

수학과 평가틀 비교 연구¹⁾

황우형²⁾ · 구자형³⁾

본 연구의 목적은 현재 국내외의 대규모 평가에서 사용하고 있는 평가틀을 비교 분석하여 앞으로 평가틀 연구의 기초를 구축하는데 초점을 두고 있으며, 이를 위해서 6 가지의 평가틀을 서로 비교분석하였다. TIMSS, PISA, NAEP, NECAP의 평가틀에서 볼 수 있는 특징들을 기술하였으며, 우리나라, 미국, 그리고 국제연구의 평가틀을 비교 분석하였다.

주요용어 : 평가틀, 수학과 평가틀, 행동 영역

I. 서론

교육의 질을 높이기 위해서는 여러 가지 조건들이 충족되어야 하지만 중요한 요건 중의 하나가 올바른 교육 평가이다. 학교 교육에서 평가는 학습의 질적 개선을 위한 필수 과정의 하나로서 중요한 교육적 위치를 차지한다. 그런데 실제 평가를 하는 상황에서는 시간적·공간적 제약이 있기 마련이다. 따라서 평가의 일관성 유지를 위해 상황에 크게 구애 받지 않도록 하고 교사의 주관적 판단의 개입이 최소가 될 수 있도록 하는 절대적인 기준이 필요하게 된다. 이러한 절대적 기준이 되는 것이 “평가틀”이다. 평가틀이 있을 때 균형 있는 평가가 이루어지고 있는지를 알 수 있다. 또한 올바른 평가틀은 균형 있는 평가도구 개발하는 데도 도움을 주며, 평가 내용과 방법을 계획하고 결정해주는 지침서가 된다. 평가틀의 제시로 평가 영역의 구분을 엄밀히 하는 것이 오히려 요소간의 연계성을 깨뜨리는 상황이 될 수도 있으나, 평가도구의 일관성 유지를 위해서는 틀을 설정하여 유지하는 것이 필요하다.

수학과 평가틀의 구조는 평가의 성격에 따라 다를 수 있으나, 대개는 내용 영역과 인지적 행동 영역으로 구성된 이차원 구조를 사용한다고 할 수 있다. 그런데 주어진 문항의 내용 영역과 행동 영역을 결정하는 것이나 내용 영역과 행동 영역에 맞추어 문항을 출제하는 일이 간단하지만은 않다. 예컨대, “좌표평면 위의 세 점 A(-1, 2), B(2, 4), C(5, 1)을 꼭지점으로 하는 삼각형 ABC의 넓이를 구하여라.” 이 문항은 좌표평면 위에 세 점을 나타내어 이 세 점으로 이루어지는 삼각형의 넓이를 묻는 문제이다. 내용 영역⁴⁾(제7차 교

1) 고려대학교 특별연구비에 의해서 수행되었음.

2) 고려대학교 (wwhang@korea.ac.kr)

3) 고려대학교 교육대학원 (math0215@naver.com)

육과정)의 측면에서 이 문항을 분류하는 상황에 있다고 가정해보자. ‘삼각형에 대해 묻는 문항이다’라는 관점으로 이 문항을 본다면 ‘도형’에 해당될 것이며, ‘넓이에 대해 묻는 문항이다’라는 관점으로 본다면 ‘측정’의 영역으로 분리될 수 있을 것이다. 또한 ‘좌표평면 위에 나타내는 능력을 측정하는 문항이다’라는 측면으로 해석하면 ‘규칙성과 함수’의 영역에 해당될 것이다. 실제로 이 문항은 7-가 단계의 교과서⁵⁾ ‘규칙성과 함수’ 영역에 실린 것으로, 세 점을 좌표평면 위에 나타낸 후 적당한 직사각형을 이용하여 넓이를 구할 수 있는지를 평가한다. 따라서 이 문항의 초점은 삼각형을 인식할 수 있는가 혹은 넓이를 구할 수 있는가가 아니라 ‘좌표를 찾을 수 있는가’에 있다고 할 수 있다. 제7차 교육과정에서, 중학교 이상의 측정영역에서는 넓이를 구하는 내용은 다루지 않으며 삼각형을 인식하는 것은 초등학교에서 배운 내용이므로 교육과정의 순서상 이 문항은 ‘규칙성과 함수’ 영역으로 분리되어 있는 것이다.

이렇듯 같은 문제도 교육과정상의 순서나 평가 문항의 초점에 따라서 혹은 평가 대상에 따라 내용 영역의 구분이 달라질 수 있는 소지가 있다는 것이다. 그런데 우리나라의 경우에는 교육과정에서 내용 요소 중심으로 분류하고 있을 뿐, 학생들이 습득해야 할 능력이나 인지 행동에 대해서는 별도의 규준으로 기술하고 있지 않다. 위에서 살펴본 예와 같이 교과서의 구분이 확실히 되어 있는 상황에서도 때로는 내용 영역의 구분이 모호해질 수 있다고 할 때, 분명한 영역의 구분조차 되어있지 않은 행동 영역의 경우는 그 구분이 더욱 어려울 것이다. 이러한 점에서 학생들이 길러야 할 능력에 대한 깊이 있고 체계적인 연구가 선행되어야 함은 신성균 외(2005)연구에서도 지적된 사항이었다. 이에 기존 대규모 평가에서 사용하고 있는 평가틀을 살펴보는 것은 매우 가치 있는 일이라 판단되었다.

평가틀의 제시는 전문가들의 의사소통과정을 통해 합의하여 결정이 되는 것으로 의사결정의 문제인 것이다. 국가적·국제적인 대규모의 평가에서는 평가 결과에 대한 분석의 영향이 큰 평가인 만큼 평가틀을 마련하기 위한 노력을 하고 있다. 여러 평가체제에서 제시하고 있는 국내외 평가틀에서는 교육의 동향과 평가의 특성에 맞추어 계속적으로 보안·유지되고 있는 것을 알 수 있다. 본 연구는 이러한 맥락에서 시도되는 기초연구의 하나로 현재 국내외의 대규모 평가에서 사용하고 있는 평가틀을 분석하는데 초점을 두었다. 그리고 현 교육과정에서 내용 영역은 분류하고 있으므로, 본 연구에서는 내용 영역보다는 행동 영역에 비중을 두었음을 밝혀둔다.

II. 수학 교육의 평가 동향

이 장에서는 최근 수학 교육의 경향을 잘 보여주는 연구에서 제시하는 평가에 관련된 사항을 살펴보고자 한다. 먼저 1980년대 이후 세계 수학 교육의 방향을 선도해온 NCTM에서 제시하는 평가원리를 구체적으로 살펴본다. 그리고 1970년대부터 시작된 네덜란드의 현실적 수학 교육을 바탕으로 하는 평가 연구의 대표적인 모델이라고 할 수 있는 Jan de Lange의 수학 학습 평가틀에 대해 고찰한다.

4) 현 제7차 교육과정에서는 내용영역을 수와 연산, 도형, 측정, 확률과 통계, 문자와 식, 규칙성과 함수의 6개로 나눈다.

5) 이 문항은 (주)도서출판 디딤돌 7-가 단계 p. 162에서 발췌한 것이다.

1. NCTM의 평가원리

1980년대 이후 세계 수학 교육의 방향은 ‘미국 수학교사협의회(The National Council of Teachers of Mathematics, 이하 NCTM)’가 선도해 왔다고 해도 과언이 아니다. NCTM에서는 학교수학의 방향에 대한 연구를 진행하여 1989년에 처음으로 1990년대의 학교수학의 교육의 방향을 제시하는 책자를 발간하였는데, 이것이 ‘학교수학의 교육과정과 평가 규준(Curriculum and Evaluation Standards for School Mathematics, 이하 Standards(1989))’이다. Standards(1989)에서는 평가에 직접 사용되는 평가도구를 판단하기 위한 원리를 3가지 - 일관성, 정보의 다양한 출처, 적절한 평가 방법과 사용 -로 제시하였다. 그리고 이 원리는 모든 수준의 평가에 적용이 가능하고, 학생의 수학 성취도 평가에 대한 타당성을 제공하며, 학습과정의 평가에도 적용할 수 있다고 설명한다. Standards에서 각 평가의 원리를 다음과 같이 구체적으로 제시한다(강옥기, 2000).

- (1) 원리1 (일관성) : 학습한 내용의 평가를 위해서는 평가도구가 교육과정과 일관성이 있어야 한다. 평가도구가 교육과정과 일관되지 않으면, 평가의 결과로 교육과정을 평가하는 것은 불가능 할 것이다. 또한 학교 교육은 교육과정을 기준으로 이루어지기 때문에 학교 교육의 평가는 교육과정을 잘 반영하여야 한다. 평가도구는 세 가지 측면 - ‘목적, 목표’, ‘수학적 내용’, ‘계산기, 컴퓨터, 조작 자료를 포함하는 지도 방법과 활동’ -에서 교육과정과의 일관성을 살펴볼 수 있다.
- (2) 원리2 (정보의 다양한 출처)⁶⁾ : 학생의 학습에 관련된 판정은 다양한 소재로부터 얻은 정보의 수렴을 근거로 하여 만들어져야 한다. 이 소재들은 다른 종류의 수학적 사고를 요구하는 과제 또는 같은 수학적 개념이나 절차를 다른 맥락, 문제 상황에 표현하는 과제를 포함하여야 한다.
- (3) 원리3 (적절한 평가 방법과 사용) : 평가 방법과 도구는 찾고 있는 정보의 유형, 그 정보의 사용, 학생의 발달 수준과 성숙을 근거로 선택되어야 한다.

2. Jan de Lange⁷⁾의 수학 학습 평가틀

Standards 이외에 최근 수학 교육의 경향을 잘 보여주는 연구로 네덜란드의 현실적 수학 교육을 꼽을 수 있다. 현실적 수학 교육은 H. Freudenthal(1973)의 인간 활동으로서의 수학에 기초를 둔 것으로, 현실에 바탕을 두고 학생들이 수학을 이끌어내어 조직화하는 수학화의 전 과정을 직접 경험하는 것을 강조한다(우정호, 2000). 1970년대부터 시작되어 오늘날 여러 측면에서 성과를 드러내고 있는 현실적 수학 교육의 핵심은, 학생의 현실 속에서 수학이 출발하고 머물러야 하고, 그래야만 학생은 수학의 유용성을 깨닫게 될 것이며, 학생의 마음속에 수학이 진정 중요한 것으로 자리 잡을 수 있다는 것이다(정영옥,

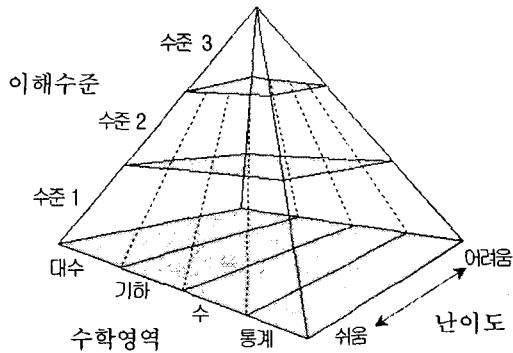
6) 여기서 ‘정보’란 문제에 사용된 정보가 아니라 평가 결과에 대한 정보를 말한다. 따라서 ‘정보의 다양한 출처’ 대신 ‘다양한 평가 방법’이라는 의미도 가능할 것이다.

7) Jan de Lange는 네덜란드의 위트레흐트(Utrecht) 대학의 정교수이며, MSEB(Mathematical Sciences Education Board)의 구성원이다. 또한 TIMSS에도 관여하였으며, 현재 PISA연구의 의장을 맡고 있다.

2004).

최근 국제적인 수학 학습 평가의 틀을 제공하고 있는 'Jan de Lange의 피라미드 모델⁸⁾'은 "수학 교육에 맞는 적절한 평가가 무엇인가?"에 대한 반성으로 시작된 것으로, RME 철학을 바탕으로 하고 있으며 OECD의 PISA 평가틀을 위한 기초를 제공하면서 계속 수정 보완되고 있다. Jan de Lange는 평가를 위해 고려해야하는 요인으로 수학내용, 이해수준, 난이도를 제시한다. 그리고 이 세 가지 요인을 삼차원으로 이루어진 피라미드 모델로 제시하여 설명한다([그림 II-1]).

참고로, 처음 de Lange가 제안한 평가틀은 삼각형의 평면모델이었다. 평면모델로 사각형이 아닌 삼각형을 제시한 것은 해당 수준에 있는 문항의 수, 배점 그리고 사용되는 시간 등을 고려한 것이라 할 수 있다. 그런데 어려운 문제가 반드시 이해수준이 높은 것은 아니며, 쉬운 문제라고 반드시 낮은 이해수준이 요구되는 것은 아니라는 지적이 제기되었다. 이에 이해수준과 난이도를 별개의 것으로 구분하여 입체모델로 수정된 것이다. 이 모델에서는 다양한 능력을 동시에 발휘하는 것을 측정할 수 있도록 개개의 능력이 아닌 이해수준에 따라 분류하였다. 이해수준에 해당하는 능력은 재생, 연결, 반성으로, 재생(reproduction)은 재생, 절차, 개념 및 정의를, 연결(connection)은 문제 해결을 위한 연결과 통합을, 반성(reasoning)은 수학화, 수학적 사고와 추론 및 일반화와 통찰력을 가리킨다.



[그림 II-1] Jan de Lange의 Pyramid 모델

III. 평가틀의 실태 분석

흔히 평가 문항을 출제하는 상황에서 교사는 자신이 문항의 특성을 바르게 이해하고 각 문항으로부터 학생의 어떠한 능력을 측정할 수 있는지를 잘 알고 있다고 판단한다. 그런데 이러한 판단의 잣대로 사용되는 것은 그동안의 교사의 관념이거나 경험인 경우가 대다수라 할 수 있다. 그러므로 교사의 가치관이나 교사마다의 경험에 의하여 동일한 문항이라 할지라도 그 문항의 성격에 대하여 판단의 결과는 다를 수 있다. 이에 대해 Dekker, T.와 Querelle, N.은 네덜란드의 수학 교육자 Jan de Lange가 제시한 평가틀을 토대로 하여 평가 설계에서 교사가 익연중에 사용하게 되는 근거나 논의를 밝히고자 하였다. 'Great assessment problems(GAP)'라는 제목의 이 책에서는 대부분의 교사들이 흔히 다음과 같이 문항을 분석한다고 Dekker와 Querelle은 서술하고 있다.

"이것은 간단한 문제라 모든 아이들이 잘 할 수 있어. 이건 충분한 연습을 했거든.", "이것은 어려운 문제지, 비록 우리가 이것에 대해 배우기는 했어도", "이것은 문제를 잘 읽어야

8) Jand de Lange의 수학 학습 평가틀은 'The great assessment picture book(2003)'을 토대로 정리한 것이다.

수학과 평가를 비교 연구

하는데, 난 이것까지 기대하지는 않아.”, “이건 정말 어려워. 대다수의 학생들이 수학적 이유를 잘 못 적으니까”, “이건 거의 모든 아이들이 못 풀 거야. 이건 정말, 다른 수학적 개념을 통합해야 하고 문제 해결을 위해 창조적인 방법을 찾아야 하거든”

그리고 Dekker와 Querelle은 위와 같이 교사들이 추측하는 것이 상당 부분 예측과 동일한 결과가 나오기는 하지만 반드시 그런 것만은 아님을 언급하면서, “존은 케비넷이 낡아서 페인트를 새로 칠하려고 한다. 이 때 빨강, 주황, 노랑, 파랑, 초록의 서로 다른 5개 색의 페인트가 있다. 이 중 존은 두 가지 색을 사용하여 케비넷을 칠하고자 한다. 존이 선택할 수 있는 두 가지 색의 쌍은 몇 가지인가?”라는 구체적인 예를 제시한다. 이 문항에 대하여 교사는 모든 종류를 나열하여 쌍의 수를 실제로 세어보는 학생이 있거나 $5 \times 4 \div 2 = 10$ (쌍)이라고 계산을 통해 구하는 학생이 있을 거라고 대부분 예상한다. 그런데 어떤 한 학생이 “그건 불가능해요. 존은 색을 섞을 수 있어요. 그래서 셀 수 없이 많아요.”라고 대답하는 경우가 있을 수 있음을 지적한다. 이런 경우 이 문항 자체가 어려운 수학적 개념을 사용했다기보다는 문항에서 사용되는 말을 제대로 이해하지 못한 것이라고 볼 수 있다고 설명한다.

이렇듯 교사의 임의적인 판단이 언제나 맞을 수 없다. 위와 같은 오해의 소지를 만들지 않도록 문항을 출제하는 것도 필요하겠으나 그에 앞서 학생의 능력을 제대로 판단하기 위한 올바른 기준이 명시될 필요가 있다. 평가의 설계 단계에서 명시된 틀이 없다면 올바른 평가를 할 수 없으며 그뿐만 아니라 올바른 결과 분석이 불가능하다. 평가틀(Assessment Framework)은 어떤 능력을 측정할 것인지, 원하는 능력을 측정하기 위해서 어떤 문항들을 어떤 비율로 구성해서 검사 도구를 만들 것인지, 학생들의 성취도는 어떤 수준으로 나눌 것인지, 각 수준이 의미하는 바는 무엇인지 등에 대한 정보를 포함한다. 그러므로 평가틀은 평가가 평가 목적과 일관성을 유지하도록 하는 데 있어서 기초가 된다.

그런데 일반적으로 학교에서 시행하는 정기고사인 중간고사·기말고사에서는 특별히 평가틀을 마련하여 시행하는 곳은 흔치 않고, 대부분 경력교사의 경험에 많이 의존하여 고사를 출제하고 있다. 교과서의 구성에서 내용 영역이 분명히 구분되어 있고, 범위를 나누어 시험을 치르게 되므로 일선 학교에서 시행되는 크고 작은 평가에서도 내용 영역에 대한 고려는 어느 정도 이루어지고 있다고 할 수 있다. 그런데 행동 영역의 경우는 제7차 교육과정에 따로 명시되어 구분하고 있지 않으며 교과서의 구성을 보면 대체로 ‘계산 → 이해 → 추론 → 문제해결 → 의사소통’을 따른다고 볼 수도 있지만 반드시 그런 것만은 아니고 이렇게 구분지은 행동 영역의 문항비율⁹⁾이 일정하게 구성되어 있다고도 할 수 없다.

이에 본 장에서는 평가틀을 구축하여 평가를 진행하고 있는 대규모 평가 6가지를 중심으로 살펴보고자 한다. 우리나라의 평가에서 제시한 평가틀, 미국에서 사용하는 평가틀, 국제 평가체제에서 사용되고 있는 평가틀을 구체적으로 살펴봄으로써 평가의 특성에 따라 어떠한 평가틀이 적용가능한지 모색하고자 한다. 우리나라의 평가에서는 ‘대학수학능력시험’과 ‘국가수준 학업성취도 평가’에서 제시한 평가틀, 미국에서는 ‘NAEP’과 ‘NECAP’의

9) 황혜정·최승현(1999)연구에 따르면 초등학교 단계에서는 ‘계산’ 요소가 ‘추론’이나 ‘문제해결’ 요소 못지않게 중요하고, 고등학교 단계에서는 계산, 이해 등을 수반하는 통합적 차원의 ‘문제해결’ 요소가 중요하다.

평가틀, 국제 평가체제에서는 ‘TIMSS’와 ‘PISA’에서 사용되고 있는 평가틀을 구체적으로 살펴보았다.

1. 우리나라

1) 대학수학능력시험¹⁰⁾

한국교육과정평가원에서 주관하는 대학수학능력시험에서는 내용 영역과 행동 영역을 구분한 평가틀을 사용한다. 수리영역시험은 시험 출제 범위에 따라 ‘가’형과 ‘나’형으로 구분되며, 각 시험의 평가 목표의 내용 영역은 시험 출제 범위에 속하는 교과목의 내용으로 하며, 제7차 교육과정에 명시된 6개의 영역 - 수와 연산, 도형, 측정, 확률과 통계, 문자와 식, 규칙성과 함수 - 으로 구분하고 있다.¹¹⁾ 행동 영역으로는 계산 능력, 이해 능력, 추론 능력, 문제해결 능력을 두고 있으며, 구체적인 내용은 다음 <표 III-1>과 같다.

<표 III-1> 대학수학능력시험의 평가틀

행동 영역		내 용
가. 계산 능력		<ul style="list-style-type: none"> • 연산의 기본 법칙이나 성질을 적용하여 주어진 식을 간단히 하는 능력 • 수학의 기본적인 공식이나 계산법을 적용하는 능력 • 수학의 전형적인 풀이 절차인 알고리즘을 적용하는 능력
나. 이해 능력		<ul style="list-style-type: none"> • 수학적 용어, 기호, 식, 그래프, 표의 의미와 관련 성질을 알고 적용하는 능력 • 주어진 문제와 관련된 수학적 개념을 파악하고 적용하는 능력 • 교과서에 나오는 기본 예제 문제나 정형화된 응용문제를 해결하는 능력 • 주어진 문제 상황을 수학적으로 표현(수학적 용어, 기호, 식, 그래프, 표 등) 하는 능력 • 수학적 표현(수학적 용어, 기호, 식, 그래프, 표 등)을 교환하여 표현하는 능력
다. 추론능력	발견적 추론능력	<ul style="list-style-type: none"> • 나열하기, 세어보기, 관찰 등을 통해 문제 해결의 핵심 원리를 발견하는 능력 • 유추를 통해 문제 해결의 핵심 원리를 발견하는 능력
	연역적 추론능력	<ul style="list-style-type: none"> • 수학의 개념·원리·법칙을 이용하여 참인 성질을 이끌어 내거나 주어진 문제의 참/거짓을 판별하는 능력 • 주어진 정의를 이해하고 참인 성질을 이끌어 내는 능력 • 반례를 들어 주어진 문제가 거짓임을 판단하는 능력 • 증명능력 - 조건 명제의 증명, 삼단논법에 의한 논리적 추론, 반례에 의한 증명, 모순법, 동치 명제의 증명, 수학적 귀납법에 의한 증명 등을 이해하는 능력 - 주어진 증명을 읽고 결론을 도출하는 능력
라. 문제해결 능력	수학내적 문제해결 능력	<ul style="list-style-type: none"> • 두 가지 이상의 수학적 개념·원리·법칙의 관련성을 파악하고 종합하여 문제를 해결하는 능력 • 두 단계 이상의 사고 과정을 거쳐서 문제를 해결하는 능력

10) 대학수학능력시험의 평가틀은 ‘2006학년도 대학수학능력시험 준비하기 자료집(www.kice.re.kr 수능 게시판 탐재)’을 중심으로 정리한 것이다.

11) 신성균 외(2005)연구에 따르면, 초등학교는 수와 연산, 도형, 측정, 확률과 통계, 관계의 5개 영역으로, 중·고등학교는 수와 식, 방정식과 부등식, 함수, 기하, 확률과 통계의 5개 영역으로 바꾸어 구분할 계획이다.

수학과 평가를 비교 연구

행동 영역	내 용
수학외적 문제해결 능력	<ul style="list-style-type: none"> • 실생활 상황에서 관련된 수학적 개념·원리·법칙 등을 파악하고 이를 적용하여 문제를 해결하는 능력 • 타교과의 소재를 사용한 상황에서 관련된 수학적 개념·원리·법칙 등을 파악하고 이를 적용하여 문제를 해결하는 능력

2) 국가수준 학업성취도 평가

국가수준 학업성취도 평가는 우리나라 현 교육과정에 기초한 평가이므로 내용 영역은 교육과정에서 구분하고 있는 것 - 6개의 영역 (수와 연산, 도형, 측정, 확률과 통계, 문자와 식, 규칙성과 함수) - 과 동일하게 구분하고 있다. 그리고 교육인적자원부 교육과정 자료인 '제7차 교육과정에 의한 성취기준 평가기준¹²⁾'에 근거하여 문항을 출제하고 있다. 국가수준 학업성취도 평가의 평가틀에서는 행동 영역의 구분을 강조하고 있지는 않는다고 할 수 있지만 그 구분은 다음 <표 III-2>와 같다. 지금껏 우리나라 수학 교과에서 주로 사용되어 왔던 '계산(또는 지식)', '이해', '문제해결'의 3개 영역과 함께, 대학수학능력시험과 미국의 국가수준교육향상평가(NAEP), 수학·과학 성취도 국제 비교 연구(TIMSS), 뉴질랜드 등에서 강조하여 다루고 있는 '의사소통'과 '추론' 영역을 적극 반영하여 구성한 것으로, 계산, 이해, 추론, 문제해결, 의사소통 5개로 구분된다.

<표 III-2> 국가수준 학업성취도 평가의 평가틀

행동 영역	내 용
계산	<ul style="list-style-type: none"> • 여러 가지 계산 능력에 관한 것 • 문제 해결에 이르기 위한 명확한 절차(알고리즘)를 능숙하게 구사할 수 있는 능력에 관한 것
이해	<ul style="list-style-type: none"> • 기본적인 수학적 개념, 원리, 법칙 및 그 관련성의 이해를 바탕으로 유의미한 개념적 사고를 형성할 수 있는 능력에 관한 것
추론	<ul style="list-style-type: none"> • 귀납, 유추, 추측에 의해 수학적 법칙과 문제의 해법을 발견할 수 있는 능력에 관한 것 • 삼단논법에 의한 연역적 추론, 반례, 간접증명법, 수학적 귀납법 등을 이용하여 수학적 명제를 이해하고 증명할 수 있는 능력에 관한 것
문제 해결	<ul style="list-style-type: none"> • 수학의 여러 가지 내용 사이의 개념, 원리, 법칙 등의 관련성이 요구되는 수학 내적 인 문제를 해결할 수 있는 능력에 관한 것 • 수학과 일상생활 및 타 교과 내용과의 관련성의 파악이 요구되는 통합교과적인(수학 외적인) 문제를 해결할 수 있는 능력에 관한 것
의사 소통	<ul style="list-style-type: none"> • 계산, 개념 이해, 추론, 또는 문제해결 영역에 관한 문제를 해결하는 상황에서 해당 내용과 관련된 수학적 용어, 기호, 문장 등을 이용하여 그 해결 과정의 근거 및 이유를 표현할 수 있는 능력에 관한 것

2. 미국

1) NAEP¹³⁾

12) 이는 초·중등학교 교육과정(교육부 고시 제1997-15호)에 의거하여 개발된 국가수준의 '절대평가 기준'이다.

NAEP(National Assessment of Educational Progress)는 우리나라의 국가수준 학업성취도 평가와 그 목적이 유사한 것으로 미국 국가교육향상평가이다. 1969년부터 실시된 NAEP은 평가 영역이나 평가 방법이 변화한 것처럼, 그 평가틀도 몇 차례 수정 보안하여 개정하였다. NAEP의 내용 영역은 5개로 구분되는데, 영역 명은 몇 차례 수정되어 2006년 현재 사용하고 있는 영역명은 수 성질과 연산(Number properties and operations), 측정(Measurement), 기하(Geometry), 자료 분석과 확률(Data analysis and probability), 대수(Algebra)이다.

<표 III-3> NAEP의 평가틀

수준	특 징
낮은 복잡도	<ul style="list-style-type: none"> · 사실, 용어, 특성 등을 회상하기 · 개념의 예를 인식하기 · 합, 차, 곱, 몫 등을 계산하기 · 동치의 표현을 인식하기 · 특정한 절차를 밟기 · 주어진 변수에 대해 방정식이나 공식에서 식을 평가하기 · 문장제를 해결하기 · 단순한 도형을 그리거나 측정하기 · 그래프, 표, 도형에서 정보를 이끌어내기
중간 복잡도	<ul style="list-style-type: none"> · 한 가지 이상의 방법으로 수학적 상황을 표현하기 · 상황과 의도에 따라 다양한 표현을 선택하고 사용하기 · 다단계를 요구하는 문장제를 해결하기 · 도형이나 문제들을 비교하기 · 풀이과정에서 단계의 정당성을 제공하기 · 시각적 표현을 해석하기 · 패턴을 확장하기 · 그래프, 표, 도형에서 이끌어낸 정보를 다단계를 요구하는 문제해결에 사용하기 · 자료와 조건이 주어지면 형식화하여 설명하기 · 단순한 주장을 해석하기
높은 복잡도	<ul style="list-style-type: none"> · 상이한 표현이 다양한 목적으로 어떻게 사용될 수 있는지 기술하기 · 다단계와 여러 의사결정을 하는 절차를 수행하기 · 절차와 개념의 유사성과 차이점을 분석하기 · 패턴을 일반화하기 · 상황이 주어지면 원래 문제를 일반화하기 · 참신한 문제를 해결하기 · 하나 이상의 방법으로 문제를 해결하기 · 해를 설명하고 정당화하기 · 해결 방법을 묘사하고 비교·대조하기 · 복잡한 상황에 대한 수학적 모델을 일반화하기 · 수학적 모델에서의 가정을 분석하기 · 연역적 주장을 분석하거나 생산하기 · 수학적 정당화를 제공하기

13) NAEP의 평가틀은 'Mathematics Framework for the 2005 National Assessment of Educational Progress'를 중심으로 정리한 것이다.

그리고 최근 변경된 평가틀에서는 ‘문항의 수학적 복잡도(mathematical complexity of items)’라는 새로운 개념을 제시한다. NAEP에서 사용하는 문항의 수학적 복잡도는 세 단계로 구분되며 Low Complexity, Moderate Complexity, High Complexity으로 명명되는데, 본 연구에서는 각각 낮은 복잡도, 중간 복잡도, 높은 복잡도로 번역¹⁴⁾한다. 복잡도의 범주에 따른 특징은 아래와 같고, 자세한 사항은 <표 III-3>에 제시하였다.

- 낮은 복잡도(Low Complexity) : 이 범주의 문항은 이전에 배운 개념·원리의 회상과 인식을 주로 요구한다. 정형적인 문항으로, 때로는 기계적으로 어떤 절차를 수행한다. 이 범주의 문항에서 독창적인 방법으로 해를 찾는 학생은 없다.
- 중간 복잡도(Moderate Complexity) : 낮은 복잡도의 문항보다는 사고의 유연성과 대안의 선택을 요구한다. 습관적이거나 특정하지 않은 한 단계 이상의 반응을 필요로 한다. 학생에게 추론과 문제 해결 전략의 비형식적 방법을 사용해서 풀이 방법을 결정하는 것, 여러 영역에서 기술과 지식을 통합하는 것을 기대하는 범주이다.
- 높은 복잡도(High Complexity) : 학생에게 요구하는 것이 가장 많은 범주로, 학생은 더욱 추상적인 추론, 계획, 분석, 판단, 창의적 사고를 해야 한다. 그리고 추상적인 생각을 세련된 방법으로 표현하기를 요구한다.

2) NECAP¹⁵⁾

(1) 개관

NECAP(New England Common Assessment Program)은 2005년에 뉴잉글랜드주에서 시작된 것으로 최근의 동향을 볼 수 있는데 살펴볼 가치가 있을 것이다. 이는 최근 시행된 것이기 때문에 우리나라에서 소개된 적이 없으므로 요약하여 평가의 개관을 설명하면 다음과 같다.

2001년 미국 정부에서 제정한 낙오학생방지법(No Child Left Behind Act)의 일환으로, 2005년부터 미국 북동부의 뉴잉글랜드(New England) 지역에서는 각 주의 교육국장을 중심으로 ‘The New England Compact(이하 NEC)’을 형성하여 ‘뉴잉글랜드주공동평가의 학년 기대치(New England Compact Assessment Program Grade-Level Expectations)’를 개발하였다. NEC에서는 뉴햄프셔(New Hampshire), 로드아일랜드(Rhode Island), 버몬트(Vermont)의 교육정부가 공동으로 각 학년의 기대되는 수준인 ‘학년 기대치(Grade-Level Expectations, 이하 GLES)’와 이러한 기대에 기초된 ‘뉴잉글랜드주공동평가(New England Compact Assessment Program, 이하 NEACP)’를 개발하고 이행하는데 중점을 두었다. 주(state) 수준에서 교육이 이루어지는 미국 교육체계에서 NEC를 형성하여 공동 평가를 마련한 주된 이유는 각 주의 활동의 교환을 확립하는데 있으며, 함께 교육 정책을 탐구하고 정보를 공유하는데 있다. 그리고 NEC의 형성으로 대규모 평가가 이루어지고 협력이 가능하게 되어 각 주에서는 상대적으로 비용의 절약이 가능해질 것이고, 이 차입자본을 다시 연구력에 투입할 것을 꾀하는 것이다.

NECAP의 대상 과목은 수학과 읽기, 쓰기이다. 미국의 공교육 체계는 낙오학생방지법이

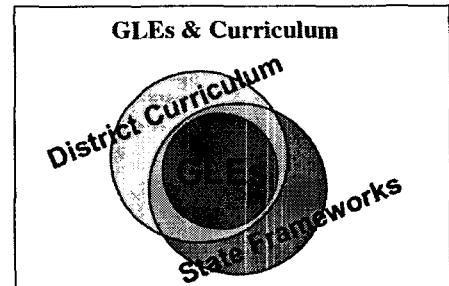
14) 이는 조영미(2004)연구에서 사용된 번역 용어를 재사용 한 것이다.

15) NEACP와 관련된 모든 사항은 NEC 사이트(<http://www.neccompact.org/>)와 뉴햄프셔 주의 교육관련 사이트(<http://www.ed.state.nh.us/>)에서 발췌한 내용을 중심으로 구성한 것이다.

제정된 이후 전반적으로 재검토되고 있다. 특히 읽기와 수학의 성취도를 제고하는 것에 중점을 두고 있는데, NEACP 역시 수학과 읽기, 쓰기를 평가 대상 과목으로 선정하고 있다.

GLEs 개발의 시작은 낙오학생방지 정책에 있었으며, 공동으로 개발된 GLEs는 교육과정 안내서 역할을 한다. 그러나 단순히 수학이나 읽기의 교육과정을 기술하데 그치지 않고, 뉴햄프셔, 로드아일랜드, 버몬트의 3개 주에서 시행되는 공동 평가(NEACP)에 대한 ‘평가틀 개발’에 보다 중점을 두었다. 따라서 GLEs는 [그림 III-1]과 같이 각 학년의 과정을 대표하는 것으로 학년 사이의 목표를 구분 짓고 각 주의 교육과정을 제한하지 않는 범위에서 평가틀을 정렬시키는 데 목적이 있다. 또한 공동으로 개발하는 만큼 교육과정을 편협하지 않게 구성되도록 하였으며, 좋은 교수법도 제공할 수 있도록 GLEs를 설계하였다. 그리고 추후 개발된 GLEs에 기초를 두고 각 주에서는 자체 교육과정을 수정, 개발할 예정이다.

이러한 평가 체제를 통해 NEC는 다음과 같이 밝히고 있다. GLEs의 개발은 체계적이고 통일된 틀을 제공함으로써 올바른 평가와 지도를 가능하게 할 것이다. 그리고 학년별로 기대치가 뚜렷이 구별되고 명확히 표현되어 있어서 학교 현장에서 학습 지도 때나 수업을 계획할 때 도움을 줄 것이다. 그리고 학습 목표가 분명해져서 학습 결과의 측정 또한 용이할 것이다. 그리고 NEC는 이를 위해 교사는 동료교사와 GLEs를 수업에 어떻게 삽입할지, 어느 정도 학생에게 지도할지를 충분히 의논할 것을 당부하고 있다.



[그림 III-1] GLEs의 설계 근원

(2) 평가틀

GLEs 개발이 주에서 공동으로 시행되는 NEACP에 대한 ‘평가틀 개발’에 보다 중점을 둔만큼 NEACP의 평가틀을 살펴보는 것은 의미있는 일이라 할 수 있다. NEACP에서는 내용 영역을 4가지로 구분하고 있는데, 수와 연산(Number and Operations), 기하와 측정(Geometry and Measurement), 대수와 함수(Algebra and Functions), 자료, 통계, 확률(Data, Statistics, and Probability)이 그것이다.

행동 영역은 수준 1부터 수준 4까지 네 단계로 구분되며 Recall, Skill/Concept, Strategic Thinking, Extended Thinking으로 명명되는데, 본 연구에서는 각각 회상, 기능/개념, 문제해결적 사고, 응용적 사고로 의역한다¹⁶⁾. 각 수준에 따른 특징은 아래와 같고, 수준 1부터 수준 3까지의 자세한 사항은 <표 III-4>에 제시하였다.

- 수준 1 (회상 : Recall) : 절차를 회상하고 사용하거나 알고리즘/공식을 적용하는 것이다. 또한 내용 기준에서 특정한 1단계 문장제와 다른 내용도 포함한다.
- 수준 2 (기능/개념 : Skill/Concept) : 모델과 설명을 통해 개념적 이해를 나타내고 정보를 분류하며 단순한 그래프로부터 자료를 해석하는 1단계 이상을 포함한다.

16) 이를 직역할 경우 수준 3은 전략적 사고, 수준 4는 확장적 사고가 될 것이지만, NEACP의 평가틀에서 제시한 하위요소에 근거하여 위와 같이 의역한 것이다.

수학과 평가틀 비교 연구

- 수준 3 (문제해결적 사고 : Strategic Thinking) : 추론하고 계획하고 증거를 사용하는 것을 포함한다. 학생들은 수준3에서 추측하고 검증하며, 복잡한 그래프에서 정보를 해석하고, 복잡한 문제를 해결하고, 개념을 설명하고, 수학적 정당화를 제공하도록 요구 받는다.
- 수준 4 (응용적 사고 : Extended Thinking) : 장기간에 걸친 복잡한 추론, 계획, 사고를 요구한다. 수학에서 수준4의 지식 깊이는 주 단위의 평가가 아니라 국소적으로 평가될 것이다.

<표 III-4> NECAP의 평가틀

수준	특 징
회상	<ul style="list-style-type: none"> • 사실, 정의, 용어를 회상하거나 인식한다. • 잘 알려진 알고리즘을 적용한다. • 공식을 적용한다. • 도형과 길이가 주어진 직사각형/삼각형의 넓이/둘레를 결정한다. • 평면이나 입체를 인식한다. • 길이를 측정한다. • 특정한 절차를 수행한다. • 표현을 평가한다. • 1단계 문장체를 해결한다. • 표나 그래프에서 정보를 도출한다. • 표현, 수(분수, 소수, 페센트)사이 또는 친숙한 측정에 대하여 회상하고 확인하고 이야기한다. • 수직선이나 좌표에 위치를 정한다. • 일차방정식을 푼다.
기능/개념	<ul style="list-style-type: none"> • 평면도형과 입체도형을 분류한다. • 단순한 그래프에서 정보를 해석한다. • 수학개념을 표현하는데 모델을 사용한다. • 다단계, 여러 개념의 적용을 요구하는 문제를 해결한다. • 그림이나 진술을 비교한다. • 도형을 비교하고 대조한다. • 풀이과정에서 단계에 대한 정당화를 제공한다. • 패턴을 확장한다. • 표, 그래프, 그림에서 정보를 뽑아 다단계를 요구하는 문제를 해결한다. • 표, 그래프, 설명을 상정적 표기로 바꾼다. • 문제상황을 상정적 표기로 다시 표현한다.
문제해결적 사고	<ul style="list-style-type: none"> • 복잡한 그래프로부터 정보를 해석한다. • 사고를 설명한다. • 추측을 하거나 이를 정당화한다. • 개념에 대해 논리적 논의를 전개한다. • 문제를 해결하는데 개념을 사용한다. • 다단계와 여러 결정으로 절차를 수행한다. • 패턴을 일반화한다. • 해결 방법을 묘사하고 비교하고 대조한다. • 복잡한 상황에 대해 수학적 모델을 형식화한다. • 수학적 정당화를 제공한다. • 답을 정당화하는 수학적 설명이 포함된 다단계 문제를 해결한다. • 문제상황을 재해석하여 상정적 표기로 다시 표현한다.

3. 국제평가

1) TIMSS¹⁷⁾

1960년대부터 시작된 TIMSS 연구의 평가틀은 평가 문항의 출제에서 뿐만 아니라 평가 결과의 분석에서 매우 큰 비중을 차지한다. TIMSS 2007에서 사용되고 있는 평가틀은 이전 TIMSS 2003의 평가틀¹⁸⁾을 보안·수정하여, 내용 영역과 행동 영역으로 구분하여 사용한다. 4학년과 8학년을 대상으로 평가하는 TIMSS 2007은 학년에 따라 내용 영역의 다르게 설정하여, 4학년의 경우는 3가지 - 수, 도형과 측정, 자료 표현 - 로, 8학년의 경우는 4가지 - 수, 대수, 기하, 자료와 가능성 - 로 구분한다.

행동 영역은 세 단계로 나누어 Knowing, Applying, Reasoning으로 명명되는데, 본 연구에서는 각각 알기, 적용하기, 추론하기로 번역한다. 행동 영역의 구분은 4학년과 8학년에 동일하게 사용되지만, 문항 출제 비율은 나이와 경험의 차이를 반영하여 다른데, 4학년의 경우 알기가 40%, 적용하기가 40%, 추론하기가 20%이고 8학년의 경우 알기가 35%, 적용하기가 40%, 추론하기가 25%로 구성되어, 고학년에 추론의 비율이 높아짐을 알 수 있다. 각 수준에 따른 특징은 아래와 같고, 자세한 사항은 <표 III-5>에 제시하였다.

- 알기(Knowing) : ‘알기’는 학생들이 수학적 문제 해결을 위해 알아야하는 사실, 절차, 개념을 포함한다. 수학을 이용하는 솜씨, 수학적 상황에 대한 추론은 수학적 지식과 수학적 개념에 친숙한 정도에 달려 있다. 학생들이 회상할 수 있는 적절한 지식이 많을수록, 이해하고 있는 개념의 범위가 넓을수록 보다 폭넓은 문제 해결 상황에 참여하고 수학적 이해를 발달시킬 가능성이 더 커진다.
- 적용하기(Appling) : ‘적용하기’는 문제를 풀거나 질문에 답하기 위한 지식·개념적 이해를 적용하는 학생의 능력에 초점을 맞춘다. 문제 해결은 학교수학의 중심 목표인 동시에 수단이다. 따라서 문제 해결과 문제 해결의 바탕이 되는 기능(예를 들면, 선택하기, 표현하기, 모델링하기)은 지식과 개념적 이해를 ‘적용하기’ 영역에서 뚜렷한 특징이 된다. 학생들은 사실, 기능, 절차에 대한 수학적 지식이나 수학적 개념의 이해를 적용하여 표현을 하고 문제를 해결할 필요가 있다. 아이디어를 표현하는 것은 수학적 사고와 의사소통의 핵심을 이루고 있으며, 동치인 표현을 만들어 내는 능력은 수학 교과에서 성공하는 데 기초가 된다.
- 추론하기(Reasoning) : ‘추론하기’는 정형적인 문제의 해결을 넘어서서 친숙하지 않은 상황, 복잡한 맥락, 다단계 문제를 해결해 내는 것이다. ‘추론하기’는 논리적이고 체계적인 수학적 사고 능력을 포함한다. 이것은 비정형적인 문제를 해결하는 데 사용할 수 있는 패턴과 규칙성에 토대를 둔 직관적인 추론과 귀납적인 추론을 포함한다. 비정형적인 문제는 학생들에게 낯선 문제 일 가능성이 많은 문제들이다. 이런 문제들은 정형적인 문제를 풀 때 필요한 것 이상의 능력을 요구한다. 비정형적인 문제는 순수하게 수학적인 문제이거나 실생활 맥락에서 나온 문제들이다. 두 가지 유형의 문제들 모두 새로운 상황에 지식과 기능을 전이하는 것을 포함하며, 추론 기능들 사이의 상호작용이 일어나는 것이 특징이 된다.

17) TIMSS 사이트(<http://timss.bc.edu/>)에 탑재된 TIMSS 2007에 사용되는 평가틀을 중심으로 정리한 것이다.

18) TIMSS 2003에서는 내용 영역을 5개 - 수, 대수, 측정, 기하, 자료 - 로 구분하였으며, 행동 영역은 4단계 - 사실과 절차지식, 개념 활용, 정형적인 문제해결, 추론 - 로 구분하여 사용하였다.

수학과 평가를 비교 연구

<표 III-5> TIMSS의 평가틀

수준	평가 요소	평가 내용
알기	1. 회상하기	• 정의, 용어, 수의 성질, 기하학적 성질, 대수적 표기법(예 $a \times b = ab$, $a + a + a = 3a$)에 대해 회상한다.
	2. 인식하기	• 수학적 대상, 모양, 수, 식을 인식한다. • 수학적으로 동치인 수학적 대상을 인식한다. (예: 같은 값을 분수, 소수, 백분율로 표현함, 보는 방향에 따라 단면이 다른 도형)
	3. 계산하기	• 사칙연산과 자연수, 분수, 소수, 정수의 혼합된 알고리즘적 절차를 실행한다. • 계산 결과를 예측하기 위해 값을 어렵다. • 기계적인 대수적 절차를 실행한다.
	4. 도출하기	• 그래프, 표 등의 자료에서 정보를 도출하고 간단히 나타낸다.
	5. 측정하기	• 측정 도구를 사용하고, 적합한 측정 단위를 사용하며, 측정값을 어렵다.
	6. 분류/정렬하기	• 공통적인 성질에 따라 대상, 모양, 수, 식을 분류하고, 그룹의 특징에 관해 올바르게 결정하며, 속성에 따라 수와 대상을 정렬한다.
적용 하기	1. 선택하기	• 정형적인 알고리즘이나 풀이 방법이 있는 문제의 해결을 위해 효과적이고 적합한 연산 방법이나 전략을 선택한다.
	2. 표현하기	• 그림, 표, 차트, 그래프에서 수학적 정보나 자료를 나타내고, 주어진 수학적 대상이나 관계에 대해 동치인 표현을 만든다.
	3. 모델링하기	• 정형적인 문제를 해결하기 위해 방정식이나 그림 같은 적합한 모델을 만든다.
	4. 이행하기	• 수학적 지시의 형태를 따라서 실행한다. • 주어진 진술에 따라 모양이나 형태를 끌어낸다.
	5. 정형적인 문제 해결	• 정형적인 문제(즉, 학생들이 수업에서 접해봤음직한 것과 유사한 문제)를 해결한다. 예를 들어, 문제 해결을 위해 기하학적 속성을 사용한다. • 자료의 여러 가지 표현을 비교하고 찾아내며(8학년), 정형적인 문제 해결을 위해 차트, 표, 그래프, 지도의 자료를 활용한다.
추론 하기	1. 분석하기	• 수학적 상황에서, 변수나 대상 사이의 관계를 도출하고 설명하거나 활용한다. • 비례적 추론을 사용한다(4학년). • 문제의 해결을 단순화하기 위해 도형을 분해한다. • 생소한 입체의 전개도를 그린다. • 3차원 도형의 변환을 시각화한다. • 같은 자료의 여러 가지 표현을 비교하고 찾아낸다(4학년). • 주어진 정보로부터 타당한 추론을 한다.
	2. 일반화하기	• 수학적 사고와 문제 해결 결과를 좀 더 일반적이고 더 폭넓게 적용할 수 있는 용어로 재진술함으로써 결과의 적용 가능한 영역을 확장한다.
	3. 종합/통합하기	• 결과를 수립하기 위하여 (여러 가지의) 수학적 절차들을 결합하고, 그 이상의 결과를 이끌어내기 위해 결과들을 결합한다. • 다양한 지식의 원리와 관계된 표현을 연결 짓고, 관계된 수학적 아이디어를 연결시킨다.
	4. 정당화하기	• 수학적 결과나 속성을 통해 명제의 참, 거짓의 정당화를 제공한다.
	5. 비정형적인 문제 해결	• 학생들이 접해 본 적이 없는 수학적 상황이나 실생활 상황 문제를 해결하고, 생소하고 복잡한 상황에서 수학적 절차를 적용한다. 예를 들어 비정형적인 문제를 해결하기 위해 기하학적 속성을 사용한다.

2) PISA¹⁹⁾

PISA 2003의 수학 평가틀은 수학적 내용, 수학적 과정, 상황과 맥락의 삼원적 요소로 구성된다. 수학적 내용은 특별한 아이디어들을 통합하여 문제를 해결할 때에 가장 중요하게 사용되고 조직되어야 하는 수학적 내용을 가리키고, 수학적 과정은 문제가 발생한 상황을 수학과 연결하고 문제를 해결하는 데에서 활성화되어야 하는 능력과 관련되며, 상황과 맥락은 문제가 제시되는 배경을 말한다.

PISA 2003의 평가틀에서는 영역간의 과도한 구분을 지양하고자 한다. ‘실제 상황’에서 수학을 할 때는 다양한 내용과 능력들이 동시에 요구되며, 개별적인 요소를 개별적인 문항으로 평가하고자 하는 시도는 수학적 소양 안에 불필요한 구획화와 인위적인 과제를 초래하게 된다. 따라서 이러한 영역 통합적 개념아래 수학적 내용인 영역은 네 가지 - 공간과 모양(space and shape), 변화와 관계(change and relationships), 양(quantity), 불확실성(uncertainty) - 로 구분한다.

그리고 수학적 과정은 수학적 능력으로 설정된 8가지 능력을 크게 세 수준으로 구분하여 사용하는데, 수학적 능력으로는 수학적 사고와 추론, 수학적 논쟁, 수학적 의사소통, 모델링, 문제 제기와 문제해결, 표현, 상징적·형식적·기법적인 언어와 조작의 활용, 보조교구와 도구의 사용을 제시한다. 이 8가지의 능력은 세 가지의 능력군 - 재생군(The reproduction cluster), 연결군(The connections cluster), 반성군(The reflection cluster) - 으로 다시 구분되는데, 해당하는 능력군에 맞는 8가지의 능력을 제시한다. 이러한 재생, 연결, 반성은 수학화의 정도에 따라 세 수준으로 구분된 것이며, 같은 수학적 능력도 능력군에 따라 그 수준이 다르게 평가된다는 것이다. 각 능력군에 따른 특징은 아래와 같고, 이를 요약하면 <표 III-6>와 같다.

- 재생군(The reproduction cluster) : 재생 능력군의 핵심어는 획득한 지식의 재생과 기계적 조작의 수행이라고 할 수 있다. 재생 문항은 학생들이 이미 학습한 내용의 회상과 인식에만 한정된 수학적 사고와 추론을 요구하며, 수학적 연산의 측면에서는 전형적으로 숙달된 기계적 절차의 적용만을 요구한다. 다시 말해서, 재생은 학생들로 하여금 자신들의 행동을 정당화하거나 수학적 논증을 구성할 것을 요구하지 않는다. 또한, 재생문항은 맥락이 수학적으로 곧바로 변형되는 기계적인 문제와 같은 간단한 문제 풀기를 포함한다.

재생에 포함되는 능력으로는, 사실에 대한 지식, 일반적인 문제 표현에 대한 지식, 동치의 인식, 친숙한 수학적 대상과 성질의 회상, 기계적 절차의 수행, 표준 알고리즘과 기능의 적용, 표준 형태의 공식과 기호를 포함하고 있는 식의 조작, 계산의 실행 등을 들 수 있다.

- 연결군(The connections cluster) : 연결 능력군의 핵심어는 숙달된 소재의 통합, 연결, 적절한 확장이라고 할 수 있다. 연결 능력에는 서로 다른 수학적 내용 영역들이나 다른 영역 통합적 개념들 사이에서 관련성을 이끌어 내어 문제를 해결하는 능력, 문제를 해결하기 위해 정보를 조합하고 통합하는 능력이 포함된다. 또한 연결은 문제 해결 전략을 선택하고 진행하는 능력, 수학적 도구를 선택하는 능력, 수학화 과정에서 다양한 방법과 다양한 단계를 사용하는 능력, 해의 의미를 해석하여 해의 타당성을 점검할 수 있는 능력과 관련된다.

연결 능력은 기계적이지는 않지만 지나치게 낯설지 않은 상황의 문제 해결에 필요한 능력이며, 재생 능력을 토대로 한다. 연결 문항에서는 학생들로 하여금 문제를 해결하는 과정에서 상황과 목적에 따라 다양한 표현 방법을 다룰 것을 요구하며, 학생들이 정의, 주장, 예제, 조건적인 주

19) PISA 사이트(<http://www.pisa.oecd.org>)에 탑재된 내용과 PISA 2003의 수학적 소양에 대한 자료(<http://www.pisa.oecd.org/dataoecd/38/51/33707192.pdf>)를 기초로 하여 정리한 것이다.

수학과 평가틀 비교 연구

장, 증명 등과 같은 수학적 용어들을 구분하고 관련지읃 것을 요구한다.

연결 능력을 지닌 학생들은, 어떤 수학적 성질에 대해 추론하고, 설명하고, 자신의 주장을 제기할 수 있어야 하며, 문제를 모델링하거나 적어도 몇 단계의 모델링 단계는 수행할 수 있어야 한다. 또한 학생들은 문제 해결 과정에서 다양한 표현을 사용하는 능력에 대한 증거를 제시할 수 있어야 하며, 문제의 내용을 변형하여 적절한 수학적 언어로 나타낼 수 있어야 한다.

- 반성군(The reflection cluster) : 반성 능력군의 핵심어는 고등 추론, 논증, 추상화, 일반화 새로운 맥락에 적용될 모델링이라고 할 수 있다. 반성 능력은 문제 해결에서 요구되거나 사용되었던 과정에 대해 학생의 입장에서의 반성을 필요로 하며, 학생들로 하여금 문제를 수학화 할 뿐만 아니라 독창적인 해결 모델을 개발할 것을 요구한다. 반성 능력을 지닌 학생들은 문제를 해결할 뿐만 아니라 문제를 제기할 수도 있어야 하며, 수학적 일반화와 주장과 증명을 제시하고 분석하고 해석하고 반성할 수 있어야 한다.

따라서 반성에 관련된 문항들은 수학적 사고, 추론, 논증에 대한 상당한 정도의 학생들의 조절력을 요구한다. 또한 수업 시간에 배운 바를 확장하여 문제를 모델링하고 해결하고 의사소통을 하는 능력, 적절한 언어를 이해하고 이를 조작할 수 있는 능력을 요구한다. 그리고 더 나아가서는 학생들로 하여금 수학과 다른 교과에서의 문제 해결과 관련된 응용 사이의 관련성을 만들 것을 요구하기도 한다.

<표 III-6> PISA의 평가틀

재생	연결	반성
<ul style="list-style-type: none">• 표준적인 표현과 정의• 기계적 계산• 기계적 절차• 기계적 문제	<ul style="list-style-type: none">• 모델링• 표준적인 문제해결의 변형과 해석• 다중적으로 잘 정의된 방법	<ul style="list-style-type: none">• 복잡한 문제의 해결과 문제의 제기• 반성과 통찰• 독창적인 수학적 접근• 다중적으로 복잡한 방법• 일반화

IV. 평가틀의 비교

본 장에서는 지금까지 살펴본 여러 문헌 연구의 평가틀을 Standards(1989)의 평가도구 판단 원리와 Jan de Lange의 평가틀에 비추어 살펴보고자 한다. 또한 각 평가틀을 비교하여 서로 간에 공통적인 기준으로 사용되는 내용은 무엇이며 두드러지는 차이점은 무엇인지를 분석하여 새로운 평가틀 개발에 기초가 되고자 한다. 그런데 평가틀을 비교한다는 것이 약간의 억지스러운 면이 분명 있다. 평가마다의 취지도 목적으로 다르기 때문에 어느 것이 더 좋은 평가틀이라고는 판단할 수 없기 때문이다. 그럼에도 불구하고 비교하는 목적은 서로 간에 어떻게 다르고 같은지 살펴보며 각 성격에 맞는 평가틀을 보안해 나갈 수 있게 하는 데 있다. 또한 학교 현장에서도 평가에 대한 올바른 이해가 되었을 때 올바른 평가가 이루어 질 수 있으므로 체계적 평가가 이루어질 수 있는 기초 자료가 되고자 함에 그 의의가 있다고 하겠다.

1. 수학 교육의 평가 동향에 따른 비교

앞에서 살펴본 평가틀에 대한 사항을 열거하는데 그치는 것이 아니라 평가틀을 보다 깊게 이해하기 위해 서로 비교해 보고자 한다. 이를 위해 II장에서 제시한 수학 교육의 평가 동향에 기초하여 III장에서 구체적으로 살펴본 평가틀을 살펴보고자 한다.

먼저 NCTM에서 제시한 Standards의 평가도구를 판단하기 위한 원리 3가지 - 일관성, 정보의 다양한 출처, 적절한 평가 방법과 사용 -에 비추어 살펴보면 다음과 같다. 첫째, 일관성에 관련하여 국내의 평가틀을 살펴보면 두 평가틀 모두 제7차 교육과정에 그 기초를 두고 있음을 알 수 있다. 미국의 경우는 주별 교육과정이 있기 때문에 국가적인 기본틀이 되는 Standards에 그 뿌리를 두고 있음을 알 수 있다. 국제 평가의 경우에는 평가의 취지에 따라 전 세계의 공통적인 과정에 기초하여 국제적인 목표를 제시하고 이에 응하는 틀을 제시하고 있음을 알 수 있다. 다시 말하자면 6가지의 평가틀 모두 각 평가의 목표에 맞고 적절한 평가도구를 개발하기 위한 틀을 설정하고 있어, 일관성을 가지고 있다고 판단된다. 둘째, 정보의 다양한 출처에 관련하여 살펴보면, 본 연구의 대상이 된 평가 중에는 국내의 대학수학능력시험을 제외하고는 모두 선택형과 서술형 모두를 다루고 있었다. 그런데 NECAP에서는 온전한 학생의 능력을 알아볼 수 있는 평가 방법인 '포트폴리오'의 방법까지 고려하여 평가틀에 명시해두고 있음을 볼 수 있다. 셋째, 적절한 평가 방법과 사용면에서 대부분의 평가는 학생에게 피드백을 주기도 하지만 이보다는 국가나 국제적인 평가흐름에 피드백을 주어 교육과정이나 교육정책에 영향을 미친다.

다음으로 Jan de Lange의 평가틀에 비추어 살펴보자. 앞서 II장에서 살펴본 바와 같이 Jan de Lange 수학 학습 평가틀의 이해수준에서는 재생, 연결, 반성의 3수준으로 그 단계를 위계적으로 나누고 있다. III장에서 구체적으로 살펴본 평가에서 PISA의 평가틀이 Jan de Lange의 피라미드 모델의 이론이 구현된 평가라 할 수 있을 것이다. 그리고 세 단계로 이해수준을 구분한 TIMSS, 미국의 NAEP, NECAP 평가틀과도 매우 유사하다는 것을 볼 수 있다. 우리나라의 경우는 행동 영역의 설정에서 '추론, 문제해결, 의사소통의 능력'이 '계산과 이해의 능력'보다는 상대적으로 우위에 있다. 그런데 추론, 문제해결, 의사소통의 영역간에는 서로 구분을 하고 있지만 이들 능력 간에는 위계적인 성향이 크게 드러나지는 않는다. 이는 행동영역의 내용명시가 모호하거나 중복된다고 하기 보다는 주어진 문항을 해결하기 위해 추론, 문제해결의 능력이 동시에 사용되는 경우가 있기 때문이라 추정된다.

2. 평가틀 간의 비교 분석

수학과 평가틀에서 제시하는 행동 영역은 수학적 사고력, 사고능력, 이해수준의 의미로 해석 가능할 것이다. III장에서 살펴본 6가지의 평가에서 사용하고 있는 평가틀에서 제시된 사고수준이 상호간에 일맥적인 계통을 지니지 않는 것도 상당수 있어 이와 같이 정리하는 것이 무리가 따르지만, 그 맥이 크게 다른 것은 아니기에 [그림 IV-1]과 같이 요약해 볼 수 있을 것이다. [그림 IV-1]에서 제시한 수준간의 경계선을 달리한 것은 해당 영역에서 제시하고 설명에 의한 것이다.

수학과 평가틀 비교 연구

우리 나라	대수능:	계산능력	이해능력	추론능력	문제해결능력	
	성취도:	계산	이해	추론	문제해결	의사소통
미국	NAEP:	낮은 복잡도	보통의 복잡도		높은 복잡도	
	NECAP:	수준 1	수준 2	수준 3	수준 4	
국제	TIMSS:	이해	적용		추론	
	PISA:	재생	연결		반성	

[그림 IV-1] 평가별 평가틀의 비교

어느 사항까지를 계산 능력, 계산, 낮은 복잡도, 수준 1, 이해, 재생이라고 명명하는지는 평가틀마다 조금씩 다름에 따라 나타낸 것이다. 구체적으로 설명하자면, NECAP의 경우는 ‘패턴을 확장한다’라고 명시된 사항이 ‘기능/개념’의 수준 2에 해당하지만, NAEP의 평가틀의 경우 유사한 내용인 ‘패턴을 확장하기’라고 명시된 사항을 ‘높은 복잡도’로 구분하였다. 즉, 다음의 문항은 NECAP에서 Practice Test²⁰⁾로 제시한 것으로, 주어진 패턴을 확장할 수 있는지를 측정하고 있다. 이는 NECAP의 평가틀에서 ‘수준 2 (기능/개념)’로 분류하고 있는데 이 문항에서 일반항 n 번째로의 일반화까지를 요구하고 있다면 이는 NAEP에서 높은 복잡도(3수준)으로 분류될 것이다. 이러한 이유로 NECAP의 수준 2의 막대 길이가 NAEP의 높은 복잡도의 막대까지 걸쳐있게 나타낸 것이다.

[NECAP] (Practice Test - 2)

Look at this pattern.

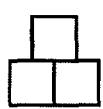


Figure 1

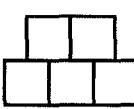


Figure 2

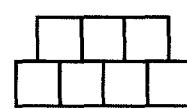


Figure 3

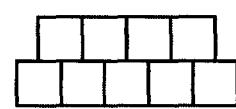


Figure 4

Which expression represents the number of boxes used of make Figure 99?

- A. $99 + 98$
- B. $99 + 1 + 99 + 1$
- C. 2×99
- D. $99 + (99 + 1)$

20) NECAP는 2004년에 GLE의 기초를 마련하고 2005년 시행되기 시작하였기 때문에, 본 검사의 문항이 아직 공개되지 않은 상태이다. 이에 본 검사를 실시하기에 앞서 시행된 8학년의 수학 예비검사 (Student Practice Test)의 문항으로 예시문항을 선정하였다.

이렇듯 유사한 사항을 같은 단계에 제시하기도 하였고 평가틀에 따라서는 같은 사항을 다음 단계의 영역에 해당시키기도 하였다. 이에 따라 그 구분선의 길이를 달리 표현하여 도식화 한 것이다. 이러한 내용을 바탕으로 네 단계로 나누어 모든 평가에서 제시하고 있는 설명을 모두 정리하여 다음 <표 IV-1>에 제시하였다. 이는 모든 평가에서 다룬 내용만을 정리한 것이 아니라 중복되는 내용을 기준으로 하되, 어느 것 하나에서라도 ‘명시’하고 있는 내용은 모두 첨가하여 정리한 것이다. 그리고 각 수준의 내용 요소에 대한 이해를 돋기 위해 각 수준에 따라 출제된 몇 개의 문항을 소개²¹⁾하였으며, 제시한 문항은 다음 장의 결론에서 밝히고 있는 비교분석한 특징을 이해하는 데에도 도움이 될 것이다.

<표 IV-1> 평가틀의 수준별 평가내용

1 수준	2 수준	3 수준	4 수준
<ul style="list-style-type: none"> • 회상 • 개념인식 • 동치인식(분수, 소수, 백분율) • 알고리즘-절차 • 계산 • 측정 • 자료정보도출 • 문제이해 • 이행하기 	<ul style="list-style-type: none"> • 표현하기(모델링) • 표현선택 • 자료이용 • 풀이정당화 	<ul style="list-style-type: none"> • 패턴 • 유추 • 분석 • 종합 • 문제해결 • 일반화 • 추측·수학적 정당화 	<ul style="list-style-type: none"> • 포트폴리오

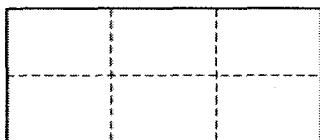
아래 3개의 문항은 “1수준”的 문항으로, 각각 NAEPI의 ‘개념인식’, TIMSS의 ‘동치인식’, NECAP의 ‘자료정보도출’에 해당된다. 순서대로 첫번째 문항은 주어진 사각형의 1/3을 표시할 수 있는지를 묻는 문항으로 분수의 개념의 예를 인식할 수 있는가를 측정하며, 두 번째 것은 비율개념을 적용하여 실제 상황의 문제를 해결할 수 있는지를 질문하는 문항이다. 5분의 단축시간을 수학적으로 동치인 수학적 대상, 백분율로 표현할 수 있어야 한다. 그리고 세 번째 문항은 시간에 따른 거리의 그래프를 보고 누가 가장 빠른지를 파악하는 문제로, 그래프에서 정보를 도출하는 능력을 평가한다.

21) 지면 사정으로 각 수준별로 두 개 정도의 대표 문항만을 제시하였다. 각 평가에 따른 상세한 대표 문항은 구자형(2006)연구를 참조하기 바란다. 본고에서는 그동안 소개된 적이 없었던 NECAP의 문항을 주로 수록한다.

수학과 평가를 비교 연구

[NAEP] (2005 - 8M3 - 9)

Shade $\frac{1}{3}$ of the rectangle below.



[TIMSS] (M022139)

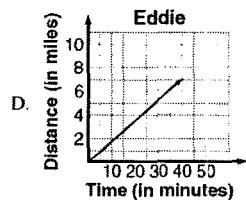
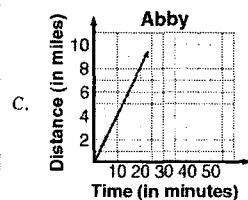
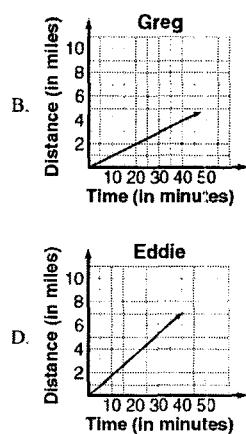
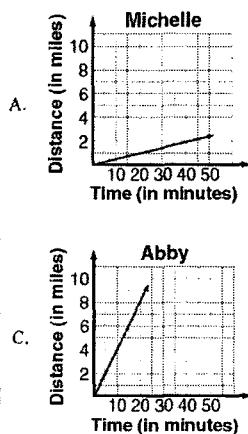
새 고속도로가 건설되면 어느 한 도시에서 다른 도시까지 버스로 가는데 걸리는 시간이 25분에서 20분으로 단축된다. 두 도시 사이를 여행하는 데 걸리는 시간은 몇 % 단축되겠는가?

- ① 4% ② 5% ③ 20% ④ 25%

[NECAP] (Practice Test - 3)

Four students rode to school on their bikes. The distance that each student traveled over time was graphed.

Which graph represents the student who rode fastest?



다음으로 제시한 아래의 문항은 “2수준”의 문항으로, NAEP의 ‘표현선택’, NECAP의 ‘풀이 정당화’에 해당한다. 전자는 주어진 방정식을 통한 그래프를 해석하여 적절한 표현을 선택하고 사용하는 능력을 측정하는 문항이며, 후자는 전형적인 시간-거리-속도의 문항으로 다단계, 여러 개념의 적용을 요구하는 문제를 해결하는지를 측정하는 문제라고 할 수 있다.

[NAEP] (2005 - 8M3 - 10)

In the equation $y = 4x$ if the value of x is increased by 2, what is the effect on the value of y ?

- A) It is 8 more than the original amount.
 B) It is 6 more than the original amount.
 C) It is 2 more than the original amount.
 D) It is 16 times the original amount.
 E) It is 8 times the original amount.

[NECAP] (Practice Test - 7)

Sean and Barry are participating in a 30-mile bike race. Sean averages 20 miles per hour and Barry averages 15 miles per hour. How many miles will Barry be from the finish line when Sean finishes the race? Show your work or explain how you know.

[NECAP] (Practice Test - 12)

Write four different numbers for which both of these state statements are true.

- The range is 8.
- The mean and median are the same number.

[NAEP] (1996 - 8M3 - 12)

Jaime knows the following facts about points A, B, and C.

- Points A, B, and C are on the same line, but might not be in that order.
- Point C is twice as far from point A as it is from point B.

Jaime concluded that point C is always between points A and B.

Is Jaime's conclusion correct?

Yes No

In the space provided, use a diagram to explain your answer.

마지막으로 제시한 위의 문항은 “3수준”의 문항으로, NECAP의 ‘문제해결’, NAEP의 ‘수학적 정당화’에 해당한다. 전자는 문제를 해결하는데 개념을 사용해야 한다. 즉 평균(mean)과 중앙값(median)의 개념을 확실히 이해하고 있어야만 주어진 조건에서 1, 3, 5, 7의 서로 다른 네 수를 구할 수 있는 문항으로 다양한 해결 전략을 볼 수 있는 개방된 참신한 문제이다. 후자는 점 A, B, C가 한 직선에 있으며, CA의 길이가 CB의 길이의 2배일 때 점 C의 위치를 결정하는 문제로, 연역적 주장을 분석하거나 생산하기 항목으로 분류될 수 있을 것이다.

V. 요약 및 결론

현재 대규모 평가에서 사용하고 있는 행동 영역의 평가들을 분석하는데 초점을 둔 이 연구에서는, 국내외의 6가지의 평가들을 서로 비교 분석하여 수학 학습 평가들에 대한 이해를 돋는 것을 목표로 하였다. 각 평가의 특성과 함께 평가들에 대해 열거한 후, 6가지의 평가들을 비교 분석하여 <표 IV-1>와 같이 공통된 성질을 이끌어 내었다. 또한 분석과정의 결과를 통해 다음과 같은 아홉 가지 사항을 이끌어 낼 수 있었다.

첫 번째, 우리나라의 평가들에서는 회상이나 개념인식, 동치인식을 따로 명시하고 있지 않는 점에서, 낮은 수준의 개념을 계산, 식을 간단히 하기 정도로 인식하고 있다고 볼 수 있다. 우리나라의 평가에서 회상이나 인식에 관계된 문항을 다루지 않는 것은 아니지만 다른 평가들에서 명시하고 있음은 보다 가중치를 둔 사항으로 해석될 수 있으며 평가문항 개발에도 영향을 미치게 될 것이다. 2장에서 최근의 평가 동향에서도 살펴본 바와 같이 수학적 개념은 무엇보다 중요한 사항이며 수학적 사고의 기초라 할 수 있다. 또한 Jan de Lange의 평가들의 이해수준에서도 재생영역에서 재생과 절차 뿐 아니라 ‘개념 및 정의’라

수학과 평가를 비교 연구

는 항목을 명시하여 중요성을 강조하고 있다. 이렇듯 상기력, 동치개념, 알아차리는 기초적인 힘도 간과해서는 안되는 수학적 능력인 점에서 우리나라의 평가를 측면에서 반성해보아야 할 사항이라 판단된다.

두 번째, 국내에서는, 또한 측정이나 자료의 정보 도출능력을 따로 항목으로 두고 있지 않다. TIMSS 2007의 수정된 평가틀에서는 ‘측정하기: 측정도구를 사용하고, 적합한 측정 단위를 사용하며, 측정값을 어림한다’라고 명시하고 있는데, 이는 뉴이(J. Dewey)의 관점이 크게 반영된 것으로 보인다. 우리나라의 경우에도 제7차 교육과정에서 ‘측정’이라는 내용 영역명을 새로 도입·구분하여 지도하는 만큼 그 중요성은 인식하고 있으며 또한 전혀 평가하고 있지 않는 것은 아니다. 그러나 아직까지 평가틀에 따로 명시하고 있지는 않다는 점이다.

세 번째, TIMSS의 평가틀의 기술 형식에 특기할만한 점이 있다. TIMSS의 평가틀에서는 행동 영역을 3단계로 분리하고, 그 아래 5~6개의 하위 항목을 한 단어로 명명해두었다. 각각의 하위 목록에 서술된 내용의 강조사항에 해당되는 내용을 대표할 수 있는 한 단어로 명명해둔 것이다. 내용을 기술하는 측면에서 한 단어로 명시하는 것도 염두에 두어볼만한 사항이라 판단된다.

네 번째, PISA의 평가틀의 특기할만한 사항은, 세 가지(재생, 연결, 반성)의 능력군 설정에 앞서서, 8가지의 능력을 제시한다는 점이다. 모든 능력을 모든 수준에서 평가 가능한 틀이라는 점에서 주목할 만한 사항이다.

다섯 번째, 정보의 출처까지 평가틀에서 강조하여 다루는 평가는 PISA의 연구이다. 상황이라는 카테고리 하에 개인적, 교육적, 직업적, 공적, 학문적으로 나누어 문항을 특성을 분류하고 있다. 5가지의 상황에 따라 문항을 구분함으로써 각 영역에 따른 소재를 다양화 할 수 있다는 점에서 가치가 있다고 하겠다.

여섯 번째, 다단계를 요구하는 문장제를 해결하기의 경우는 거의 가장 높은 수준으로 분리되고 있다. 그런데 미국의 경우는 다단계 문장제에서 풀이과정 절차를 수행해야 하는 경우는 수준 3으로 보고 있으나 답만 구하는 정형적인 형태인 경우는 수준 2로 분리한다는 것이 특색 있었다. 자신의 풀이를 정당화하는 과정이 없는 문항으로는 수준 3의 해결력을 가지고 있다고 판단할 수 없다는 이유인 것으로 해석된다.

일곱 번째, 추론하는 능력을 요구하는 패턴을 확장하는 문제에서 NAEAP의 경우는 높은 복잡도에, NECAP의 경우는 기능/개념 단계에 해당한다. ‘패턴을 확장하기’에서 확장의 범위가 어디까지냐에 따라 그 기준이 달라질 수도 있음을 알 수 있다. 따라서 평가틀에서 패턴을 확장한다고 명시할 때는 확장의 범위까지도 요구된다 하겠다.

여덟 번째, ‘해결 방법을 묘사하고 비교하고 대조한다’라고 명시하고 있는 것은 미국의 평가뿐이다. 다른 평가보다는 문제해결력, 다양한 해결방법을 강조하여 항목으로 설정해둔 것으로 보인다.

아홉 번째, 정당화의 능력이 단순히 수준이 높은 학생에게만 요구되는 것은 아니라는 점이다. 미국의 NAEAP의 중간 복잡도, NECAP의 기능/개념 수준에서 ‘풀이과정에서 단계의 정당화를 제공한다’라고 기술 되어있음을 볼 때, 단지 표현해내는 정도의 차이가 있는 것이며, 낮은 수준에서부터 숙달되는 것이 필요할 수 있다고 판단된다.

이러한 분석 결과가 학교현장에서의 수학 학습 평가틀을 계획하고 수준별 수업에 따른 평가를 계획할 때 기초 자료가 될 수 있을 것이다. 본 연구는 어떠한 평가가 올바른가가 아니라 평가틀의 비교에 그 목적이 있었으며 평가방법의 비율적인 부분은 추후 검토해볼

만하고 판단된다. 이 논문에서는 여러 평가제도의 평가틀을 체계적으로 정리·비교하는 데 초점을 두었다. 왜냐하면 평가에 관한 연구는 이루어지고 있으나 ‘평가틀’ 자체에 대한 논의는 그다지 활발하지 않고, 2000년 이후 많은 평가제도에서 평가틀을 수정·보안 하였으므로 이 논문의 내용이 이후 평가틀에 관련한 연구를 수행할 때 기초 자료로써 충분히 가치가 있을 것으로 생각되었기 때문이다. 한편 내용을 정리하는 데 치중하다 보니, 몇 가지의 분석결과는 이끌어 내었으나 비판적으로 고찰하는 부분이 부족하기도 하였으며, 새로운 평가틀을 마련하여 제시·사용해보지 못하였다. 이 논문의 내용을 기초로 학교 현장에서 사용 가능한 평가틀에 대해서는 별도의 연구가 진행되어야 할 것이다.

참고문헌

- 강옥기 (2000). 수학과 학습지도와 평가론. 경문사.
- 교육부 (1998). 수학과 교육과정. 서울: 대한교과서주식회사.
- 나귀수 · 이봉주 · 한경혜 · 최석진 · 김경희 · 양명희 · 이명희 (2002). 2002년 국가수준 학업 성취도 평가 연구(I) - 수학 -. 한국교육과정평가원 연구보고 RRE 2002-1-4.
- 박정 · 정은영 · 김경희 · 한경혜 (2004). 수학 · 과학 성취도 추이변화 국제비교 연구 - TIMSS 2003 결과 보고서-. 한국교육과정평가원 연구보고 RRE 2004-3-2.
- 박정 · 최석진 · 이미경 · 김경희 · 정은영 · 꽈영순 · 민경석 (2004). 학업성취도 국제비교연구 성과와 과제 - 2003 TIMSS와 OECD/PISA 결과를 중심으로. 한국교육과정평가원 연구자료 ORM 2004-26.
- 박정 · 정은영 · 김경희 · 한경혜 (2004). TIMSS 2003 공개문항 분석 자료집. 한국교육과정 평가원 연구자료 ORM 2004-27.
- 신성균 · 황혜정 · 김수진 · 성금순 (1992). 교육의 본질 추구를 위한 수학 교육 평가 체계 연구(III) - 수학과 평가 도구 개발 -. 한국교육개발원 연구자료 RM 92-5-2.
- 신성균 · 고정화 · 권점례 · 박선화 · 이대현 · 이봉주 · 최승현 · 조영미 (2005). 수학과 교육과정 개선 방안 연구. 한국교육과정평가원 연구보고 RRC 2005-6.
- 우정호 (2000). 수학 학습-지도 원리와 방법. 서울대학교 출판부.
- 이준열 · 장훈 · 최부림 · 남호영 · 이상은 (2005). 수학 7-가. 디딤돌.
- 정영옥 (2004). RME의 수학학습 평가틀에 대한 고찰 - Jan de Lange의 수학 학습 평가 틀을 중심으로-. 대한수학교육학회지 수학교육학연구, 14(4), 347-366.
- 조영미 · 이대현 · 이봉주 (2004). 2003년 국가수준 학업성취도 평가 연구 - 수학 -. 한국교육과정평가원 연구보고 RRE 2004-1-4.
- 조영미 (2004). 미국의 NAEP 수학 평가틀에 관한 고찰. 교육과정평가연구, 7(1), 107-130
- 한국교육과정평가원 (2001). 제7차 교육과정에 의한 성취기준 평가기준 - 중학교 수학 -. 교육인적자원부 교육과정 자료 114.
- 한국교육과정평가원 (2002). 제7차 교육과정에 의한 성취기준 평가기준 - 고등학교 수학 -. 교육인적자원부 교육과정 자료 134.
- 황혜정 · 최승현 (1999). 수학과 평가틀에 관한 고찰. 대한수학교육학회지 수학교육학 연구, 9(2), 459-471.
- NAGB. (2004). Mathematics Framework for the 2005 National Assessment of

수학과 평가를 비교 연구

- Educational Progress. [On-line] available <http://www.nagb.org>
- NCTM. (1989). Curriculum and Evaluation Standards for School Mathematics. VA: NCTM.
- De Lange, J. (1999). Framework for classroom assessment in mathematics. Utrecht: Freudenthal Institute & National Center for Improving Student Learning and Achievement in Mathematics and Science.
- Dekker, T. & Querelle, N. (2003). Great assessment problems(GAP). Utrecht: Freudenthal Instituut.
[On-line] available http://www.fi.uu.nl/catch/products/GAP_book/intro.html
- Freudenthal, H. (1973). Mathematics as an Educational Task. Dordrecht: D. Reidel Publishing Company.
- IEA (2005). TIMSS 2007 Assessment Frameworks.
[On-line] available http://timss.bc.edu/TIMSS2007/PDF/T07_AF_chapter1.pdf.

황우형 · 구자형

A Comparison Study on Mathematics Assessment Frameworks

Whang, Woo Hyung²²⁾ · Ku, Jahyung²³⁾

Abstract

The purpose of the study was to construct future research basis for large scale assessment framework by analyzing current domestic and international large scale assessment frameworks. Six frameworks were compared and analyzed. The characteristics of TIMSS, PISA, NAEP, NECAP were described in detail, and three frameworks, Korea, the United States, and International, were compared and analysed.

Key Words : Assessment Framework, Mathematics framework, Cognitive domain

22) Korea University (wwhang@korea.ac.kr)

23) Korea University, Graduate School (math0215@naver.com)