

## 수학과 학업성취도 평가를 위한 수행평가의 과제와 전망

성 태 제 (이화여자대학교)

권 오 남 (이화여자대학교)

### I. 서 론

교육은 인간의 행동을 변화시키는 데 목적을 두고 있다. 교육을 뜻하는 그리스어의 Pedagogy의 어원은 paidagōgos로써 paidos와 agōgos를 합성하여 '어린이를 이끈다'를 의미한다. 라틴어의 Education은 e와 duco의 합성어로 '밖으로 꺼내다'의 의미를 지니며, Simpson(1987)은 draw out, lead out, raise up, bring up, rear a child로 해석한다. 김정환·강선보(1997)는 교육을 한글의 語義로 '가르치다'와 '기르다'의 합성어로 분석하며 교육은 피교육자의 발전 가능성, 교육자의 교도 훈련성, 그리고 인격적 매개성을 전제로 하고 있다고 주장한다.

교육은 학생들의 무한한 잠재성과 가르치고 배우는 자와의 인격을 매개로 자연적 상황보다는 인위적 상황에서 이루어지기 때문에 목적을 추구하는 행위이다. 즉 교육은 인간의 행동 특성인 인지적 특성, 정의적 특성, 그리고 심동적 특성의 행동 변화를 유도하여 교육목표에 도달하게 하는 데 목적이 있으며 목표 달성 여부를 판단한다. Tyler(1992)는 교육목표의 달성 여부를 판단하는 행위를 교육평가라 하였다. 현대 교육에서 교육평가는 교육과 관련된 모든 것에 대한 가치 판단으로 의미가 확대되고 있지만 교육평가는 학습과정에 도움을 주어 학습을 극대화하는데 목적을 두고 있다.

교육평가 방법은 평가의 내용과 목적에 따라 매우 다양하다. 특히 가르치고 배운 내용을 얼마만큼 알고 있는지를 평가하는 학업성취도 평가 방법은 학업 내용과 특성, 그리고 목적, 교육적 상황에 따라 다양하다. 최근에는 '새 물질 교육'이니 '새 교육문화 창조'라는 구호 아래 수행평가에 대한 관심이 고조되고 있다. 성태제(1998a)는 적지 않은 교원연수와 연구

보고서에서 선다형이 아닌 모든 평가방법을 수행평가라 하고 있으며 수행평가가 모든 상황에서 바람직하다는 오해가 있음을 지적하고 있다.

본 연구는 학업성취도 평가 방법과 변화, 관심의 대상이 되고 있는 수행평가 특히 수학교육을 위한 수행평가의 정의, 특징, 장단점, 그리고 수행평가의 과제와 전망에 대하여 설명한다.

## Ⅱ. 학업성취도 평가방법의 변화

교육의 대상인 인간의 특성은 인지적 특성, 정의적 특성, 심동적 특성으로 분류되며 각 특성을 평가하는 일반적 방법은 <표 1>과 같다.

<표 1> 인간의 세 가지 특성 평가 방법

특성	평가 방법
인지적 특성	대화, 구술, 지필검사(논술, 단답, 괄호, 진위, 배합, 선다), 실험·실습
정의적 특성	자연 관찰, 관찰 기록, 면접, 관찰 점검표, 자기 서술, 질문지
심동적 특성	관찰에 의한 수행과정 평가, 지필검사

인지적 능력을 측정하기 위해 최초로 실시한 방법은 대화였다. 종이도 연필도 없었을 테니, 그리고 세부적으로 묻는 것보다는 거시적 질문에, 제자들이 여유 있게 알고 있는 모든 것을 몇 날을 두고 대답하였을 것이다. 그러나 많은 제자와 제한된 시간 때문에 대화에 의한 평가보다는 구체적 질문을 부여하고 그에 맞추어 대답하는 구술의 형태로 변화하였을 것이다. 대화와 구술은 기록하여 보관할 수 없는 문제점 때문에 파피루스가 만들어지면서 논술형의 질문이 사용되었고, 논술형 문항에 의한 지적 능력 평가는 채점의 신뢰도 문제를 야기한다.

산업사회로 진입하면서 경쟁은 평가 결과의 타당도보다는 신뢰도를 강조하게 되고 신뢰도를 높이기 위하여 구조적인 형태의 평가 방법을 택하게 되었으며 행정적 기능을 강조하는 평가일 경우는 일반적으로 선택형 문항에 의하여 지적능력을 측정한다. 즉 경쟁이 치열한 사회일수록 인간의 능력을 측정할 때 거시적 평가보다는 미시적 평가를 실시하게 된다. 선택형 문항에 의한 미시적 평가는 개인의 특성을 전체적으로 조망하기 어렵고 정답 여부에 관심이 있으므로 문제해결 과정을 측정하지 못하는 단점이 있다. 그러나 많은 경쟁 대상, 상대비교에 의한 판정 등의 목적을 달성하기 위해서는 선택형 문항에 의한 지적능력 평

가가 체계적이고 신뢰할 수 있고 효율적이기 때문에 보다 널리 사용된다. 미국의 경우도 1930년대 말부터 대단위 형태의 검사 프로그램이 구술시험이나 서술시험에서 표준화 형태의 선다형 문항의 검사로 대체되기 시작하였다(Mislevy, 1997).

심동적 특성 평가는 어떤 행위를 전개하는 과정과 결과 모두를 평가하는 수행평가를 사용하여 왔다. 수행평가(performance assessment)는 학습한 내용이나 기술을 얼마나 잘 수행하는지, 즉 행위로 드러나는지를 평가하는 방법으로, 예를 들어 발레나 다이빙 자세 등을 평가할 때 사용된다. 심동적 특성 평가를 위한 수행평가는 **앎(knowing)**보다는 **행함(doing)**에 초점을 두고 있다. 또한 예전에는 수행평가에서 수행의 범위를 제한하지 않았으며 실제 생활에서 수행하는 정도를 평가하였으나 평가 실시의 한계 때문에 자연적 상황보다는 인위적 상황, 그리고 수행범위를 제한하는 경향으로 흐르고 있다.

아는 것과 행위 하는 것이 함께 요구되는 지식산업사회에서 인간에 대한 평가는 아는 정도는 물론 행위 하는 수준까지 실시하게 된다. 예를 들어 워드프로세서를 사용하는 방법을 아는 정도와 실제로 워드프로세서로 문서를 작성하는 정도를 평가하는 것이 총체적으로 능력을 평가할 수 있기에 수행평가를 교육의 모든 영역에 적용하는 추세이다.

Herman, Aeschbacher & Winter(1992)는 선다형 문항에 의한 평가방법을 전통적인 평가방법이라 하고 이에 대응되는 평가방법을 대안적 방법이라 하여 <표 2>와 같이 두 가지 평가 방법을 비교하였다.

<표 2> 전통적 평가방법과 대안적 평가방법

특 성	전통적 방법	대안적 방법
학습관 학습자관	학습결과에 관심 수동적 관점 분리된 지식과 기술을 평가	학습과정과 결과 관심 능동적 관점 통합된 지식과 기술평가 메타인지적 관점
평가형태	지필검사	수행평가 참평가 포트폴리오
평가실시	일회적 평가	지속적 평가
평가내용	단일 속성	다원적 속성(여러 측면)
평가대상	개인 평가 강조	집단 평가 강조(협동)

선택형 문항 위주의 전통적 평가방법은 문제 풀이 과정을 알지 못하고 결과만 알 수 있으며 지필검사에 의존하고 있는 특징이 있다. 반면 대안적 평가방법은 학습과정을 알 수 있어 메타인지적 관점에서 학습자를 보며 평가 방법으로 수행평가, 참평가, 그리고 포트폴리오를 예로 들고 있다. 전통적 평가방법은 단일 속성을 평가하는 반면 대안적 평가방법은 문제를 풀어 가는 과정 중에 필요한 모든 능력을 평가하는 특징을 지니며 평가과제가 다양한 정보를 다루어 복합적인 능력과 기술을 요구하는 경우가 많기 때문에 집단을 단위로 과제가 제공될 수 있다. 이러한 경우는 개인평가보다는 공동으로 문제해결에 참여하는 집단평가를 강조한다.

앞에서 설명한 것처럼 지적능력을 평가하는 최초의 방법은 대화나 구술이었고 과학교과인 경우는 실험·실습이 주된 평가방법이었다. 그러나 평가의 편리성과 객관성이 강조되면서 평가방법은 선택형 위주의 지필검사를 실시하게 된다. 지필검사 중 선택형 위주의 평가방법이 지난 수십 년 지적 능력을 평가하는 방법으로 사용되었으나, 다시 수행평가, 포트폴리오 등이 새로운 대안적 평가방법으로 사용된다.

### Ⅲ. 수행평가의 개념

#### 1. 수행평가와 수학과 수행평가의 정의

교육평가 방법의 순환적 특성을 이해하지 못하고 선택형 문항에 의한 지필검사를 전통적 평가방법으로 규정하고 그 외 다른 모든 평가방법을 수행평가로 정의하여, 우리 나라에서 실시하는 선다형 문항 위주의 평가방법은 모두 잘못된 것이므로 수행평가를 실시해야 한다는 주장은 최소한 두 가지 이유에서 오류를 범한다. 첫째, 수행평가의 정의에 대하여 올바르게 이해하고 있지 못하기 때문이고, 둘째, 수행평가와 선택형 문항에 의한 평가의 기능과 목적이 다름을 이해하지 못하기 때문이다. 앞에서 설명한 것과 같이 수행평가는 전혀 새로운 평가방법이 아니라 심동적 영역의 행동 특성을 평가하기 위하여 사용되던 평가방법을 인지적 영역 행동 특성을 평가하기 위하여 인지적 행동 특성에 맞게 평가 방법을 적용한 것이다. 그러므로 인지적 영역에서의 수행평가는 기존의 선택형 검사에 근거한 평가방법을 보완하는 측면에서 해석하는 것이 옳다. 남명호(1995)도 수행평가는 관련 있는 과제에 대해 학생들의 실제적인 수행(performance)을 검사하고 판단하는 평가로써 최근에 처음 사용되는 평가방법이 아니라고 하였으며, 오래 전부터 실시되었던 예·체능계의 실기평가와 과학실험평가, 그리고 의과대학의 실습 등이 대표적인 수행평가의 예라고 하였다.

성태제(1998b)는 수행평가와 혼용되어 사용되고 있는 참평가, 포트폴리오, 직접평가, 대안적 평가에 대한 용어를 정리하여 <표 3>과 같이 설명하고 있다.

<표 3> 수행평가와 관련된 용어 정의

용어	정의
수행평가	지식이나 기능, 혹은 기술의 수행정도를 측정하는 평가
참평가	실제 상황에서 수행정도를 측정하는 평가
포트폴리오	개인의 작업이나 작품을 모아둔 자료집이나 서류철
직접평가	표출되는 행위를 직접 관찰에 의하여 실시하는 평가
대안적 평가	기존의 어떤 평가 방법을 대치할 수 있는 평가

수행평가(performance assessment)는 배운 내용이나 지식, 그리고 습득한 기술이나 기능을 행위로 나타내는 정도를 측정하여 판단하는 평가방법을 말한다. 참평가(authentic assessment)는 실제 생활에서 실시되는 수행평가로써 평가의 목적에 따라 인위적으로 고안되지 않는 특징을 지니고 있다. 그러므로 Mayer(1992)는 교육 현장에서 의미 있는 평가 수행평가와 참평가를 구분해야 한다고 주장한다. 정의에 의하면 참평가는 모두 수행평가가 될 수 있으나 모든 수행평가가 참평가가 될 수 없다. 포트폴리오(portfolio)는 작업 결과나 작품 혹은 어떤 수행의 결과를 모아 놓은 자료집이나 서류철을 보고 평가하는 방법으로 자료집에 넣을 작업 결과를 학생이 스스로 선택한다. 그러므로 학생이 일차적으로 평가의 주체가 될 수 있으며 오랜 시간의 수행결과를 평가할 수 있으므로 평가대상의 수행 정도뿐 아니라 발달 과정까지 평가할 수 있는 장점을 지닌다. 직접평가(direct assessment)는 간접평가와 반대되는 개념에서 쓰인 용어이다. 인간의 능력은 잠재되어 있기 때문에 잠재되어 있는 속성을 직접 측정하기가 불가능하므로 검사라는 도구를 사용하여 간접측정을 실시하게 된다. 그러나 수행평가는 검사에 의한 간접 측정이라기 보다는 어떤 행위를 보면서 평가하기 때문에 직접평가라는 용어를 사용한다. 대안적 평가(alternative assessment)는 평가방법의 특성을 염두에 두고 사용된 용어가 아니며 기존의 어떤 평가방법을 대치할 수 있는 평가방법을 말한다.

<표 3>에 설명된 용어를 볼 때 의미상으론 대안적 평가는 수행평가와 동의어가 아니다. 수행평가, 참평가, 포트폴리오, 그리고 직접평가가 동의어이나 이 모두를 포함하여 광의의 수행평가라 한다. 그러므로 주목하여야 할 것은 수행평가의 정의가 <표 3>에서 설명한 본래의 수행평가의 정의가 있고 확산된 개념으로 광의의 수행평가에 대한 정의가 있음을 알 수 있다. 수행평가를 이해함에 있어 본래 수행평가에 대한 정의를 기초로 수행평가에 대한 범위를 넓히는 것이 수행평가를 이해하기가 용이하다.

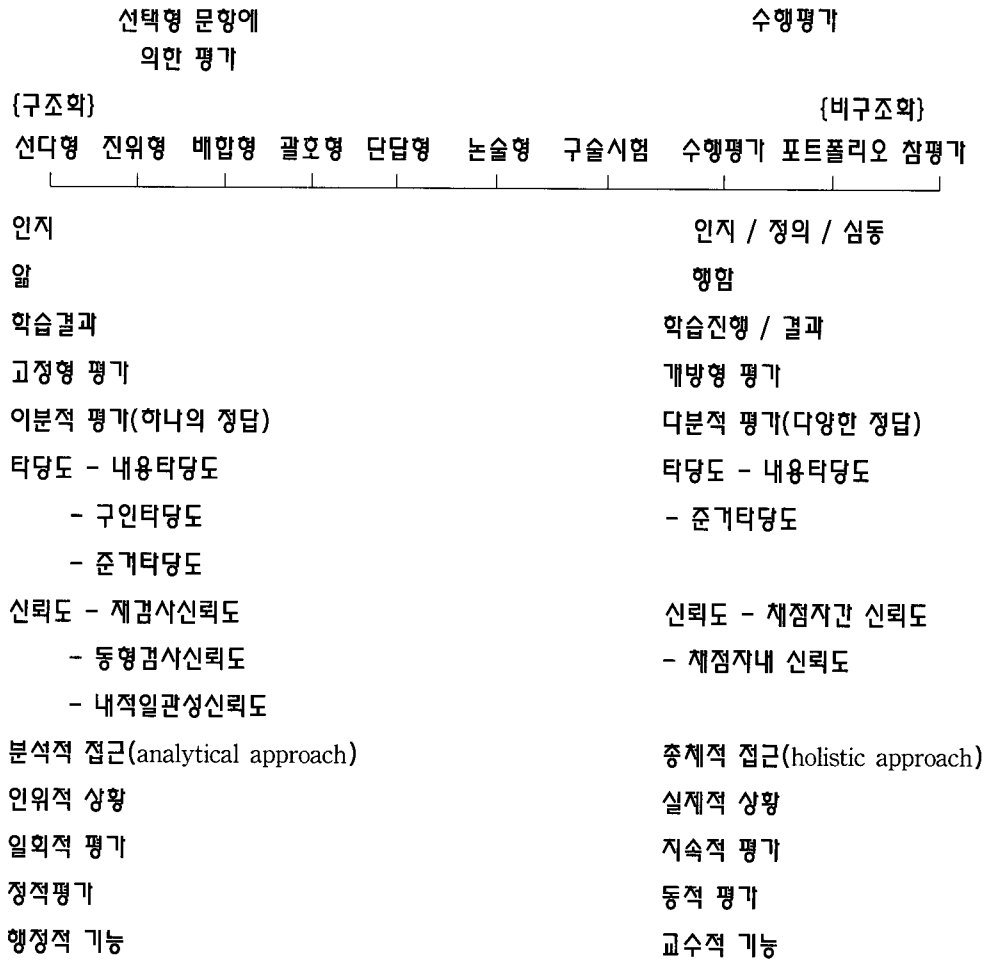
본래 의미의 수행평가는 심동적 행동특성을 평가하기 위하여 음악이나 체육 등과 같은 분야에서 주로 사용하던 평가방법으로 학습한 지식이나 습득한 기능, 기술을 얼마나 잘 수행(doing, performing)하느냐를 판단하는 평가방법으로 일반적으로 관찰에 의존하여 행위를 수행하는 모든 과정과 수행이 끝났을 때의 결과를 종합적으로 판단하는 평가방법이다. 그러므로 본래 의미의 수행평가는 행위의 정도를 보여주는 분야에서 개발된 평가방법이라 할 수 있다.

수행평가의 장점 때문에 심동적 행동특성을 평가하던 수행평가는 인지적 행동특성뿐 아니라 정의적 행동특성을 평가하는 방법으로 이용되었으며 그러므로 수행평가의 정의는 확대되었다. 확대된 수행평가는 참평가, 포트폴리오, 직접평가 등을 모두 포함하고 있으며 다양한 평가방법을 사용한다. 확산된 수행평가의 정의는 습득한 지식, 기능이나 기술을 실제 생활이나 인위적 평가상황에서 얼마나 잘 수행하는지(doing, performing) 혹은 어떻게 수행할 것인지(how to do, how to perform)를 관찰, 면접 등의 방법을 통하여 종합적으로 판단하는 평가방법으로써 지식이나 기능에 의한 정답 여부나 산출물에만 관심이 있는 것이 아니라 수행과정이나 그 결과를 평가한다(최연희·권오남·성태제, 1998).

수행평가가 많은 교과영역에 적용되면서 선택형 문항에 의한 평가를 제외한 모든 평가방법을 수행평가라고 하는 경향이 있다(교육부, 1998). 백순근(1998)은 수행평가는 외국에서 수입한 것이 아니라 우리 고유의 평가 방법이라 하며 수행평가에 서술형 및 논술형 검사, 구술시험, 토론, 실기시험, 면접, 관찰, 자기평가 및 동료평가 보고서, 연구보고서, 포트폴리오, 나아가 역할놀이, 프로젝트, 신문활용교육, 작품감상, 만들기 등도 수행평가라 하며, 평가목적이 학습과정을 개선하고 개별 학생에게 지도, 조언, 충고하기 위한 목적으로 사용되면 어떤 평가 방법도 수행평가에 포함될 수 있다고 주장한다. 이와 같은 주장은 학교 교사들에게 수행평가에 대한 오해와 개념의 혼란을 야기할 가능성이 있다. 일부에서는 선다형 문항이외의 모든 평가 방법을 수행평가라 하여 종전에 출제되던 논술형 문항, 단답형마저도 수행평가라 하고, 특히 단순한 계산형의 문제도 수행평가라 하는 오류를 범하고 있다.

이와 같은 혼돈은 평가방법을 임의적으로 그리고 인위적으로 구분하려 하기 때문이다. 면접은 면접이고 시험은 시험이다. 실험·실습이나 실기시험은 수행 능력을 평가할 수 있으므로 넓은 의미의 수행평가에 포함시킬 수 있으나 경기까지 수행평가라 할 수 있는지 의문이다. 이런 이유로 연구자는 인지적 행동특성을 평가하는 선택형 문항의 평가와 수행평가를 구조화의 정도에 따라 연속선상의 개념으로 보고 <그림 1>과 같은 특징을 설명한다.

선다형 문항에 의한 평가로부터 참평가에 이르는 연속선상에 있는 두 평가방법은 강조점이 다르다. 선다형의 선택형 문항에 의한 평가는 무엇을 알고 있는지 여부를 판단하는 데



〈그림 1〉 선택형 문항의 평가와 수행평가의 연속적 개념

초점이 있으나 참평가는 인위적 상황보다는 생활에서 알고 있는 것을 할 수 있느냐에 관심을 둔다. 그러므로 선택형 문항에 의한 평가는 고정된 형태의 평가방법에 의존하나 수행평가는 개방된 형태의 평가방법을 사용하며 결과뿐 아니라 진행과정까지 평가한다. 선택형 문항에 의한 평가는 일반적으로 정답과 정도의 차이를 나타내므로 평가도구의 타당도와 신뢰도를 객관적으로 추정할 수 있는 장점이 있다. 그러나 수행평가는 평가항목이 구조화되어 있지 않으므로 내용타당도와 준거타당도에 의존하여 평가도구의 타당도를 검증하고 관찰에 의하여 평가가 이루어지므로 관찰자인 채점자내, 그리고 채점자간 신뢰도에 의하여 평가 결과의 신뢰도를 검증한다.

등위판정, 서열화 등의 행정적 기능이 강조될 때 측정의 오차와 관련된 신뢰도를 높여야

하므로 선택형 문항에 의한 평가방법이 사용된다. 반면, 행정적 기능보다는 교수적 기능이 강조되고 채점자내와 채점자간 신뢰도에 보다 자유스러우며 수행평가를 실시할 수 있다. 만약 상호경쟁이 치열하여 평가 결과의 공정성과 신뢰성이 절대적이라면 수행평가 결과를 사용하는 데는 한계점이 있다.

연속 선상에 있는 선다형 문항에 의한 평가와 참평가 사이에서 어떤 평가방법을 수행평가로 볼 수 있는 기준은 논술에 의한 평가 형태 이상의 수준이라 말할 수 있으나 논술에 의한 평가를 모두 수행평가라 하지 않는다. 평가방법의 유형으로 수행평가 여부를 판단하는 확실적 기준은 없으나 무엇을 평가하느냐 즉 알고 있는 지식의 정도를 행위로 나타내거나, 행위로 나타내지 못하더라도 어떻게 할 것이라는 내용(how to do)을 서술할 수 있는지를 수행평가의 판단 기준으로 삼는 것이 바람직하다.

수행평가가 전통적인 평가의 대안으로 등장하게 된 궁극적인 쟁점은 평가도구의 형식보다는 평가의 내용이 무엇인가, 곧 무엇을 평가하는가에 있기 때문이다. 곧 수행평가는 전통적인 평가가 잘 다루지 못했던 복합적인 사고과정, 기술이나 기능 등의 고등정신 능력을 평가하고자 하기 때문에 수행평가의 개념 정의는 평가형식보다는 평가의 내용이나 목적을 근거로 이루어져야 할 것이다. 이런 의미에서 계산형 문항의 평가형태는 학생들에게 무엇을 요구하는가에 따라 전통적인 평가가 될 수도 있고 대안적인 평가가 될 수도 있다. 수학과에서의 수행평가는 학습자의 사고과정을 측정하고 아는 것을 행하는 정도까지 측정하여 평가하여야 하고 개방된 평가 유형에 의한 평가방법으로 고차원적이고 복합적인 능력을 측정하여야 한다.

수행평가의 정의에 의하여 NCTM(1995)은 평가기준(assessment standards)에서 수행을 '그 과제를 하는 동안 그 사람의 지식이나 판단을 드러나게 하는 물리적인 활동을 완성하거나 의미 있는 산물을 이루는 것'으로 규정하고 있고, Sternmark(1991)도 수학과에서의 수행평가란 학생들에게 수학적 과제, 프로젝트나 탐구 등의 수행 과제를 하도록 하여 그 산출물을 근거로 학생들이 실제로 아는 것과 학생들이 할 수 있는 것이 무엇인지를 평가하는 것이라고 정의했다. 수행평가는 예·체능이나 문학의 경우 학생들의 수행이나 활동 과정을 중요시하고, 그 과정 속에서 학생의 수행을 평가하겠다는 의미가 크다. 그러나 음악이나 문학에서처럼 실기나 활동의 비중이 크지 않은 중등 수학과에서는, 주로 문제해결의 사고 과정과 수학적 의사소통, 수학적 추론, 수학적 개념 등에 초점을 맞추고 있다.

미국의 Wisconsin 주의 수행평가 개발 프로젝트(Wisconsin Performance Assessment Development Project, 1995)에서는 수학과 수행평가를 학생들의 지식이나 기술을 증명하는 산출물이나 해답을 학생들이 창조해 내도록 요구하는 평가 방법으로 보았다. 이에, 실



험·실습을 유도하며, 에세이를 쓰도록 하고, 수학적으로 계산하도록 하는 것을 포함하는 여러 가지 형식이 될 수 있다고 말한다.

수학과에서의 수행평가는 학생들이 수업을 통해 배운 자신의 수학적 지식이나 기능을 보여줄 수 있는 산출물이나 해답을 스스로 창조해 내도록 요구하는 평가 방식이다. 장경운·권오남·최명례(1996)는 수행평가란 실제 생활과 관련된 과제를 해결하기 위해서 학생들이 자신이 알고 있는 수학적 지식이나 방법을 사용하고 경우에 따라서는 물리적인 수행을 하기도 하며, 교사는 그 과제의 산출물 등을 통해 학생의 수학적 능력을 평가하는 방법이라 정의하고 있다. 이상에서 살펴본 정의를 참고로 하여 본 연구에서는 수학과 수행평가를 다음과 같이 정의한다.

수학과 수행평가란 수학적 사고능력이나 문제해결 능력 등을 평가하기 위하여 학습자가 지니고 있는 수학적 지식이나 수학과 관련된 기능을 얼마나 알고 있으며 얼마나 잘 수행(performing)하는가와 어떻게 잘 수행 할(how to perform)것인가를 총체적으로 평가하는 방법으로써, 수학적 문제를 해결하는 과정과 결과, 수학적 원리를 찾아내는 과정과 결과, 그리고 실제 생활에 수학적 지식이나 기능을 적용하는 능력을 측정한다. 수학과 수행평가의 정의에서 볼 수 있듯이 수학과 수행평가의 특징은 수학적 지식이나 기능에 의한 산출물이나 정답여부에만 관심이 있는 것이 아니라 이를 도출하는 과정에 대한 관심이다. 그러므로 수행평가를 실시하기 위한 수학과 수행평가 과제는 실제 생활과 관련이 있거나 실생활에서 일어날 수 있는 문제 혹은 다양한 해결과정과 반응을 유도하는 열린 관제(open-ended task)의 성격을 지닌다. 이와 같은 수행평가의 예로 프로젝트, 포트폴리오, 관찰 및 면담 등을 들 수 있다.

## 2. 수학과 수행평가의 특징

Linn과 Baker(1996), Baron과 Boschee(1995) 그리고 Herman, Aeschbacher와 Winter(1992)는 수행평가의 공통적인 특징을 다음과 같이 설명하고 있다. 첫째, 개방형의 과제에 대하여 학생들이 반응하는 특성을 지니고 있으며, 둘째, 학생들에게 문제를 제기하고 해결하며, 분석하고, 연구하는 등의 다양한 활동을 허용하며 학생들의 이러한 활동은 판단력, 문제 해결력, 고등사고 능력, 의사소통과 같은 복합적인 기술을 요구하고, 셋째, 과제 수행을 위하여 충분한 시간을 필요로 하며 넷째, 학생 개인뿐 아니라 학생집단에 의하여 과제를 수행하도록 설계되기도 한다. 다섯째, 학생들은 교사와 함께 과제를 만들 수도 있고, 과제수행에 대한 평가의 준거도 논의할 수 있으며 여섯째, 점수 부여는 학생들의 수행

결과뿐 아니라 과정에도 근거한다는 점이다. 그러므로 수행평가에서는 점수부여 기준이나 준거를 개발하고 채점자를 훈련하는 일이 중요하다.

수학과 수행평가의 특징을 결정짓는 가장 중요한 요소는 과제라고 할 수 있다. 수학과 수행 과제(performance task)는 수행평가를 실시하기 위해 사용되는 과제로써 학생들에게 해결하도록 제시하는 문제이면서 동시에 해결 과정을 묻는 질문지이기도 하다. 답이 여러 개 있는 문제, 해결 접근법이 여러 가지인 문제, 실생활과 관련되고 적용이 가능한 문제, 개량된 주관식형의 문제 등이 여기에 해당된다.

수행과제를 논할 때는 과제의 진실성을 빼놓을 수 없다. 진실성이란 참평가와 관련된 용어으로써 성태제(1998b)는 참평가를 수행평가의 한 종류로써 실제 생활 속에서 평가하는 자연적 평가라 설명한다. Lesh & Lamon(1992)은 수행 과제는 참된 수학적 활동(authentic mathematical activities)의 성격을 띠어야 한다고 보았다. 참된 수학적 활동이란, 실제수학(real mathematics), 현실적 상황, 실생활에서 일어났직한 이슈나 질문들, 실생활적인 도구(realistic tools)나 자료들과 관련되는 것이다. 여기에서 '실제수학'이란 전통적으로 수학을 공식의 암기나 조건과 법칙의 목록일 뿐이라고 보았던 것에 비하여, 수학이 변화와 패턴을 탐구하고 예측하는 학문이라는 관점을 반영한 것이다.

Sternmark(1991)에 의하면 수학과 수행과제는 다음과 같은 기준을 만족시키는 것이라고 한다. 과제는 교육과정의 핵심적 내용을 다루어야 하고, 다른 질문이나 문제를 이끌어 낼 수 있어야 하고, 고차적 사고를 유발할 수 있게 지속성을 가져야 한다. 또한 과제는 학생이 자신의 사고 및 의사의 결정자가 되도록 하고 다른 학생과 상호 작용할 수 있게 하는 것이어야 하며, 그 수업 내에 끝낼 수 있는 것이며 안전하고 적절한 것이어야 한다. 과제는 학생들이 긍정적 태도를 가지도록 하는 것이어야 되고, 하나 이상의 해답을 가지고 모든 학생들에게 받아들여질 수 있는 다양한 접근 방법이 있는 것이어야 한다.

Lesh & Lamon(1992)은 참된 수학적 활동의 평가는 단편적 지식만을 측정하는 좁은 평가여서는 안 되고, 계산능력, 암기력만을 측정하는 등 평가 영역 상의 균형을 잃고 편중되어서도 안되며, 과거의 고전적인 이론과 방법을 강조하는 낡은 것이어서도 안 된다고 하였다. 예컨대, 교수·학문적으로 수학의 본질이나 유용함과 관련되는 실제적인 문제해결의 본질을 다루는 것이어야 한다. 즉, 수학과 수행평가를 통해 학생들은 공식이나 절차를 외워서 정답을 재생·기억해 내는 것보다는 당면한 문제의 해결 과정을 표현할 것을 요구받음으로써 자신의 수학적 능력, 문제 해결력 등을 평가받게 된다. 실제 생활과 관련된 문제를 해결하면서 수학적 가치를 인식하게 될 것이며, 완성된 수학적 지식의 습득이 아니라 스스로 만들어 가는 수학을 경험할 수 있게 될 것이다. 학생들은 자신의 수행을 돌아봄으로써

부족한 점을 이해하게 되고 이는 이후 학습에 많은 도움을 줄 수 있다. 또한 교사는 학생들의 상태를 좀 더 정확하게 파악하게 되어 학생들에게 적합한 교수를 제공할 수 있고, 자신의 수업을 반성할 수 있다. 이러한 관점에서 수행평가는 수업의 연속 선상에 있는 것이다.

이와 같은 사항들을 종합하여, 본 연구에서는 수학과 수행평가의 특징을 다음과 같이 요약한다. 첫째, 수학적 지식과 방법을 사용하여 나타난 활동과 산출물로 평가한다. 둘째, 문제해결의 과정이 드러나도록 하여 평가한다. 셋째, 창의력이나 논리적 사고력 등과 같은 고차적 정신능력을 평가한다. 넷째, 과정 중심적이고 평가 목표 지향적인 채점 방법을 사용한다. 다섯째, 실제 생활과 밀접한 관련이 있는 문제 상황(수학의 내적·외적 연결성을 강조함)에서 평가한다. 여섯째, 협동학습과 개별학습이 평가와 병행하여 이루어진다. 일곱째, 평가가 수업의 과정 속에 이루어진다. 여덟째, 한두 시간에서 수일에 걸치는 프로젝트나 탐구과제를 수행하는 것이다. 아홉째, 자와 컴퍼스 등과 같은 도구, 컴퓨터와 계산기 등과 같은 테크놀로지를 비롯한 실 생활적인 도구(realistic tools), 구체적 조작물(manipulative)을 사용할 수 있는 평가라고 할 수 있다.

### 3. 수학교과에서의 수행평가의 예

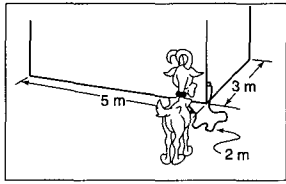
수행과제와 채점표의 예는 최연희·권오남·성태제(1998)의 연구보고서에서 인용한 것이다.

- 평가방법 : 지필법 및 관찰법
- 평가대상 : 중 1
- 수행과제

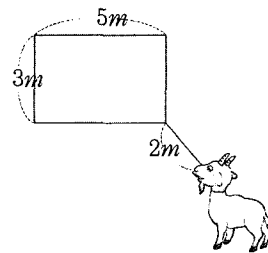
다음 수행과제는 염소가 활동할 수 있는 범위를 학생들이 예측해서 그 부분의 넓이를 구해야 하는 과제로 <그림 2>와 같으며 전체 내용은 최연희·권오남·성태제(1998)의 보고서에 수록되어 있다. 염소가 직사각형, 정사각형, 정삼각형 모양의 기둥에 묶여 있을 때, 풀을 뜯을 수 있는 영역의 면적을 계산하는 것이 이 과제의 주된 내용이다. 여기서는 45분 이내에 과제를 해결하기 위하여 직사각형과 정삼각형, 정사각형의 경우만이 제시되었지만, 정오각형, 정육각형, 더 나아가 정  $n$ 각형까지도 확장해서 제시할 수 있다.

〈그림 2〉 수행과제의 예

[질문 하나] 다음 그림처럼 직사각형 모양의 기둥에 염소를 매어 둔다면 염소가 풀을 뜯을 수 있는 넓이는 얼마나 되나요?

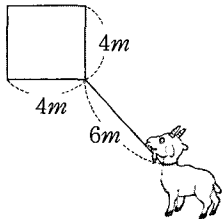


◆ 염소가 다닐 수 있는 곳을 그림으로 표시하세요.



◆ 염소가 다닐 수 있는 곳의 넓이를 계산하세요.  
(풀이 과정도 쓰세요.)

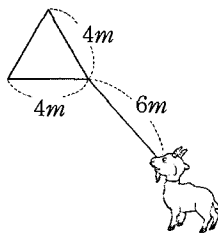
[질문 둘] 염소를 다음 그림과 같이 정사각형 모양의 기둥에 매어 둔다면 염소가 풀을 뜯을 수 있는 넓이는 얼마나 되나요?



◆ 염소가 다닐 수 있는 곳을 그림으로 표시하세요.

◆ 염소가 다닐 수 있는 곳의 넓이를 계산하세요. (풀이 과정도 쓰세요.)

[질문 셋] 염소를 다음 그림과 같이 정삼각형 모양의 기둥에 매어 둔다면 염소가 풀을 뜯을 수 있는 넓이는 얼마나 되나요?



◆ 염소가 다닐 수 있는 곳을 그림으로 표시하세요.

◆ 염소가 다닐 수 있는 곳의 넓이를 계산하세요. (풀이 과정도 쓰세요.)

• 채점기준

수학적 의사소통 능력과 문제해결 능력에 대한 채점 기준을 설정하였고, 그 내용은 다음과 같다.  
가. 수학적 의사소통에 대한 채점 기준

수 학 적 의 사 소 통	도형과 각의 크기, 면적의 표현이 분명하고 명확함	1 점	· 도형과 각의 크기, 면적 등의 표현이 미숙하고 틀린 경우 · 문제를 해결하는데 있어 도형과 각의 크기, 면적 등을 표현하지 못하고 언어적인 설명으로만 하는 경우
		2 점	도형과 각의 크기, 면적의 표현을 하고 있기는 하지만, · 각의 크기, 변의 길이가 실제적으로 맞지 않은 경우 · 도형과 각의 크기, 면적의 표현이 정확하지 않고 모호한 경우
		3 점	도형과 각의 크기, 면적의 표현 등이 세련되고 완전하며 상세하여 보기가 쉽고 명쾌하다.
	식의 표현	1 점	식을 나타내는데 있어, 기호나 단위의 사용이 미숙하거나 없다.
		2 점	식의 표현이 정확함

나. 문제 해결 과정에 대한 채점 기준

문 제 해 결 의 과 정	과제에 대한 이해	0 점	무응답
		1 점	문제를 완전하게 이해하지 못한 경우 : 문제에서 제시되지 않은, 관련되지 않은 식이나 생각들을 비논리적으로 나열해놓은 경우
		2 점	문제를 부분적으로만 이해한 경우 : 얻어진 자료의 수치들을 그대로 베껴놓은 경우
		3 점	이해된 경우
	해결의 접근 방안, 전략, 절차 등의 선택	1 점	전반적으로 상황에 알맞지 않은 전략을 사용하거나 논리적으로 부적절한 절차를 사용한 경우
		2 점	부분적으로 적절한 접근이나 절차를 사용한 경우 : 염소가 움직인 범위를 줄의 길이와 맞지 않게 생각하여 구한 경우
		3 점	해결 전략을 알맞게 정하고, 절차가 모두 맞는 경우
	전력의 실행	1 점	자신이 세운 전략을 실행하지 못한 경우
		2 점	넓이를 표현하고 계산하는 과정에서 계산적인 착오가 있어 틀린 경우
	전력의 실행	3 점	정답을 구했으나 단위가 틀렸거나 문제 상황에 맞게 해석하지 못한 경우
		4 점	조건에 맞게 해결하고, 문제 상황에 맞게 해석한 경우

#### 4. 수행평가의 문제점

수행평가의 장점은 <그림 1>에서 설명한 것과 같이 인지적 능력, 정의적 특성, 심동적 특성을 모두 평가할 수 있는 총체적 접근이고 문제 풀이 과정도 분석할 수 있으며 개방형 형태의 평가 방법은 다양한 사고 능력을 함양시킬 수 있다. 또한 수행평가는 과제의 성격상 협동학습을 유도하므로 타인과 협동하여 나아가는 전인교육도 도모하는 장점이 있다.

수행평가 실시에 따른 장점 못지 않게 단점도 적지 않다. Wolfe & Miller(1997)는 미국의 16개 중등학교 교사들이 포트폴리오 방법을 사용할 때 나타나는 가장 큰 문제점을 시간과 점수화 과정이라고 밝히고 있다. 포트폴리오 평가방법 설계와 적용, 채점에 소요되는 많은 시간, 점수화 하는 기준, 점수 부여 방법, 학생과 학부모에게 통보, 그리고 점수 사용에 어려운 점이 있다고 보고하고 있다. 특히 교사가 수행평가 결과에 대하여 학생이나 학부모와 의사를 교환하는 데 애로점이 많다고 토로하고 있다.

수행평가를 실시하는 데 소요되는 비용도 만만치 않다. Reckase(1995)는 포트폴리오에 의한 평가 결과의 신뢰도는 낮으므로 훈련된 채점자에게 평가를 의뢰하여야 하며 Miami 대학에서 포트폴리오 하나를 분석하는 데 \$17.5의 비용이 소요된다고 보고하고 있다. Stecher & Klein(1997)은 과학교과에서 수행평가가 생각했던 것보다 훨씬 많은 비용이 소요된다고 결론지으며, 학기당 한 학생을 위하여 드는 수행평가 비용은 \$34이고 신뢰도를 .8로 유지하려면 \$102가 소요되며, 이는 표준화검사를 실시하는 비용의 약 20배 정도가 된다고 발표하였다. 또한 Wisconsin 주의 경우 수행평가가 3년간 연구과제로 선정되어 진행되다가 1년 반만에 주정부의 결정으로 연구가 중단되는 사례도 있었다. 미국은 우리나라에 비하여 학급 크기가 작고 교사들의 업무가 적으며 교육에 투자하는 비용이 많은 데도 불구하고 교사들은 수행평가를 실시할 때 소요되는 시간과 비용의 문제를 제기하고 있다. 이상의 연구들을 정리하면 수행평가의 문제점은 다음과 같다.

첫째, 수행평가 도구 개발에 어려움이 있다. 수행평가 도구를 개발하기 위하여는 교과 내용은 물론 학습자들의 인지구조, 그리고 학습과제들을 실생활에 적용하는 범위까지 고려하여야 하므로 전통적 방법에 의한 평가 문항 개발보다 수행평가 도구 개발이 어렵다.

둘째, 채점기준 즉 점수부여 기준 설정이 용이하지 않다. 수행평가는 수행과정까지 고려하여 점수를 부여하게 된다. 이런 경우 어느 정도까지 몇 점을 주어야 하는지에 대한 판단이 필요하며 이 점수부여 기준 설정에 어려움이 있다.

셋째, 채점자내 신뢰도와 채점자간 신뢰도 확보에 어려움이 있다. 전통적 평가방법보다 주관이 개입될 소지가 많은 수행평가에서 일관성 있게 점수를 부여하는 것은 쉬운 일이 아

니다. 논술형 문항을 채점하더라도 채점자 한사람이 시간의 변화에 상관없이 일관되게 채점하는 것이 쉽지 않으며 나아가 여러 채점자가 서로 유사하게 채점을 하는 것 역시 쉬운 일이 아니다.

넷째, 시간이 많이 소요된다. 평가도구 개발, 점수부여 등에 많은 시간이 소요된다. 평가도구 개발, 채점표 작성, 채점 등에 소요되는 시간이 늘어날 수밖에 없다. 컴퓨터에 의하여 기계적으로 채점하는 선다형 문항에 비하여 수행평가는 검사는 물론 채점에 필요한 시간이 늘어난다.

다섯째, 많은 비용이 든다. 수행평가의 전략과 채점의 방식에 따라 소요되는 시간과 필요한 교사의 수가 달라질 수 있으므로 실제적인 비용을 정확히 산출하기 어렵다 하더라도 전통적 평가 방법에 비하여 많은 비용이 소요된다.

여섯째, 점수결과 활용에 어려움이 있다. 학생과 학부모가 평가 결과를 인정하지 않을 경우 점수결과를 활용하는데 많은 문제가 야기된다. 나아가 교사를 불신하는 일까지 일어나게 된다.

최연희·권오남·성태제(1998)도 수행평가를 실시함에 있어 교사들에게 어려운 점은 수행평가로써의 의미를 지닐 수 있는 문제상황을 만들기가 어려웠으며, 수행결과에 대한 평가의 객관성 확보, 그리고 평가결과에 대한 학생들의 동의 거부 등의 문제를 지적하였다. 그러나 수행평가를 차시적으로 실시할 때 학생들의 질문이 자연적으로 유도되었고, 학생들의 이해 정도를 보다 정확히 이해할 수 있었으며, 즉각적 피드백은 학생과의 공감대를 형성하여 수업분위기가 좋아졌으며, 학생들의 수학적 주장이나 수학적 의사소통 능력이 발전하였다고 보고하고 있다.

#### IV. 수학과 수행평가의 연구 과제와 전망

수학 교육을 위한 수행평가를 성공적으로 실시하기 위하여 수학 교육 분야와 교육평가 분야에서 해결하여야 할 과제가 있으며 동시에 학교 교육환경의 변화도 요구된다. 성태제(1995)는 수행평가를 실시하기 위하여 첫째, 교육평가에 대한 학생 그리고 학부모의 인식 변화, 둘째, 인지심리학에 대한 연구 활성화, 셋째, 교과교육 연구 활성화, 넷째, 교사들에게 수행평가 연수, 다섯째, 수행평가를 위한 타당도와 신뢰도 증진방법, 여섯째, 교과별 혹은 단원별 수행평가 연구 및 도구 개발, 일곱째, 교육과정과 교과내용 개편, 여덟째, 학급 단위 축소, 아홉째, 수행평가 실시를 위한 경비 지원을 들고 있다.

이와 더불어 수학 교과 분야에서 수행평가를 실시하기 위해서는 학습자의 수학의 습득과 이해 등에 대한 수학적 사고 구조에 대한 인지적 관점에서 접근한 교재가 제작되어야 한다. 또한 수학 인지 사고 방법에 의한 교수가 실시되어야 할 것이다. 그러므로 수학 교육을 위한 인지학습 이론에 대한 연구 뿐 아니라, 수학 수행평가를 위한 올바른 이해와 수행평가 도구 개발이 원활히 이루어 져야 한다. 아울러 수학 수행평가를 위한 채점기준을 명확하게 수립할 수 있는 방법이 구안되어야 한다. 각 교과 영역에서의 수행평가의 성공 여부는 수행평가의 내용에 달려 있다고 보아도 과언이 아니다.

수행평가를 위하여 교육평가 분야에서 연구할 과제로는 점수부여를 위한 준거개발, 다른 과제에 의한 점수조정, 채점자간 신뢰도와 채점자내 신뢰도 검증을 위한 일반화 가능성 이론(generalizability theory), 수행평가 도구의 질을 분석하는 다분문항 반응 이론(polytomous item response theory)을 들고 있다(성태제, 1998b).

수행평가를 실시하기 위하여 변화해야 할 우리의 교육환경은 교과 교사당 학생 수를 대폭 줄여야 하며 교사의 잡무를 대폭 감소하여야 한다(성태제, 1998c). 미국의 경우 수행평가를 실시하는데 따른 교사들의 불만에서도 알 수 있듯이 수행평가를 실시하기 위하여 많은 시간이 소요되므로 교사들이 수행평가를 위하여 많은 시간을 할애할 수 있는 배려가 있어야 한다. 수행평가의 가장 큰 문제점으로 채점자간 신뢰도와 채점자내 신뢰도를 제기하였다. 미국의 경우도 채점 결과에 대하여 학생이나 학부모가 동의하지 않을 경우 채점 결과를 활용하기는커녕 의사도 교환되지 않는다 하였다. 우리 나라와 같은 경우 수행평가에 대한 점수 결과가 많은 영향을 준다면 수행평가 결과에 대하여 의의를 제기할 뿐더러 또 다시 교사를 불신하는 교육문제를 유발할 수 있다. 그러므로 교사들의 수행평가 결과를 신뢰할 수 있는 교육적 환경을 사전에 구축하는 것이 선결 과제이다. 교사들의 행동이 의심받는 교육환경에서 수행평가 결과는 교사들에 대한 오해를 불러일으킬 가능성이 있으므로 수행평가를 실시하기 전에 많은 준비가 요구된다.

우리 나라의 경우 수행평가가 고등정신 능력을 함양하고 학교 교육의 모든 문제를 해결할 수 있는 유일한 방법이며 최선의 방법으로 생각하여 수행평가를 현장에 적용하도록 독려하고 있으나 수행평가의 결과가 학생에게 미치는 영향을 간과해서는 안 된다. 그러므로 수행평가가 무엇인지, 교과별 수행평가는 무엇인지, 교과별로 어떻게 수행평가 도구를 만드는지, 수행평가를 어떻게 실시하고 채점하는지, 수행평가 결과는 어떻게 보고하고 사용하는지에 대한 체계적인 연수가 필요하다.



## 참 고 문 헌

- 교육부(1998). 수행평가의 이해.
- 김정환, 강선보(1997). *교육학 개론*. 서울:박영사.
- 남명호(1995). 수행평가의 타당성 연구. 고려대학교대학원 박사학위논문(미간행).
- 백순근(편)(1998). *수행평가의 이론과 실제*. 서울:원미사.
- 성태제(1998a). 수행평가 그리고 과제와 전망. *사대뉴스* 12월호.
- 성태제(1998b). 교육평가 방법의 변화와 결과타당도. 한국교육평가학회: 21세기 한국교육 평가의 과제와 전망, 125~147.
- 성태제(1998c). 2002학년도 대학입학전형제도 개선에 따른 교육환경의 변화와 연구과제. 한국교원대학교 교육과학연구소: 제2회 현장교육개혁 연구발표·토론 대회 발표 논문.
- 장경윤, 권오남, 최명례 (1996). 중학교 수학 수행평가 문항의 개발 및 그 활용 가능성의 탐색. *교과 교육 공동 연구 결과 보고서*.
- 최연희, 권오남, 성태제 (1998). 중학교 영어·수학 교과에서의 열린 교육을 위한 수행평가 적용 및 효과 분석 연구. 교육부 초등교육정책과 열린 교육 연구 과제 연구 결과 보고서.
- Baron. M. A., & Boschee. F. (1995). *Authentic assessment*. Basel, Switzerland : Technomic Pub. Co Inc.
- Herman, J. L., Aeschbacher, P. R., & Winters, L.(1992). *A practical guide to alternative assessment*. Alexandria : Association for Supervision and Curriculum Development.
- Lesh, R., & Lamon, S. (Eds.) (1992). *Assessment of authentic performance in school mathematics*, Washington, DC: AAAS Pub. Co.
- Linn. R. L., & Baker. E. L. (1996). Can performance-based student assessments be psychometrically sound? In J. B. Babon & D. Wolf(Eds.), *Performance-based student assessment: challenges and possibilities*. Chicago: University of Chicago Press.
- Mayer(1992). What's the difference between authentic and performance assessment? *Educational Leadership*.
- Mislevy, R. J.(1997). Assessing student learning. In H. J. Walberg & G. D. Haertel(Eds.), *Psychology and educational practice*(pp.176-195). Berkeley:

- MrCutrhan Publishing Corporation.
- National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.
- Reckase, M., D.(1995). Portfolio assessment : A theoretical estimate of score reliability. *Educational Measurement : Issues and Practice*, 14(1), 12-14.
- Simpson, D. P.(1987). *Cassell's Latin-English English-Latin Dictionary*. London: Cassell Publisher.
- Stecher, B. M., & Klein, S. P.(1997). The cost of science performance assessments in large-scale testings programs. *Educational Evaluation and Policy Analysis*, 19(1), 1-14.
- Sternmark, J. K. (1991). *Mathematics assessment*. Reston, VA: National Council of Teachers of Mathematics.
- Tyler, R. W.(1942). General statement on evaluation, *Journal of Educational Research*, 35, 492-501.
- Wisconsin Student Assessment System (1995). *Performance assessment sampler*. Wisconsin Department of Public Instruction; Madison, Wisconsin.
- Wolfe, E. W., & Miller, T. R. (1997). Barriers to the implementation of portfolio assessment in secondary education. *Applied Measurement in Education*, 10(3), 235-251.

## Future Directions and Perspectives for Performance Assessment in Mathematics

Oh-Nam Kwon(Ewha Womans University)

Tae-Jae Seong(Ewha Womans University)

The first part of this paper examines the assumptions underlying traditional types of standardized testing compared with assumptions underlying innovative types of performance assessment. The second part focuses on directions for the future. The third part gives examples from closed related projects which were designed to find practical ways to implement recommendations.