

**초등학생의 자연과 수행평가 실태조사 및
초등학교 자연과 수행평가도구의 개발 I**

**임영득 · 조혜경 · 한안진 · 박현주 · 송민영 · 김은진 ·
홍석인 · 강호감 · 노석구**

인천교육대학교

**Survey on Authentic Performance Assessment
for Elementary Science Education**

**Lim, Young-Deug · Cho, Hea-Kyung · Hann, Ann-Chin · Park,
Hyun-Ju · Song, Min-Young · Kim, Eun-Jin · Hong, Seok-In ·
Kang, Ho-Kam · Noh, Suk-Goo**

Inchon National University of Education

ABSTRACT

The purpose of this study was to investigate about current status of authentic science performance assessment in elementary schools. A total of 707 elementary teachers was involved and the data from questionnaire was gathered. The questionnaire was comprised two categories: items on the epistemological aspect & items on the actual condition.

The major findings are as follows:

1. There are some elementary teachers who don't have a clear understanding on what authentic science performance assessment is even though most of teachers admit the importance of authentic performance assessment in their science teaching.
2. Many teachers have chosen and have used experimental activities or performance tasks rather than they have developed tasks of their own.
3. Many teachers have pointed out 'the lack of time for performance assessment preparation' and 'the limitation of class time' as the problems of working on science performance assessment in elementary schools.

I. 서론

오늘날의 과학교육은 학습자의 특성 및 개성을 존중하는 총체적이며 자율적인 교육을 지향한다. 최근 과학교육 연구자들의 관심은 과학교육과정의 목표 및 교수법에 부합되고, 학습자의 문제 해결력과 탐구능력, 정의적 영역, 그리고 창의성을 평가할 수 있는 '수행평가'에 집중되고 있다(Herman, et al., 1992). 이것은 지난 60년간 교육계를 지배해왔던 평가가 1990년에 들어서면서 새로운 전환 국면을 맞았음을 의미한다(Stiggins, 1991).

1990년대 이전의 과학교육의 평가는 주로 선다형이나 수학적 문제풀이와 같은 표준화된 검사의 형태를 발전시켰다. 잘 계획되고 주의 깊게 개발된 선택형 문제는 과학 지식에 대한 학습자의 복합적 이해 정도를 측정할 수 있다. 그러나 그것은 근본적으로 학습자의 사고 과정을 '수렴적'으로 제한한다. 이것이 선택형 검사의 기본적인 한계로써, 문항들을 통해 측정할 수 있는 능력이나 지식의 범위가 한정되어 있음을 의미한다(Haertel, 1991).

또한, 최근 연구 결과에 의하면, 지필 평가를 통한 과학적 능력과 실험과정에서의 탐구 능력은 학습자에 따라 차이가 있는 것으로 나타났다(김동찬, 1991). 이와 같이 지필 평가와 실험 과정평가의 결과가 다른 것은 지필 평가 문항은 문제 상황이 단순하고 구체적으로 제시되는 반면, 실제 실험은 전체적인 맥락에서 과학 개념과 종합적 사고가 요구되기 때문이다(최병순 외, 1994). 따라서, 지필 평가는 학습자들의 완전학습 능력, 참여, 창의성, 자신감, 지속성, 새로운 사고의 생성 등과 같은 자연과 학습 요인들(박현주, 1998)에 대한 평가의 기회를 제공하지 못한다.

수행평가(performance assessment)는 학습자들이 '무엇(what)'을 알며, 주어진 현상을 이해하기 위해 '어떻게(how)' 과학적 개념들을 사용하는가를 총체적으로 이해하는데 그 목적을 두고 있다(Chittenden, 1991; Lazzaro & Park, 1994;

Meisels, 1993; Perrone, 1990; WCEA, 1993).

현재 한국교육과정평가원과 각 시·도 교육청 등과 같은 교육 기관에서 수행평가의 중요성을 인식하고 교육 현장에서의 실행을 적극적으로 장려하여(국립교육평가원, 1996), 초등학교 현장에서 수행평가의 실행이 점차 확산되고 있는 추세이다. 그러나 수행평가가 단순히 한 가지 형태의 평가이라기보다 학습과정/평가에 대한 패러다임의 변화를 요구하기에 구성주의, 수행평가, 수행평가의 방법 등에 대한 올바른 이해가 필수적이다. 따라서 본 연구는 현재 초등학교 현장에서 실시되는 자연과 수행평가에 대한 실태를 초등학교 교사를 대상으로 조사함으로써, 학교 현장에서 적용 가능한 타당도와 신뢰도가 높은 수행평가 도구를 개발하기 위한 기초 자료를 얻는데 그 목적이 있다. 이를 위한 세부적인 연구 문제는 다음과 같다.

첫째, 초등학교 교사의 자연과 수행평가에 대한 인식은 어떠한가?

(1) 수행평가란 무엇인가?

(2) 초등학교 자연과에 있어서 수행평가가 필요한가?

(3) 수행평가의 방법에는 무엇이 있는가?

둘째, 현재 자연과 수행평가의 실행 실태는 어떠한가?

(1) 현재 사용하고 있는 수행평가 방법은 무엇인가?

(2) 평가의 비율은 어떻게 하고 있는가?

(3) 현장에서 수행평가 실행상의 문제점은 무엇인가?

II. 이론적 배경

1. 수행평가의 정의

수행평가에 대한 정의는 학자에 따라 다양하게 내리고 있다. Gitomer(1989)는 학생들의 수행 과제 해결 과정의 질을 판단하는 활동을 수행평가로 정의한다. Berk(1986)는 "수행평가는 개인에 대한 의사결정을 위해 체계적인 자료를 수집

하는 과정"이라고 정의하면서 수행평가의 특징을 다섯 가지로 설명하고 있다. 첫째, 수행평가는 과정(process)이지 검사나 어떤 단일한 측정도구가 아니다. 둘째, 수행평가 과정은 다양한 도구와 전략들을 사용하는 자료 수집에 초점을 둔다. 셋째, 자료는 체계적 관찰이라는 수단을 이용하여 수집한다. 넷째, 수집된 자료는 의사결정이라는 목적을 위해 통합된다. 다섯째, 의사결정의 대상은 프로그램이나 결과물이 아니라 피고용자나 학생과 같은 개인이다(남명호, 1995).

국내 연구에서 백순근(1998)은 수행이 구체적인 상황하에서 실제로 행동을 하는 과정이나 그 결과를 의미한다고 언급하면서, 수행평가를 학생 스스로가 자신의 지식이나 기능을 나타낼 수 있도록 산출물을 만들거나, 행동으로 나타내거나 답을 작성(구성)하도록 요구하는 평가방식이라고 정의한다. 남명호(1995)는 수행평가는 학습과정, 결과물(일기, 글짓기, 작품집, 전시물 등), 수행(연구, 음악공연, 토론 등)을 모두 포함하는 광의의 개념이라고 정의한다.

한편, 수행평가는 직접 평가(direct-assessment), 참 평가(authentic assessment), 대안 평가(alternative assessment), 수행 평가(performance assessment) 등으로 지칭된다. 직접 평가는 간접적이고 대리적인 과제(검사 문항이 대부분)에 의존하는 전통적인 평가와 대비되는 면을 강조하고, 참 평가는 학생 행동의 작은 단편들을 표집하여 실시하는 전통적인 검사와 대비시켜 학생이 "참으로 가치 있는(authentic)" 과제를 어떻게 수행하는가를 평가하는 것을 의미한다. 그리고 이러한 평가들이 전통적인 검사에 대한 대안으로서 제안되었기에 대안 평가라는 이름으로 불리기도 한다. 이러한 용어들은 강조점이 약간씩 다르나, 모두 전통적인 선다형 지필 검사의 대안, 그리고 일상 생활과 관련된 과제에 대한 학생의 수행을 직접 평가한다는 공통적인 특징이 있다.

본 연구자들은, 위에서 언급한 여러 학자들의 다양한 정의들을 기초로 하여 수행평가를 다음

과 같이 정의한다: "수행평가는 학생들이 지식을 구성해 가는 과정에 대해, 혹은 실생활에서 부딪칠 수 있는(또는 실제적인) 문제를, 과학 개념, 과학적 사고 및 탐구과정, 창의력, 적용력, 그리고 과학적 태도 등을 통해서, 해결하는 과정과 결과에 대해 다양한 방법을 통하여 총체적으로 이해하고자 하는 노력이다."

수행평가의 요소는 다음의 세 가지로 정리된다(Brown & Shavelson, 1996). 첫째는 수행과제(performance tasks)로서, 이는 학생들에게 문제를 풀도록 하는 구체적인 상황을 제공하며 더불어 구체적인 자료도 함께 제공한다. 학생들은 수행과제를 해결하기 위해 특별한 수행을 하는데, 수행평가는 이 과제 수행의 과정과 결과를 평가한다. 둘째는 반응 양식(response format)인데, 이것은 학생이 발견한 방법이나 해답 과정 등을 다양한 방법으로 기록하여 타인에게 이를 명확히 알릴 수 있도록 하는 기록지를 의미한다. 여기에는 학생이 결과를 요약하고 도표화하는 능력이라든가 상호 작용하는 능력 등이 평가요소로서 나타난다. 셋째는 채점체계(scoring system)로서, 이것은 평가자가 학생의 수행과정과 결과를 관찰하거나 검토하면서 직접 점수화 하기 위한 기준이라 할 수 있다. 이 세 번째 요소가 없어도 수행과제를 통한 수업은 가능하다. 그러나 평가를 했다고는 말할 수 없다고 그들은 주장한다.

한편, 김찬중과 김혜정(1998)은 우리 나라 교육 현장에 적합한 수행평가(포트폴리오) 형태의 구성 요소로 목표, 증거, 평가 준거를 제시하였다. 그들에 의하면, 수행평가의 구성은 교사에 따라 융통성 있게 계획되어야 하고, 학생들이 이해하기 쉬운 구체적 학습 목표와 학생들이 작성한 증거를 평가할 수 있는 타당하고 효율적인 평가 준거가 개발되어야 한다고 주장하였다..

2. 수행평가의 특징

(1) 평가 기능의 균형

과학교육에서 평가의 기능을 교육적 기능과 사회적 기능의 측면에서 살펴보면, 교육적 기능

은 학생의 과학 학습 활동 자체를 이해하고 교수 학습의 효율을 기하기 위한 것인데 반해, 사회적 기능은 학생의 선발과 배치 또는 입시 등 주로 평가의 결과를 정보로써 활용하는 것이다. 이러한 평가의 교육적 측면은 최근까지 사회적 측면에 비하여 상대적으로 그 중요성을 적게 두는 경향이였다. 그러나 수행평가는 평가의 교육적·사회적 기능의 균형을 강조한다.

(2) 평가와 학습의 일원화

기존 과학교육 평가는 과학의 교수·학습과 분리하여 학생 성취에 대한 정보를 얻는 방법으로 생각하는 경향이 지배적이였다. 그러나 수행평가는 과학의 교수·학습의 일부로, 과학 교수·학습과 구별되는 것이 아니다. 이것은 다음 두 가지의 기본 가정을 뒷받침한다: 첫째, 교수/학습의 질은 교사가 다양한 평가 과정 및 방법을 통하여 명확하게 학생들을 이해할 때 높일 수 있다. 둘째, 평가의 목표와 방법이 교수의 목표 및 방법과 다를때, 필연적으로 수업에 부정적인 영향을 미친다. 따라서 수행평가는 효율적인 학습을 위한 평가와 학습의 일원화를 의도한다 (Champagne, Lovitts, & Calinger,1990; Wiggins, 1993; Wolf, Bixby).

(3) 평가자의 다양화

Olson(1995)은 구성주의 교수·학습 활동에서 주체자의 역할을 하고 있는 학생들의 과학 학습을 평가할 때, 학생을 평가의 주체가 아닌 객체로서만 다루게 되는 현재의 평가방법에 대하여 문제제기를 한다. 학생의 과학 학습에 대한 교사의 객관적 평가와는 별개로, 학생 자신이 과학 학습의 목표를 어떻게 이해하고 있으며, 과학적 지식에 부과하는 의미는 어떠한가를 스스로 평가해야 함을 주장하고 있다. 수행평가는 학습 또는 평가에 대한 학생의 일방적인 수용의 형태에서, 교사와 학생간, 학생과 부모간, 학생과 학생간, 그리고 학생과 대상(예를 들면, 수행 과제)간의 의사소통을 강조함으로써 학생들의 창의적이고 자기 주도적인 학습과 평가를 강조한다.

(4) 평가 대상과 평가 방법

수행평가는 다양한 주체, 그리고 여러 가지 방법을 활용하여 학생의 학습 과정 및 능력에 대한 정보를 얻고, 이를 총체적으로 이해하고자 하는 것이다. 수행평가 대상으로 다음을 생각할 수 있다. 첫째, 학습 과정이다. 조각 지식의 합이 아니라, 전체적인 맥락 속에서 지식으로 구성되는 학습자 활동을 그 대상으로 한다. 학습과정의 평가를 위한 방법의 예는 면담법 및 관찰법, 혹은 체크리스트를 통하여 찬반토론, 프로젝트, 게임과 역할놀이, 실험과정평가, 야외 활동 등이 있다. 둘째, 학생 학습의 다양한 결과이다. 지식의 생산(production of knowledge)은 자기화된 진정한 표현(authentic expressions)과 기술의 연마에 기초하여 스스로 구성하여야 한다. 이것은 지식의 단순한 암기나 이해를 넘어선, 지식의 구성, 이해, 생성, 전달 등으로 나타난다. 학습 결과물에 대한 평가방법은 연구보고서, 일지, 작품, 지필 검사, 개념도, 그림 그리기, 실기 평가 등이 있다. 또한, 학생의 학습 과정과 결과를 지속적으로 평가할 수 있는 포트폴리오가 있다. 포트폴리오는 학습지, 실험노트, topic이나 project의 보고서, 실험·실습의 결과 보고서 등을 정리한 학생의 성장과 학습 과정상의 자기 반성을 포함하는 '증거 모음집'이다.

(5) 평가 시기

평가는 실시시기에 따라 연속적인(continuous) 평가와 비연속적(discrete) 평가로 나뉜다. 연속적인 평가는 교수·학습과 동시에 그리고 흔히 교수·학습에 통합되어 이루어진다. 평가가 연속이라는 것이 평가가 끊임없이 이루어지는 것을 의미하는 것이 아니라, 단지 제한된 시간에 집중적으로 이루어지는 구체적인 평가가 없음을 의미한다. 연속적인 평가의 목적은 학생들에게 자신들의 수행 정도를 알려줌으로써 학습 활동을 통제하고 수정하도록 도와주며, 교사들에게 자신들의 수업을 수정하도록 하는 데 있다. 비연속적 평가는 학기의 중간 또는 말에 이루어진다. 여기에는 전통적인 중간고사, 기말고사, 졸업고사 등이 포함된다. 비연속적 평가는, 일반적으로 학생,

학부모, 교사에게 개인의 학습에 대한 이해를 돕기 위해 정보를 제공하기 보다, 학교 혹은 제도에 성취 결과에 대한 정보를 제공하는데 그 무게를 두고 있다. 수행평가는 연속적인 평가와 비연속적인 평가의 상호보완의 중요성을 인식하고, 평가의 목적이나 상황에 부합되도록 다양한 평가 측정 방법에 따라 평가 시기를 다양화한다.

(6) 학습자의 다양한 반응에 대한 수용

Bodin(1993)은 평가가 정답 또는 정답률에만 중점을 둘 것이 아니라, 주어진 과제의 해결 절차는 물론 오답이나 무답의 경우에 대해서도 관심을 기울일 필요가 있다고 주장한다. 이것은 평가에 대한 양적 시각에서 질적인 시각으로 변화하는 것을 의미한다. 정답에 대한 강조는 정답을 찾는데 관련된 방법, 기술, 절차에 대한 학생들의 숙달을 강조한 것이다. 대체적으로, 이러한 강조 때문에 학습의 다른 문제는 이차적인 것으로 여기게 된다. 정답에 대한 완전과 정확성을 강조한 결과로 검사와 시험 조건 아래서 시행된 평가의 결과를 판단할 때, 중요한 요인은 속도, 보다 엄밀히 말해서 효율성이다. 어떤 학생이 일련의 평가 과제를 수행할 수 있는 능력은 있지만 주어진 시간 내에 해내지 못한다면, 그 학생은 능력이 부족한 것으로 판단된다. 이점은 실제와 많이 다를 수 있음이 널리 알려져 있다. 더욱

이, 이러한 평가 유형은 과학적으로는 능력이 있지만 과학적 활동의 속도가 느린 학생들을 좌절시킬 수 있다. 수행평가는 정답이나 오답뿐만 아니라, 학생들의 학습과정에서 표출되는 다양한 양상을 인정하고 의미를 부여한다.

III. 연구내용 및 방법

1. 설문지 개발

자연과 수행평가 실태조사를 위한 설문지는 수행평가의 인식과 실태의 두 영역에서 제작되었다. 초등학교 현장 교사들의 수행평가에 대한 인식은 수행평가의 정의, 필요성, 방법의 측면에서, 그리고 수행평가의 실태는 현장 교사들이 수행평가를 사용하는 평가 방법들의 종류와 빈도, 그리고 수행평가 실시상의 문제점 등에 대한 문항으로 개발하였다.

설문지 개발 과정은 설문 영역을 선택한 후, 이를 토대로 한 문항 개발로 이루어졌다. 문항은 연구진의 합동 작업으로 5차례에 걸친 수정·보완을 통해 1차 설문 문항이 완성되었고, 이를 현장교사들을 대상으로 한 예비 조사(pilot test)를 하여 나온 문제점을 수정·보완하여 최종 설문지를 완성하였다. 개발된 설문지는 총

<표 1> 자연과 수행평가의 실태 조사를 위한 설문 영역 및 문항 구성

영역	내용의 관점	세부 내용	문항 번호
수행평가에 대한 인식	정의	·수행평가에 대한 정의	1
	필요성	·수행평가의 목적	2
		·수행평가의 필요성에 대한 인식도	3
		·평가 결과의 활용	9, 15
	방법	·수행평가의 실시 방법	4
		·다양한 평가 방법에 대한 인지도—실기평가 - 연구과제(프로젝트) 평가	6 12
·정의적 영역의 평가를 위한 방법		10	
수행평가의 실태	방법	·현재 사용하고 있는 수행평가 방법—실행 빈도가 가장 높은 평가 방법	5
		·수행평가를 위한 과제 제시 방법	11
		·평가시 비중도	13
	평가의 비율	·수행평가를 사용하는 기준	7
		·평가의 등급 설정법—실험보고서	8
문제점	·수행평가 실시상의 문제점	14	

15문항으로 구체적인 설문 영역 및 세부 내용과 문항 번호는 <표 1>과 같다.

2. 연구 대상 선정

본 연구는 교육평가 연구 학교와 비 연구 학교에 재직하는 초등학교 교사들을 대상으로 하였다. 전국에서 교육평가에 대한 교육부 또는 교육청 지정 연구 학교를 파악하여, 교육평가 연구학교가 있는 시·도(서울특별시 포함 9개 시·도)의 경우에는 연구학교와 환경이나 규모 면에서 근사한 비 연구 학교를 교육 통계연보(교육부, 1997)를 기초 자료로 하여 본 연구자들의 합의하에 선정하였고, 연구학교가 없는 시·도(대구광역시 포함 6개 시·도)의 경우는 2개교를 무작위 선정하였다. 이러한 방식으로 선정된 연구 대상 학교는 제주도를 제외한 전국적인 규모로, 서울특별시와 6개 광역시 그리고 전국의 8도에서 각 2개교씩 총 30개교를 최종 선정하였다. 본 연구 대상 학교의 총 교사 수는 232명으로서, 각 학교마다 그 학교에서 재직 중인 교사 수의 설문지를 협조 공문 및 재발송 봉투와 함께 우편으로 발송한 후 수거하였다.

3. 설문 결과 분석

설문지에 응답하여 회수된 학교의 수는 26개교(회수율 86.7 %)이었고, 발송된 설문지 959장 중 회수된 것은 707장(회수율 74 %)이었다.

설문 결과의 분석을 위해 SPSS Win 7.5 를 사용하였다.

결과의 분석은 설문지 제작시 고려하였던 설문 대상의 일반적인 정보와 수행평가에 관한 인식과 실태 영역을 중심으로 각 문항 내 빈도와 문항간 관련도를 비교하고 의미를 분석하였다.

본 설문에 응답한 교사는 여자가 66.1%로 여 교사의 비율이 많았으며, 담당 학년은 교과담임 2, 특수학급 1, 기타 1인이 있었으나 1~6학년까지 거의 고른 분포를 보였다. 교직 경력은 5년 미만의 교사가 11.1%로 가장 적었고, 21년 이상 경력자가 47.5%로 거의 반수를 차지하였다. 학교 규모는 18학급 이하의 소규모 학교도 5.8% 있었으나 거의가 중·대규모의 학교였고, 담당

학생 수는 10명 이내가 0.7%, 11~20명이 2.1%, 21~30명이 2.7%, 31~40명이 40.7%이었으나, 40명 이상의 다인 수 학급이 53.8%(378명)를 차지하고 있었다.<표 2>

<표 2> 조사 대상 교사의 구성 분포 (N=707)

설문대상의 일반정보	유형	빈도	타당도 비율(%)
성별	남	238	33.9
	여	465	66.1
담당 학년	1, 2	209	30.0
	3, 4	235	33.8
	5, 6	248	35.7
	기타	4	0.5
교직 경력	5년 미만	78	11.1
	5~10년	111	15.8
	11~20년	180	25.6
	21년 이상	334	47.5
담당 학생수	10명 이내	5	0.7
	11~20명	15	2.1
	21~30명	19	2.7
	31~40명	286	40.7
	40명 이상	378	53.8
학교 규모	18학급 이하	41	5.8
	19~36학급	211	30.0
	37~42학급	257	36.5
	43학급 이상	195	27.7
학교 구분	연구학교	130	18.4
	비연구학교	578	81.6

IV. 결과 및 논의

1. 문항별 빈도 분석

교직경력, 학교 규모, 연구학교와 비 연구 학교에 따른 수행평가의 인식과 방법에 대한 차이를 알아본 결과는 <표 3>에서 보는 바와 같다.

초등학교 현장교사의 수행평가에 대한 인식 영역으로, 수행평가의 정의에 대한 현장교사들의 인지도를 알아보기 위한 문항에서는 ‘다양한 방법을 통해 학생의 능력을 총체적으로 평가하는 것’ 이라는 옳은 답에 답한 교사가 80.8%나 되었다. 이는 설문 조사 대상에서 이전에 수행평가를 실시한 경험이 있는 교사가 많이 포함되어 있었고, 그 동안 초등학교 수행평가에 대한 연수활동 등을 통하여 수행평가의 정의가 자주

언급된 것과 무관하지 않으리라 추측된다. 그러나 수행평가를 행동목표의 성취도 측정을 위하여 많이 사용해 온 '실기평가'와 같은 의미의 평가(6.8%)로 인식하고 있거나, 또는 잘 모르겠다고 답한 교사(1.3%)도 있었다.

수행평가의 목적에 대한 인식을 묻는 문항에서는 '학생들에게 학습과정(체험학습)의 중요성을 인식시키기 위한 것'이라 생각한 교사가 가장 많았으며(43.2%), 다음이 '학습 목표의 도달 정도를 확인하기 위한 것' (35.2%), '효과적인 교수-학습 환경의 조성을 위한 것'(12.9%), '후속 학습의 설계를 위한 기초 자료를 얻기 위한 것' (7.0%) 순으로 나타났다. 한편, '성적의 산출을 위한 것' 이라고 여기는 교사는 극소수에 불과하였다(0.6%). 그리고 수행평가의 실시에 대하여는 대부분 긍정적으로 여기는 것으로 나타났다(89.8%).

실기평가에 대한 인지도를 묻는 문항에서는 "정해진 수업 시간에 실시하는 실험 기구 사용 및 측정 능력 평가를 말한다"라고 바르게 답변 빈도는 낮았으며(13.6%), 오히려 "정규 수업 중에 하는 전반적인 탐구 능력 평가라든가 정규 수업 중에 하는 실험 기구 사용 및 측정 능력 평가"라고 응답한 비율이 높은 것으로 보아(65.0%) 수행평가와 실기평가를 혼동하고 있을 가능성이 높다고 분석된다.

수행평가의 방법에 대한 인식과 실태를 알아보기 위한 문항은 11 가지의 다양한 측정 방법을 제시하고 자유롭게 고르도록 하였다. 한 가지만 고른 교사부터 11 가지를 모두 고른 교사까지 다양하게 나왔으며, 그 중에서도 11가지 모두를 수행평가 방법으로 선택한 빈도가 비교적 높았다. 그리고 실기검사, 관찰법, 개인 연구 과제, 보고서 평가, 포트폴리오의 선택 빈도가 높게 나타났다. 특히 수행평가의 방법에 관한 인식면에서는 연구 학교와 비 연구 학교간에 유의한 차이가 나타났다($p < .01$). 그리고 현재 사용하고 있는 평가의 방법으로는 실기 검사·관찰법(22.0%)이 가장 많았으며, 이 경우도 연구 학

교와 비 연구 학교간에 유의한 차이가 있었다($p < .01$).

수행평가 결과의 활용 실태에 대하여는 '학습 목표의 도달 정도를 확인하기 위한 것' 이라 본 경우가 가장 많았고(24.1%), 다음으로 "학생들에게 학습 과정(체험 학습)의 중요성을 인식시키기 위한 것이다" 라고 응답한 비율이 높았다(14.1%). 특히 이 문항은 교직 경력 면에서 ($p < .05$) 그리고 연구 학교와 비 연구 학교의 경우 유의한 차이가 있었다($p < .01$). 교직 경력의 경우는 교직 경력이 높을수록 위와 같은 비율로 답한 경우가 많았다. 뿐만 아니라, 이 결과는 인식면에서 수행평가의 목적을 후자가 우선적이었던 것으로 나타났던 것을 비교해 볼 때, 현장의 교사들은 수행평가의 목적을 학생들에게 학습 과정(체험학습)의 중요성을 인식시키기 위한 것이라고 파악하고 있지만 실제로 실행할 때는 학습목표의 도달 정도를 확인하기 위한 목적으로 사용한다는 것으로 분석되었다.

또한 정의적 영역의 평가를 위한 방법으로 활용하고 있는 것은 '학생들의 행동 관찰' (68.4%)의 비율이 가장 높게 나타났는데 이 경우에 연구학교와 비 연구 학교간에 유의한 차이를 보였다($p < .05$).

다음으로 수행평가의 실태면에서, 수행평가를 위해 활용하는 과제의 종류에 대한 문항에서는 "교과서에 나와있는 실험을 그대로 활용"하거나(32.0%) "기 제작되어 있는 수행평가자료를 활용"한다(45.1%)는 교사가 많았다. 특히 이 문항의 경우는 교직 경력에 따라($p < .05$), 학교 규모에 따라($p < .05$) 그리고 연구/비 연구 학교에 따라($p < .01$) 모두 유의한 차이가 있는 것으로 나타났다. 이것으로 볼 때 현장에 배포되는 수행평가 문항들이나 자료들의 수준과 타당성을 더욱 높여야 하며, 그 제작자들은 더욱 책임감을 가지고 문항 제작에 노력해야 할 것으로 보인다.

자연과 평가시 비중도에 대해서는 '실험 설계의 정확성, 결과 분석력, 종합력 등과 같은 과학적 탐구 사고 영역'을 중요시한다는 비율이

가장 높았으며(87.6%)나 있었으며, 인지 영역(2.6%)이나 정의적 영역(3.0%), 기능 영역(4.3%)은 적었다. 이 문항은 교직 경력에 따라 또 연구 학교와 비연구 학교간에 유의한 차이를 보였는데(p<.05), 교직 경력이 많을수록 위와 같은 순위의 비중을 가지고 평가에 임하고 있는 것으로 나타났다.

한편, 일반적인 평가 결과의 활용 실태도 수행평가 결과의 활용 실태에 대한 응답과 그 우선 순위가 같았던 것으로 보아 현장의 교사들은 수행평가나 기존의 일반평가나 결국에는 같은 목적으로 활용하고 있는 것으로 분석되었다. 이 문항은 연구학교와 비 연구 학교 간에 유의한 차이가 있었다(p<.05).

<표 3> 설문대상교사의 일반정보와 문항간 유의도 (χ^2 -value)

문항내용	일반정보		
	교직경력	학교규모 (학급수)	연구학교/ 비연구학교
수행평가의 실시방법에 관한 인식			292.877**
사용하고 있는 평가 방법			77.877**
수행평가 결과의 활용도	110.489*		50.275**
정의적 영역의 평가를 위한 방법		43.292*	22.881*
수행평가를 위한 과제 제시 방법	45.667*		26.610**
자연과 평가시 중요시 하는 점	33.703*		
수행평가 실시상의 난점	56.716**		23.579*
일반 평가 결과의 활용도			50.194*

p<.05 p<.01

수행평가의 실시상의 어려움으로는 ‘사전 준비 시간의 부족’ (48.1%)이나 ‘수업 시간의 제약’ (33.3%)을 가장 많이 들었는데, 이 문항은 교직 경력에 따라서(p<.01), 그리고 연구 학교와 비 연구 학교간에(p<.01) 유의한 차이가 있는 것으로 나타났다. 이르 볼 때 교과 과정의 운영면이나 학교 행정면에서 교사들에게 교과 운영의 재량권을 더 부여해 주고, 기타 잡무를 줄여 교수 준비를 위한 절대 시간을 더 확보해 주도록 노력해야할 것으로 파악된다. 그러나 이러한 교사의 준비 시간의 부족이나 수업 시간의 제약과 같은 문제점은 단지 수행평가 실시상의 어려움이라기 보다는 교육의 변화 또는 움직임에 맞출 때마다 언급되는 문제(issues)이다. 또한 최근 연구 결과를 보면, 교육 현장에서의 많은 문제들이 교육 행정적인 문제와 함께 교사의 인식 및 교사 교육으로 야기되고 있음을 지적하고 있다(Hewson, et. al, in press; Meyer, et. al., in press; Lemberger, et. al., in press). 따라서 이러한 문제에 대한 구체적이며 실질적인 많은 연구가 요구된다.

2. 문항간 유의성 검증

수행평가의 정의에 대한 인식과 평가시 중요시하는 점간의 유의성은 <표 4>와 같다.

<표 4> 수행평가의 정의에 대한 인지도와 문항간의 유의도

	χ^2 -value	df	p
평가지 중요시하는 점	14.350	7	.045*

* p<.05

수행평가의 정의에 대한 인식과 평가시 중요시하는 점을 알아보는 문항간에는 유의한 상관관계를 보였으며, 수행평가의 정의를 잘 알고있는 교사는 ‘실험 설계의 정확성, 결과 분석력, 종합력 등과 같은 과학적 탐구 사고 영역을 중요시’ 하는 것으로 나타났다(p<.05).

수행평가의 목적에 대한 인식도와 유의한 상관관을 보인 문항에 대한 결과는 <표 5>와 같다.

<표 5> 수행평가의 목적에 대한 인식도와 문항간의 유의도

	χ^2 -value	df	p
수행평가의 결과 활용도	560.807	290	.000**
평가지 중요시하는 점	556.685	70	.000**
일반 평가 결과의 활용	392.004	310	.001**

** p<.01

수행평가의 목적에 대한 인식도는 수행평가의 결과 활용도($p < .01$), 평가시 중요시하는 점($p < .01$), 일반적 평가 결과의 활용도($p < .01$)간에 유의한 상관이 있었다.

수행평가의 방법에 대한 인지도와 유의한 상관을 보인 문항에 대한 결과는 <표 6>과 같다.

<표 6> 수행평가의 방법에 대한 인지도와 문항간의 유의도

	χ^2 -value	df	p
현재 사용하고 있는 평가 방법 정의적 영역의 평가를 위한 방법 실기평가의 정의에 대한 인지도	14650.79	10485	.000**
	3283.85	2596	.000**
	2252.885	2115	.018*

p<.05 p<.01

수행평가의 방법에 대한 인지도는 사용 중인 평가 방법($p < .01$)이나 정의적 영역의 평가를 위한 방법간에 매우 유의한 상관이 있었다($p < .01$). 즉, 수행평가의 목적에 대한 인식 정도에 따라 평가 결과를 활용하는 용도에 차이가 있었다. 뿐만 아니라 수행평가의 방법에 대한 인지도와 실기평가의 정의에 대한 인지도간에 유의한 상관($p < .05$)이 있는 것으로 보아 초등학교 현장의 교사들은 수행평가와 실기평가와의 차이를 인식하지 못하는 것으로 생각되었다. 따라서 교사의 수행평가 및 수행평가의 목적에 대한 올바른 인식이 교육현장에서의 수행평가 정착에 아주 중요하다 하겠다.

수행평가 결과의 활용 목적과 유의한 상관을 보인 문항에 대한 결과는 <표 7> 과 같다.

수행평가 결과의 활용 목적과 평가시 중요하게 여기는 점간에 유의한 상관이 있었다($p < .01$)

<표 7> 수행평가 결과의 활용에 따른 평가시 비중도

	χ^2 -value	df	p
평가시 중요시 하는 점	342.726	203	.000**

p<.01

한편, 서로 관련성이 있으리라고 예상했던 담당 학년에 따른 수행평가의 방법 및 실시상의 문제점이나, 학교 소재지(도시, 교외)에 따라

사용하고 있는 수행평가의 방법, 과제 제시법, 실시상의 문제점, 그리고 수행평가의 필요성에 대한 인식에 따른 결과의 활용 등에서는 유의성을 보이지 않았다. 통계적 유의성이 나타나지 않은 문항에 대해서는 본 논문의 지면상의 이유로 제시·논의하는 것을 생략하기로 한다.

V. 결론 및 제언

‘초등학교 자연과 수행평가도구의 개발’을 위한 기초 조사로서 초등학교 교사를 대상으로 자연과 수행평가의 실태조사의 결론은 다음과 같다.

첫째, 수행평가의 목적에 대한 인식 정도와 수행평가의 결과 활용도, 평가할 때 중요시하는 점, 일반평가(수행평가가 아닌, 교사가 현재까지 사용하여 온 평가)의 결과를 활용하는 것에 통계적으로 유의한 상관관계를 보였다. 즉 수행평가의 목적에 대한 올바른 인식을 하고 있는 교사는 수행평가 결과를 활용하는데 있어서 학생의 학습과정을 이해하기 위, 평가할 때는 학생의 학습 과정(체험 학습)에 더욱 관심을 두는 것으로 나타났다.

둘째, 수행평가의 실시 방법에 대한 인지도에 따라 사용하고 있는 평가 방법이나 정의적 영역의 평가를 위한 방법이 상관관계를 보였다. 수행평가의 다양한 방법에 대한 인지도가 높을수록 교사가 일반적인 평가 또는 정의적 영역의 평가에 여러 가지 방법을 사용하는 것으로 나타났다.

셋째, 많은 교사들이 수행평가에 대한 올바른 인식을 가지고 있다고 대답하였으나, 실기평가를 정규 수업 중에 하는 전반적인 탐구 능력 평가라고 대답한 것으로 보아 실기평가와 수행평가를 혼돈 또는 혼용하고 있는 교사가 많다는 것을 추측할 수 있다.

넷째, 수행평가의 도구는 교사 스스로가 개발하여 사용하기도 하지만, 교과서상의 실험이나 기계작된 도구를 활용하고 있는 것으로 나타났다.

다섯째, 수행평가의 실시에는 긍정적인 것으로 나타났으나, '사전 준비 시간의 부족' 이나 '수업 시간의 제약' 등을 실시상의 문제점으로 지적했다.

본 연구를 기초로 한 제언은 다음과 같다.

첫째, 수행평가의 교육 현장의 정착을 위해서는 수행평가에 대한 올바른 인식이 필수적이다. 왜냐하면, 수행평가가 단순히 한 가지 형태의 평가이라기보다 학습과정/평가에 대한 패러다임의 변화를 요구하기에 구성주의, 수행평가, 수행평가의 방법 등에 대한 올바른 이해가 필수적이다. 따라서 수행평가에 대한 구체적이며 실질적인 교사 교육이나 교사 재교육이 요구된다.

둘째, 교과서 집필진이나 수행평가 도구 제작자들의 체계적이고 과학적인 연구를 통한 도구 제작이 요구된다. 신뢰도와 타당도가 높은 수행평가 도구를 위하여 교수 자료(teaching material) 및 학습 자료(learning material)에 대한 연구, 그리고 그것과 학생들과의 상호작용에 대한 구체적인 연구가 필요하다.

셋째, 수행평가 실시상의 어려움으로 지적한 교사의 준비 시간의 부족이나 수업 시간의 제약과 같은 문제점은 단지 수행평가 실시상의 어려움이라기 보다는 교육의 변화 또는 움직임을 맞이할 때마다 언급되는 문제(issues)이다. 최근 연구 결과를 보면, 교육 현장에서의 많은 문제들이 교육 행정적인 문제와 함께 교사의 인식 및 교사 교육으로 야기되고 있음을 지적하고 있다. 따라서 이러한 문제에 대한 구체적이며 실질적인 많은 연구가 요구된다.

이 연구는 1998년 한국학술진흥재단 대학부설연구소 연구과제 1차년도 연구비에 의해 연구되었음

참 고 문 헌

1. 교육부(1997). 교육통계연보. 교육부.
2. 국립교육평가원 (1996). 수행 평가의 이론과

실제. 대한 교과서 주식회사.
3. 김동찬 (1991). 지필 평가에서 나타난 학생들의 탐구능력과 실험과정에서 보여주는 탐구능력과 관계. 한국교원대학교 대학원 석사학위논문.
4. 김찬중·김혜정(1998). 초등학교 자연과 포트폴리오 평가의 구성요소. 한국과학교육학회지, 18(2), 233-243.
5. 남명호 (1995). 수행평가의 타당성 연구. 고려대학교 대학원 박사학위 논문.
6. 박현주 (1998). What measured by science performance assessment. 인천교육대학교 과학교육연구소. 과학교육연구논총, 10, 123-137.
7. 백순근 (1998). 수행평가의 이론과 실제. 서울: 원미사.
8. 최병순 외 (1994). 화학 실험 과정에서의 탐구능력 평가 도구 개발, 한국교원대학교부설 교육교육공동연구소, 211-230.
9. Berk, R. A. (Ed.). (1986). Performance assessment: Methods and applications. Baltimore, MD: Johns Hopkins University Press.
10. Brown, J. H. & Shavelson, R. J. (1996). Assessing hand-on science: A teacher's guide to performance assessment. Thousand Oaks, CA: Crown Press, Inc.
11. Buber, M. (1978). Between Man and Man. New York: Macmillan.
12. Champagne, A. B., Lovitts, B. E., & Calinger, B. J. (1990). Assessment in the service of instruction. Washington, D. C.: American Association for the Advancement of Science.
13. Chittenden, E. (1991). Authentic assessment, evaluation, and documentation of student performance. In V. Perrone (Eds.), Expanding student assessment. Alexandria, VA: Association of Supervision and Curriculum Development.
14. Gitomer, D. H. (1989). Developing a portfolio culture that enables learners. Paper pres

- ented at the 1989 National Summit Conference on the Arts and Education, John F. Kennedy Center for the Performing Arts, Washington, D. C.
15. Gitomer, D. H. & Duschl, R. A. (1998). Emerging Issues and practices in science assessment. Barry J. Fraser & Kenneth G. Tobin (Eds.) International Handbook of Science Education, Netherlands: Kluwer Academic Publishers.
 16. Haertel, E. H. (1991). Form and function in assessing science education. In G. Kulm & S. M. Malcom (Eds.), Science assessment in the service of reform.
 17. Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). A practical guide to alternative assessment, Selecting Assessment Tasks. U. S.A : University of California.
 18. Hewson, P. W., Tabachnick, B. R., Zeichner, K. M., Blomker, K. B., Meyer, H., Lemberger, J.,
 19. Lazzaro, C. & Park, H. (1994). What' measured by science performance assessment?: A look at student competence, engagement, creativity, confidence, outgoingness, and disruptiveness as correlates of assessment. Wisconsin Performance Assessment Development Project Report.
 20. Lemberger, J., Hewson, P. W., & Park, H. (1999). Relationships between prospective secondary teachers' classroom practice and their conceptions of biology and of teaching science. Science Education., 83(3): 347-371
 21. Meisels, S. L. (1993). The work sampling system: An authentic performance assessment. Principal, 72(5), 5-7.
 22. Meyer, H., Tabachnick, B. R., Hewson, P. W., Lemberger, J., & Park, H. (1999). Relationships between prospective elementary teachers' classroom practice and their conceptions of biology and of teaching science. Science Education, 83(3): 323-346.
 23. Perrone, V. (Eds.) (1990). Expanding student assessment. Alexandria, VA: Association of Supervision and Curriculum Development.
 24. Stiggins, R. J. (1991). Facing the challenges of a new Era of educational assessment. Applied Measurement in Education, 4(4): 263-273.
 25. WECA(1993). Wisconsin Performance Assessment Development Project: Brief report on goals, plans, and progress. WI: Wisconsin Center for Educational Research.
 26. Wiggins, C. (1993). Assessment: Authenticity, context and validity. Phi Delta Kappan, 200-214.
 27. Wolf, D. P. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant(Ed.) Review of Research in Education, vol. 17, 31-74. Washington, D. C.: American Educational Research Association