄셈을 하기	47.7
A-16	실생활 문제를 통해 시간의 덧셈과 뺄셈을 바르게 하기	77.6
B-6	생활 장면의 문제를 해결하기 위해 분모가 다른 분수의 덧셈과 뺄셈을 하기	44.0

6) 자세한 내용은 평가원 보고서 「국가수준 학업성취도 평가 연구 -총론-」(정구향, 2004) 참조

나. 개선 방향

교육과정을 개정할 때 수학 교육 내용의 적정성 평가는 여러 가지 준거를 바탕으로 종합적으로 판단되어야 할 것이다. 즉, 수학 교육과정에서 다루어지고 있는 주제나 개념 등의 ‘학습량’ 측면, 다루어지는 학습 주제의 ‘난이도’ 측면, 주제나 개념이 수학교육의 목표와 관련성이 있는지 학습자들의 필요나 실생활과 관련이 있는지 등과 같은 내용의 ‘타당성’ 측면, 가르치고자 하는 교육 내용이 적절히 선정·조직되어 있는지와 관련된 내용의 ‘연계성’ 측면 등이 고려되어야 할 것이다(임재훈 외, 2004). 이 네 가지 준거 중에서 국가수준 성취도 평가를 통해 얻을 수 있는 정보는 특별히 두 번째로 언급된 학습 주제의 ‘난이도’와 관련이 깊다고 할 수 있을 것이다.

교육 내용 적정화에 필요한 ‘난이도’ 정보는 미시적일수록 그 활용도가 클 것이다. 함수 전체에 대한 난이도 보다는 함수 영역 안에서, 이를테면 일차함수의 기울기 개념에서의 난이도 정보가 더욱 긴요할 수 있는 것이다. 따라서 수학과 교육 내용 적정화에 필요한 난이도 정보를 얻기 위해서는, 전체 교육 내용을 최대한 세분화하여 요소화 하고, 각 요소별로 평가 문항을 개발하여 평가 결과를 구축하는 것이 바람직할 것이다.

그런데 현재의 국가수준 성취도 평가 체제에서는 그와 같은 요소별 정보 구축에 한계가 있다. 국가수준 성취도 평가는, 모두에서 언급한 것과 같이, 연도별 학력 변화 추이 분석을 목적으로 하고 있다. 이 목적을 효율적으로 달성하기 위해서는 매년 개발되는 검사지 간에 ‘일관성’을 유지하는 것이 중요한데, 이를테면, 가급적 동일한 성취기준 혹은 내용 요소를 평가하고, 문항별 난이도를 일정하게 유지하는 것이다. 출제 문항 수가 많지 않은 상태에서, 검

사지 간 일관성 유지라는 명목 때문에 매년 문항을 출제하는 데 있어 운산의 폭이 좁아보니, 다양한 내용 요소들을 대상으로 평가하고 그 결과들을 구축하는 것이 근본적으로 어려운 실정이다. 그러한 어려움을 해소하는 한 방안으로, 연도별 학력 변화 추이 분석 용 문항군(群)과 교육과정의 질 관리, 즉 교육 내용 적정화 용 문항군(群)으로, 평가 문항을 이원화하여 개발하는 것을 생각해 볼 수 있을 것이다. 특히 후자의 문항군과 관련해서는 매년 평가하는 내용 요소와 난이도에 약간씩 변화를 주어 문항을 출제하고 그러한 결과들을 몇 년에 걸쳐 쌓음으로써 추후 교육과정 개정에 필요한 정보를 제공해 주는 것이다.

수행평가 문항은 수학 교육 내용에 대한 학생들의 이해 상태에 관해 심층적인 정보를 제공해 주기 때문에 교육 내용 적정화에 유용한 자료가 될 수 있다. 그런데 국가수준 성취도 평가와 같이 평가 대상 학생 수가 대규모인 평가에서는 원하는 대로 유의미하게 수행평가 문항을 출제하기가 쉽지 않다. 수행평가로부터 유의미한 결과를 얻기 위해서는 앞서 살펴본 대로 정밀한 채점기준과 그에 따른 채점자 연수, 채점 과정이 뒷받침되어야 한다. 그만큼 인력과 예산이 소요되는 것이다. 인력과 예산이 지원되지 않을 경우 수행평가는 형식적으로만 이루어지거나 포기될 수 있다. 지금까지 국가수준 성취도 평가에서는 모든 학생들을 대상으로 모든 수행평가 문항을 실시해 왔다. 그 과정에서 수행평가가 의미 있게 시행된 적도 있긴 하지만, 많은 경우 단답형과 간단한 서술형 문항에 그쳤고, 가급적 단순화된 채점기준에 따라 채점하였다. 국가수준 성취도 평가에서 수행평가의 기능을 학생들의 이해 상태에 관한 심층적인 정보 확보로 규정한다면, 모든 학생들이 모든 수행평가 문항을 풀도록 하고 있는 현재의 시스

템에서 벗어나, 인력과 예산이 허락하고 평가의 신뢰도와 타당성이 확보되는 범위 안에서, 수행 평가의 대상이 되는 학생 수를 과감히 줄여, 특정 집단의 학생들만 수행평가를 보도록 할 필요가 있다. 그리고 거기서 얻어진 평가 자료는, 연도별 학력 추이 분석과 같은 목적보다는 순전히 교육과정의 질 관리, 즉 교육 내용의 적정화 목적에만 사용하는 것이다.

수학과 교육과정의 질 관리의 핵심이라고 할 수 있는 교육 내용의 적정화와 관련하여, 국가수준 성취도 평가에서 수학 담당 연구자는 항상 교육 내용과 관련하여 쟁점이 되는 사항에 관심을 갖고, 또한 연구할 만한 가치가 있는 주제를 발굴하여, 평가문항 개발 단계에서 이를 어떻게 반영할 것인지에 대하여 고민하여야 할 것이다. 또한 국가수준 성취도 평가 결과를 분석하는 데 좀더 많은 수학교육 연구자가 동참할 수 있는 방안을 마련해야 한다. 평가 문항 개발에 참여하는 인력에 비해, 결과 분석에서는 참여자가 상대적으로 적은 편이다. 초·중·고를 망라하는 이 방대한 연구를 평가원의 국가수준 성취도 평가의 수학 담당자만으로 내실 있게 분석하는 데에는 분명히 한계가 있다. 문항별 정답률이나 수행평가의 오류 비율과 같은 데이터는 아이디어를 내고 약간의 가공을 거치면 교육과정, 특히 교육 내용에 유용한 시사점을 얻을 수 있다. 국가수준 성취도 평가 결과에 대해 심층적이고 유의미한 해석을 하고 교과교육 연구의 발전에 기여할 수 있도록, 여러 교과교육연구자들이 결과 분석 단계에서 참여할 수 있는 방안을 마련할 필요가 있다.

IV. 맺음말

이 논문에서는 국가수준 성취도 평가에서 추

구하는 여러 목적 중 특별히 ‘교육과정의 질 관리’라는 목적에 한정하여, 이 목적과 관련된다고 볼 수 있는 측면으로 네 가지를 추출하고 실제 수학과 사례를 통하여 현재 그러한 목적에 어떻게 기여하고 있는지를 살펴보았으며, 향후 ‘교육과정의 질 관리’를 내실 있게 하기 위해서 개선해야 할 점을 탐색하였다.

본문에서 제기한 ‘네 가지 측면’의 현황과 각각에 대한 개선 방향을 정리하면, 먼저 국가수준 성취도 평가에서 산출하는 성취수준별 비율 결과를 통하여 수학 교육과정에 제시된 수학 내용을 학생들이 어느 정도 이해하고 있는지를 파악할 수 있다. 이 때 사용하는 성취수준의 의미를 교과 차원에서 좀더 구조화하고 명료화하는 작업이 진행되어야 한다. 둘째, 내용 영역 간 학생들의 이해 정도를 비교할 수 있는 정보를 얻을 수 있다. 그런데 그 정보에 대한 타당성을 좀더 확보하기 위하여 내용 영역별로 출제 문항 수를 현재보다 늘리는 방안을 강구할 필요가 있다. 셋째, 국가수준 성취도 평가에서는 평가문항별로 성취수준별 정답률을 산출하고 있으며, 이 자료를 통하여 성취수준별 학생들의 특징, 특히 강점과 약점을 파악할 수 있다. 그러한 특징을 좀더 체계적으로 파악하기 위해서 문항 수를 늘리는 방안을 검토해야 할 뿐만 아니라, 평가문항 개발 단계에서부터 각 성취수준별 특징이 드러나도록 문항을 개발한다는 것을 중요한 지침으로 정하고 그 방안을 구체화할 필요가 있다. 넷째, 국가수준 성취도 평가의 자료로부터 교육 내용 적정화에 유용한 기초 자료를 얻을 수 있다. 이를 위해서 평가 문항 개발 단계에서 교육과정의 질 관리만을 목적으로 한 문항군(群)을 별도로 개발할 필요가 있으며, 수행평가로부터 유의미한 심층 분석 정보를 얻을 수 있도록, 표집 대상 학생 중에서도 일부 학생만을 대상으로 수행평

가를 실시하도록 하는 방안을 제시하였다. 또한 수학과 담당 연구자는 항상 교육 내용과 관련하여 이슈가 무엇인지에 관심을 가지고 있어야 하며, 평가 문항 개발 단계와 결과 분석 단계에서 좀 더 많은 교과교육연구자가 참여할 수 있는 활로를 마련해야 할 것이다.

1998년 국가수준 성취도 평가 방안 연구에서는, '피드백이 강화된 모니터링' 체제로서, 평가의 결과를 교육과정, 교육정책, 교수-학습 등 교육체제로 다시 송환시켜 교육의 질 개선에 기여도를 높인다는 국가수준 성취도 평가의 목적을 명시하고 있다. 여러 변화와 함께 적지 않은 시간이 흐른 지금, 본문에서 보았듯이, 국가수준 성취도 평가는 교과 수준에서 교육과정에 대한 모니터링 기능을 수행할 수 있는 기본 골격을 상당 부분 갖추고 있다고 생각된다. 이제 이러한 기본 골격에 내실을 갖추는 일이 남겨져 있으며, 그 일은 분명히 '교육과정'과 '평가'의 연계를 이념으로 설립된 한국교육과정평가원의 중요한 기본 소임이다. 이 소임을 다할 수 있도록 수학교육계에서도 국내의 대표적인 교육 프로그램 연구인 국가수준 성취도 평가에 지속적인 관심을 가지며, 기회가 닿는 대로 적극적으로 참여하여야 할 것이다.

참고문헌

- 김명숙·노국향·박정·부재율·양길석(1998). 국가수준 교육성취도 평가 방안 연구. 한국교육과정평가원 연구보고 RRE 98-8.
- 김명숙(2003). 국가수준 학업성취도 평가 연구의 주요 쟁점과 발전 방향. *교육평가연구*, 16(1), 241-264.
- 김홍원(1998). 사회과 교육의 질관리를 위한 국가 수준에서의 사회과 학업성취도 평가 방안 개선 연구. *사회과교육*, 31, 357-381.
- 나귀수 외(2001). 2001년도 국가수준 교육성취도 평가 연구. 한국교육과정평가원 연구보고 RRE 2001-5-4.
- 성효석 외(1986). 제5차 초·중학교 수학과 교육과정 시안 연구 개발. 한국교육개발원 연구보고 RR 86-37.
- 양길석(2000). *준거지향평가의 기준 설정 방법 비교 -수행평가의 논술형을 중심으로-*. 고려대학교 대학원 박사학위 논문.
- 이미경(2004). PISA 2003 결과 분석 연구. 한국교육과정평가원 연구보고 RRE. 2004-2-1.
- 이봉주·조영미·나귀수(2002). 2002년 국가수준 교육성취도 평가 연구. 한국교육과정평가원 연구보고 RRE 2003-16-2.
- 이종승(1987). *전국적 학력평가의 국제동향. 교육의 질 관리와 전국 학력평가의 의의*. 중앙교육평가원 교육평가세미나보고서 4.
- 임인재(1998). 국가수준 학업성취도 평가의 제도화 방안 연구. 서울대학교 사범대학 교육연구소.
- 임재훈·이대현(2004). 수학과 교육 내용 적정성 분석 및 평가. 한국교육과정평가원 연구보고 RRC-2004-1-4.
- 조난심(1998). 국가수준 학업성취도 평가 발전 방안. 한국교육과정평가원 학술세미나 자료집.
- 조영미(2004). 2003년 국가수준 학업성취도 평가 연구 -수학과-. 한국교육과정평가원 연구보고 RRE 2004-1-4.
- 조영미(2005). 우리나라 초등학교 6학년 학생들의 수학 성취수준별 특징 탐색. *학교수학*, 7(1), 23-45.
- 채선희·나귀수·윤준채·최성연(2003). 2003년도 OECD 학업성취도 국제 비교 연구. 한국

- 교육과정평가원 연구보고 RRE 2003-2-2.
- 황혜정(2001). 2000년도 국가수준의 중·고등학교 수학과 교육성취도 평가 연구. *수학교육학연구*, 10(2), 92-110.
- 황정규(1998). *학교학습과 교육평가*. 서울: 교육과학사.
- Allen, N. L. (1999). *The NAEP 1996 technical report*. National Center For Education Statistics.
- Lindquist, M. M. (2001). NAEP, TIMSS, and PSSM: Entangled influences. *School Science and Mathematics*, 101(6), 286-291.
- Romberg, T. A. (1992). Evaluation: A coat of many colors. In T. A. Romberg (Eds.), *Mathematics assessment and evaluation* (pp. 10-36). New York: State University of New York.

How to Make National Assessment of Educational Achievement Better for Managing National Mathematics Curriculum

Cho, Young Mi (Gongju National University of Education)

One of the main objectives of National Assessment of Educational Achievement (NAEA), which is implemented by Korea Institute of Curriculum and Evaluation(KICE), is to manage the qualities of National Curriculum. Thus we considered NAEA as the research for educational program assessment and tried to find out how to make NAEA better for achieving that objective. For the purpose of this paper we analyzed the present conditions of NAEA into four perspectives. And then according to each perspective we explored schemes for improving NAEA. In conclusion, we proposed the following; Firstly, in relation to assessing students' understanding of the mathematics contents, the meanings of Achievement Levels have to become much more clear

and concrete. Secondly, in relation to assessing for comparing student' understanding according to mathematics content areas, the number of items has to be increased for elevating validities of NAEA. Thirdly, in relation to gathering the information about properties which are shown by the students in particular Achievement, it is necessary to think about how to produce items for the purpose in the step of developing items in NAEA. Finally in relation to collecting the data which can be useful for relevance of mathematics contents, it is necessary to make a special group of items for identifying relevance of mathematics contents, and to apply performance assessments on part of the students, not whole students.

* **Key words** : National Assessment of Educational Achievement(국가수준 학업성취도 평가), national mathematics curriculum(수학 교육과정), mathematics contents(수학 교육내용)

논문접수 : 2006. 7. 5

심사완료 : 2006. 8. 4