

과학탐구 실험대회의 문제점 분석

백성혜

(서울대학교)

(1995년 1월 17일 받음)

I. 서론

본 연구에서는 1993년 '과학교육의 해'에 처음 개최되어 2년째 계속되고 있는 '학생 과학탐구 올림피아드대회'의 8개 단위 대회 중 한 개인 '과학탐구 실험대회'의 문제점을 분석하고자 한다. 1993년도 과학탐구 실험대회의 문제점을 분석한 논문(백성혜, 박승재, 1994)에서는 주로 평가 문항의 문제점이 지적되었다. 즉, 지필평가의 경우에는 평가 영역의 편재, 평가 문항의 고난이도, 문항 진술의 문제점 등이 지적되었으며, 실험 평가의 경우에는 평가 기준 설정과 배점의 문제점이 지적되었다.

1993년도 실험대회의 평가에서는 대회가 끝난 후에 평가자를 선정하여 실험대회의 전반적인 운영을 파악하거나 대회 현장에 직접 참석할 기회가 없었기 때문에 그 이상의 문제점 분석이 어려웠다. 그러나 1994년도 학생 과학탐구 올림피아드 대회는 두번째로 열리는 대회이기 때문에 일 년 전부터 미리 평가자를 선정하였다. 따라서 이번 대회의 경우에는 평가자가 실험대회의 전반적인 운영이나 대회 현장에서 파악한 내용 등을 분석할 수 있었다.

특히 1994년도 실험대회에서는 1993년도에 지적된 문제점을 해소하기 위하여 평가 문항에 대한 연구가 집중적으로 이루어졌기 때문에 이번 대회의 문제점 분석에서는 평가 문항 자체의 문제보다는 평가 항목간의 관계, 학생들의 구체적인 실험 활동 내용, 운영의 문제점 등을 폭넓게 다루고자 한다.

해마다 여러 과학관련 기관들의 주최로 열리는 각종 과학대회들이 개최된 후에 이러한 문제점 분석 등에는 관심을 기울이지 않고 형식적인 보고서 제출로 마감하기 때문에,

참가자들이나 관람자들에게 '그 대회가 그 대회'라는 인식을 주며 예산 낭비를 초래하고 궁극적으로는 학교 과학교육의 개선에 아무런 도움을 주지 못하는 경우가 있다.

따라서 본 논문에서 시도한 과학탐구 실험대회의 문제점 분석은 앞으로 계속 열리게 될 실험대회와 이와 유사한 다른 과학 대회들에게 보다 나은 의사결정의 토대를 제공해 줄 것이라는 점에서 매우 필요한 시도라고 본다.

II. 이론적 배경

오늘날 과학교육에서 탐구는 매우 중요한 위치를 차지하고 있으며, 이를 학생들에게 가르치기 위한 가장 효과적인 방법으로 실험의 역할이 크게 강조되었다(Leonard, 1983; Shymansky, Kyle and Alport, 1983; Tobin, 1990). 과학탐구 실험을 학생 스스로 과학의 개념과 방법을 경험하는 것이라고 정의할 때(Leonard, 1988), 학생들이 경험하는 내용은 아는 것과 하는 것으로 구분할 수 있다.

과학을 아는 것과 하는 것은 서로 밀접한 관련을 가지고 있으면서도 매우 다른 측면이다. 그리고 이 두가지 측면은 과학교육에서 모두 중요하다. 이를 Woolnough(1989), Fairbrother(1989) 등은 과정과 기능으로 묘사하였다. 또한 그들은 과학을 잘하기 위해서 과학의 명확한 지식뿐만 아니라 수공적인 실험기능도 매우 중요하다고 지적하였다.

실험을 설계하고, 자료를 수집하고, 정리하고, 이를 해석하여 결론을 내리는 능력 등을 과정이라고 한다면, 이러한 과정은 측정하거나 기구를 조작하는 학생의 개인적인 실험 기능에 크게 좌우될 것이다. 그러나 이러한 수공적 실험 기능은 학생이나 과학자가 문제를 해결하는 과정에 관한 능력

과는 관련이 없는 독립적인 능력이다.

따라서 과학탐구능력이 뛰어난 학생들을 선발하는 목적을 가진 과학탐구 실험대회에서는 과학적 탐구과정뿐만 아니라 이러한 과정에 큰 영향을 미치는 수공적 실험기능에 관한 중요성도 부각되어야 할 것이다.

한 걸음 더 나아가 Woolnough(1989), Bates(1985) 등은 과학을 잘하기 위해서 인지적 능력과 수공적 실험 기능 이외에도 몰두, 경향, 자신감, 직관력과 같은 정의적 능력과 형식적, 비형식적 학습을 통해 형성되는 개인적 통찰력, 과학의 성질에 대한 이해 등도 중요하다고 강조하였다.

실제적으로 한 대회의 내용을 결정하는 것은 평가 방법이라고 할 수 있다. 그러나 학생들이 과학을 잘하는가를 알아보기 위하여 현재까지 개발된 평가 방법은 많은 문제점들을 안고 있다. 특히 정의적 영역이나 개인적 통찰력 등에 관한 평가 방법은 현장에 적용하기가 매우 어려운 실정이다.

과학적 탐구과정과 수공적 실험기능에 관한 평가 방법은 Doran(1980), Doran et al.(1992), Kempa(1986), Swain(1988), Tobin(1990), 백성혜, 박승재(1994) 등 많은 사람들의 연구물에서 다양한 형태로 제안되었다. 이 중 몇몇은 엄격한 평가의 규정과 정확성을 선호하고, 다른 이들은 좀더 예민하고 가변적인 평가를 선호하였다.

이러한 과학탐구에 관한 평가의 두가지 측면은 각각 장점과 단점을 가진다. 엄격한 평가의 규정을 강조하는 경우에는 보다 객관적인 결과를 얻을 수 있는 장점이 있는 반면, 지나치게 특정 활동을 강조하다가 포괄적이고 전반적인 목표를 간과할 위험이 있다. 한편 학생의 학습 결과에 가변적인 평가 방법은 이와는 반대되는 문제점을 안고 있다.

Fairbrother(1989)는 학생들이 과학을 탐구하는 과정과 그로부터 나오는 결과가 다양한 요인에 의해 영향을 받기 때문에, 규격화된 평가들에 근거한 수치만으로 학생들의 탐구능력을 평가하는 것은 적절하지 못하다고 주장하였다. 또한 그는 수공적 실험기능을 평가하는 경우에 두가지 기본적인 접근법이 있다고 주장하였다.

그 첫째는 학생들의 능력을 위계적으로 구분하여 점수화하지 않고, 단순히 학생들이 한 것이 무엇인가를 묘사하는 것이다. 그러나 이 방법은 학생들을 객관적으로 비교하는 척도를 제공해 주지 못한다는 단점이 있다.

두번째 접근법은 학생들의 실험기능을 단계별로 나누어 각 단계마다 적절한 점수를 부여하는 형태로 평가하는 것이다. 이 방법은 학생의 전체적인 점수를 결정하는데 도움을 주지만, 각 단계마다 부여한 점수가 학생들의 능력을 그대로 반영하는 것이 아니라는 문제점을 안게 된다. 즉 4점은 2점보다 2배 진전된 학생들의 능력을 의미하지 않지만, 이

러한 방법은 비례 척도의 개념으로 잘못 이해될 우려가 있다.

또한 이 접근법은 완벽한 절대 평가적인 위계를 찾는 것이 불가능하다는 문제점을 갖는다. 즉, 아무리 상세하게 학생의 능력에 대한 위계를 구분한다 하더라도 학생들은 이러한 분석틀과 똑같이 행동하지 않는다는 것이다. 학생들은 능력을 구분한 위계의 내용 중에서 일부분만 일치하는 행동을 하거나, 여러 위계에 동시에 해당하는 행동을 보이기도 할 것이다. 따라서 평가자는 실제 학생들의 행동을 평가할 때 매우 어려움을 겪게 되며, 객관적인 평가를 안에서도 자신의 주관울 내포한 평가를 내리는 경우가 생기게 된다.

이러한 문제점을 극복하기 위하여 두가지 방법이 사용되고 있다. 즉 예와 아니오의 선택 보기와 같이 단순하게 학생의 능력에 대한 위계를 설정하는 것이다. 그러나 이 경우에도 학생들의 구체적인 행동은 정확한 판단을 내릴 만큼 명확하지 못한 경우가 있기 때문에 평가자에게 어려움을 줄 수 있다. 또다른 방법은 가능한 한 많은 선택 보기를 가진 위계 척도를 개발하는 것이다. 그러나 이 경우에도 평가자에게 너무 많은 판단을 하게 하여서 잘못된 세밀감을 제공하게 될 우려가 있다. 따라서 이러한 평가 방법은 교사들의 경험과 연구가 필수적이다.

학생들의 탐구능력을 평가하는데 있어서 고려해야 할 또 다른 점은 평가자가 학생들의 행동을 평가할 수 있는 기회가 충분히 제공되어야 한다는 점이다. 매우 짧은 시간 안에 이루어지는 수공적 실험기능의 경우에는 한 평가자가 한 명의 학생을 계속 주시하면서 평가를 하지 않는 한 평가의 기회를 놓칠 우려가 있다. 대부분의 실험 평가에서 한 명의 평가자가 여러 명의 학생들을 동시에 평가하므로 이러한 점을 고려하여 실험과 평가의 내용을 계획할 필요가 있다.

이를 위한 한가지 방법은 영구히 평가할 수 있는 활동과 그렇지 않은 활동을 구분하는 것이다. 기록하거나 보고하는 활동들은 영구히 보관이 가능하므로 실험이 끝난 후에 평가할 수 있다. 그러나 수공적 실험기능에 해당하는 내용들은 매우 짧은 시간 내에 일어나며, 실험이 진행되는 과정에서만 평가가 가능하다.

또한 많은 학생들의 실험 능력을 두세 명의 평가자들이 나누어 평가하는 경우에는 학생들의 구체적인 행동들을 간략하게 묘사해 둘 필요성이 있다. 이는 여러 평가자들이 공통적인 평가 기준을 설정하는 경우에 도움을 줄뿐만 아니라, 한 평가자가 여러 학생들을 일관성 있게 평가할 수 있는 근거를 제공해 주기 때문이다. 따라서 이러한 평가 활동들은 평가의 공정성을 위해 반드시 고려되어야 한다.

학생들의 탐구 실험 능력을 평가할 때 발생하는 또다른

문제점으로는 실험이 진행되는 동안에 먼저 실험을 수행한 학생들로부터 직접적, 혹은 간접적인 방법으로 정보가 유출될 수 있다는 점이다. 따라서 학생들이 다른 학생의 탐구 실험 과정과 기능을 모방함으로써 평가의 결과에 영향을 미치지 않도록 평가의 내용을 신중하게 결정할 필요가 있다.

III. 연구 내용 및 방법

국민학교 6학년 학생 43명과 중학교 2학년 학생 40명을 대상으로 실시된 1994년도 과학탐구 실험대회는 크게 지필평가와 실험평가로 구분된다. 지필평가는 28 개의 객관식 문항(70점 배점)과 4 개의 주관식 문항(30점 배점)으로 구성되어 있으며, 실험은 2개의 실험을 각각 50분씩 실시하고 이를 보고서와 점검표로 평가하였다.

본 연구는 크게 3 부분으로 나누어진다.

첫째는 지필평가 결과의 분석이다. 여기서는 지필평가 객관식 문항의 변별도 지수를 알아보았다. 이를 위해 Johnson의 DI 방법(황정규, 1985)이 사용되었으며, 전체 학생 수에서 상위 집단 27%와 하위 집단 27%를 나누는 방법이 사용되었다. 자료 분석을 위해서는 MINITAB 통계 패키지가 사용되었다.

둘째는 실험 평가의 분석이다. 여기서는 실험 평가 항목 및 배점, 학생들의 구체적인 실험 활동 내용, 실험 평가 방법, 실험 평가 항목간의 관계 등이 분석되었다.

(1) 실험 평가 항목은 크게 수공적 실험기능과 탐구과정으로 구분하였으며, 실험 점검표와 보고서의 평가 내용유에 근거하여 분류하고 배점의 문제점을 살펴보았다.

(2) 학생들의 구체적인 실험 활동 내용에서는 실험 점검표의 평가 척도에 포함되지 않았으나 평가에 고려해야 할 학생들의 실험 활동의 특징들을 간략하게 정리하였다.

(3) 실험 평가 방법에서는 실험에 따라 다르게 이루어진 평가 방법의 문제점을 살펴보고, 동일한 실험 내에서도 평가자에 따라 평가 점수의 간격이 크게 벌어지는 점 등을 그래프로 분석하였다.

(4) 실험 평가 항목간의 관계에서는 자료 수집이 가능하였던 화학 실험 보고서와 점검표의 평가 항목 점수들을 가지고 상관관계를 구하였다. 이를 위해 SAS 통계 패키지가 이용되었다.

셋째는 실험대회 전체 평가 결과의 분석이다. 여기서는 국민학교와 중학교의 지필평가 객관식 문항과 주관식 문항, 그리고 실험 보고서와 점검표 점수들 간의 상관관계를 알아보았다. 이를 위해 역시 SAS 통계 패키지가 사용되었다.

IV. 연구 결과 및 논의

1. 지필평가 결과의 분석

실험대회의 지필평가는 객관식 28문항과 주관식 4문항으로 이루어졌다. 국민학교와 중학교 지필평가 객관식 문항의 변별도는 Johnson의 DI(황정규, 1985)로 계산하였으며, 가장 보편적인 방법으로 전체 학생수에서 상위 집단 27%, 하위집단 27%로 나누는 방법을 사용하였다. 계산한 결과를 <표 1>과 <표 2>에 제시하였다.

<표 1> 국민학교 지필평가 객관식 문항 변별도(문항수)

변 별 도	.00	.08	.17	.25	.33	.42	.50	.58	.67	.75
문항수(개)	5	4	3	2	4	4	2	1	2	1

<표 2> 중학교 지필평가 객관식 문항 변별도(문항수)

변 별 도	-.09	.00	.09	.18	.27	.36	.45	.64
문항수(개)	2	2	5	8	4	3	3	1

김창식 외 4인(1993)에 따르면 일반적으로 변별도 지수가 0.20보다 낮으면 변별력이 좋지 못하므로 검사에서 제외하거나 수정하는 것이 필요하다. 그리고 변별도 지수가 0.30 이상인 경우에 변별력이 좋다고 하였다. 국민학교 지필평가 객관식 28 문항 중에서 변별도 지수가 0.30 이상인 문항이 절반 수준인 14 문항이었으며, 평균 변별도는 0.29 이었다. 그리고 변별도 지수가 0.20보다 낮은 문항이 전체 28 문항 중에서 12 문항이었다. 따라서 객관식 문항의 변별도 지수가 다소 낮다고 말할 수 있다.

중학교 지필평가 객관식 28 문항 중에서 변별도 지수가 0.30 이상인 문항은 7개이며, 평균 변별도는 0.21 이었다. 또한 변별도 지수가 0.20 보다도 낮아 검사에서 제외할 필요가 있는 문항은 17 개나 되었다. 따라서 전반적으로 국민학교보다도 문항의 변별도 지수가 낮다고 말할 수 있다.

문항의 변별도는 난이도와 밀접한 관련이 있다. 즉, 난이도가 50%일 때 변별도는 최대가 되며 난이도가 이보다 높아지거나 낮아지면 변별도는 감소한다. 따라서 변별도 지수가 낮다는 의미는 문항의 난이도 설정에 문제점이 있었음을 의미한다.

2. 실험 평가 결과의 분석

1) 실험 평가 항목 및 배점

국민학교 실험은 생물과 지구과학으로 구분되며, 각각 점검표와 보고서를 평가하였다. 이 때 평가한 항목과 배점은 문항 출제자가 정하였으므로, 생물과 지구과학의 경우에 차이가 있었다. 본 연구에서는 구체적인 평가 항목의 내용을 탐구능력 분류틀로 분석하여 이에 해당하는 배점과 함께 <표 3>에 제시하였다.

<표 3>에서 국민학교 실험 평가 항목을 8개로 구분할 때 '실험 설계' 항목의 배점이 36점으로 가장 높으며, '가설 설정'과 '결과 해석'의 배점이 각각 6점으로 가장 낮음을 알 수 있다. '실험 설계' 항목의 배점이 매우 큰 이유 중 하나는 생물과 지구과학 실험 모두 점검표와 보고서에서 중복하여 평가하였기 때문이다.

<표 3> 국민학교 실험 평가 항목 및 배점

실험평가 평가항목	생물		지구과학		배점
	점검표	보고서	점검표	보고서	
가설 설정		6			6
실험 설계	10	6	12	8	36
자료 수집	5		3		8
자료 처리		12			12
결과 해석		6			6
결 론				8	8
적 용				16	16
수공적 실험 기능	5		3		8

중학교 실험은 물리와 화학으로 구분되며, 역시 점검표와 보고서로 구분하여 평가되었다. 평가 항목과 배점을 분석하면 <표 4>와 같다.

<표 4>의 중학교 실험 평가 항목 중에서 '자료 처리' 항목의 배점이 27점으로 가장 높으며, '가설설정'과 '적용'의 항목은 국민학교와 달리 평가되지 않았음을 알 수 있다. '자료 처리' 항목의 배점이 큰 이유는 물리 점검표와 보고서에서 각각 10점씩 중복하여 평가하였기 때문이다. 같은 항목의 중복 평가가 '실험 설계' 항목에서도 있었다. 따라서 이 항목의 배점도 25점으로 다른 항목에 비해 점수 비중이 높은 것으로 나타났다.

<표 4> 중학교 실험 평가 항목 및 배점

실험평가 평가항목	물리		화학		배점
	점검표	보고서	점검표	보고서	
가설 설정					0
실험 설계	5	5		15	25
자료 수집	10			6	16
자료 처리	10	10		7	27
결과 해석		5			5
결 론		5		4	9
적 용					0
수공적 실험기능			18		18

실험 현장에서 평가가 가능한 영역과 실험이 끝난 후 학생들이 제출한 보고서를 통해 평가가 가능한 영역을 구분하여 서로 중복된 평가를 피해야 효과적인 실험 평가가 이루어질 수 있다. 특히 실험 현장에서만 평가가 가능한 영역으로는 '수공적 실험 기능'을 들 수 있는데, 화학 실험을 제외한 다른 실험에서는 실험 현장에서 평가가 이루어지는 점검표에서 이에 대한 배점이 크지 않았다.

중복된 평가가 이루어진 경우에도 배점을 조정하여 평가 영역간의 균형을 이룰 수 있으나, 실험 시간에 2-3명의 평가자가 40명 가량의 학생들을 대상으로 1시간 이내에 점검표에 설정된 많은 항목들을 모두 평가해야 하는 어려움을 고려한다면 중복된 평가는 바람직하지 않다고 본다.

Woolnough(1989)과 Fairbrother(1989)의 주장대로 학생의 탐구 실험 능력을 과학적 탐구과정과 수공적 실험 기능으로 구분할 때, 국민학교의 경우 두 평가 영역의 점수 배점 비율이 생물 실험의 경우에는 9:1, 그리고 지구과학 실험의 경우에는 15.7:1로 매우 차이가 큼을 알 수 있다. 중학교 물리 실험의 경우에는 수공적 실험기능에 대한 평가가 전혀 없었다. 단지 화학 실험의 경우에 과학적 탐구과정과 수공적 실험기능간의 배점 비율이 1.8:1로 4개의 실험 중에서 가장 차이가 적었다. 즉 이번 대회의 경우 실험 평가에서 과학적 탐구과정에 비해 수공적 실험기능을 상대적으로 덜 강조하였다고 말할 수 있다.

국민학교 생물 실험의 경우에는 점검표 전체 배점(20점)의 25.0%에 해당하는 8점이 수공적 실험기능에 배점되었으며, 지구과학의 경우에는 전체 배점(18점)의 16.7%에 해당하는 3점이 배점되었다. 그리고 중학교 물리 실험의 경우

에는 점검표에서 수공적 실험기능에 대한 배점은 없었다. 이에 비해 중학교 화학 실험 점검표에서는 전체 배점(18점)의 100%를 수공적 실험기능에 할당하였다.

생물 실험의 경우에는 실험을 소홀히 하면서 보고서 작성에만 치중하는 경우에 실험 평가에서 높은 점수를 받게 되는 것을 방지하기 위하여 실험을 소홀히 한 학생들에게는 보고서 점수도 기본적으로 낮게 채점하였다. 그러나 이는 '자료 수집'이나 '수공적 실험기능' 등 구체적인 실험 활동에 해당되지 않으나, 탐구 실험 중에서 매우 중요한 부분들인 '가설설정', '실험 설계', '자료 처리', '결과 해석', '결론', '적용' 등의 평가 영역을 소홀히 생각하고 평가자들이 내린 조치라고 볼 수 있다.

물론 거의 실험을 하지 않은 학생들이 '자료 수집'의 내용을 조작하거나 '자료 처리', '결과 해석', '결론' 등의 평가 영역을 실험에 근거하지 않고 작성할 수 있는 경우에 대비한 조처이기는 하지만, 기본적으로 학생들의 그러한 조작이 불가능한 실험 문제를 출제하는 것이 보다 바람직하다고 본다.

화학실험 점검표의 경우에는 다른 실험과는 달리 보고서의 평가 영역과 점검표의 평가 영역을 명확히 구분하였으며, 점검표에서는 학생들의 수공적 실험기능에 대한 평가만이 이루어졌다. 이는 화학 실험 내용이 세가지 물질의 용해도를 알아보는 것으로, 실험을 설계하는 과정은 평이하나 실제로 자료를 수집할 때 이루어지는 수공적 실험기능의 차이에 따라 실험 자료의 수집과 결과의 해석, 그리고 결론이 매우 달라질 수 있었기 때문이다. 따라서 수공적 실험기능에 대한 배점을 전체 화학 실험 평가 배점(50점)의 36.0%에 해당하는 18점으로 설정하고 점검표에서는 집중적으로 이에 대한 평가를 하였다.

2) 학생들의 구체적인 실험 활동

모든 실험의 경우에 평가는 세 명의 심사자들에 의해 이루어졌는데, 모든 실험이 동시에 실시되었으므로 한 평가자가 4개의 실험에 모두 참석하는 것은 불가능하였다. 본 연구자는 화학 실험에만 참가하였으며, 이에 관한 자료만을 분석하여 정리하였다.

실제 학생들의 실험 활동은 화학 실험 점검표에 제시된 평가 척도에는 구체적으로 명시되지 않은 내용들이 포함되었다. 따라서 이에 대한 평가자들의 판단에는 주관이 개입될 가능성이 있었다. 최종적으로 평가자들의 평가 내용을 비교함으로써 평가 결과에 공정성을 기하기 위하여 학생들의 구체적인 실험 활동을 점검표 점수 옆에 기재하였다. 그 내용을 정리하면 다음과 같다.

화학 실험의 제목은 '주어진 세 물질의 용해도 측정'이었

는데, 실험 점검표는 5항목으로 구분되었다. 그 중 첫째 항목은 저울 사용법에 관한 것이다. 이에 관한 실험 점검표의 평가 척도는 <표 5>와 같이 설정되었다.

<표 5> 화학 실험 점검표 첫째 항목의 평가 척도

저울 사용법이 올바른가?(3점 배점)
- 0점 조정, 분동의 이용 등 저울 사용법이 바른 경우(3점)
- 저울 사용법에서 1가지가 잘못된 경우(1점)
- 저울 사용법에서 2가지 이상 잘못된 경우(0점)

저울 사용법은 질량을 측정해야 하는 모든 실험의 기본이다. 특히 이 화학 실험은 중학교 학생들을 대상으로 하였으며 대회에 참가한 학생들이 대부분 과학 우수자로 시,도 교육청의 추천을 통해 선발되거나 다른 실험대회의 성적 우수자가 선발되었다. 따라서 실험 점검표의 첫째 항목은 매우 수준이 낮은 실험평가 영역으로 과학 탐구 실험에 우수한 학생을 변별하는 능력이 떨어지리라고 예상하였다.

그러나 주어진 50분 안에 같은 실험을 여러 번 반복하여 주어진 세가지 물질의 용해도를 계산해야 하는 실험이었으므로, 시간이 흐를수록 학생들은 당황하기 시작하여 저울 사용법의 가장 기본적인 수칙마저도 무시하곤 하였다. 이는 이 대회에 참가한 많은 학생들이 평소 많은 실험을 통해 이러한 기본적인 실험기능을 몸에 익히지 못하였음을 의미한다.

그 대표적인 예로 실험이 진행되면서 많은 학생들이 시료를 용지에 담아서 양팔 저울의 한 쪽에 올려놓은 다음, 이에 대한 질량을 측정하기 위하여 분동을 올려놓은 다른 쪽 접시에는 같은 크기의 용지를 함께 놓지 않았다. 이러한 변인 통제에 대한 기본적인 실수는 매우 정밀한 측정 값을 요구하는 이번 실험에서 오차를 유발하는 요인이 되었으나, 이 점을 자신의 오류로 지적한 보고서조차 없었다.

학생들의 양팔 저울 사용법에 대한 또다른 대표적인 실수는 실험을 처음 시작할 때에는 영점 조절을 하였으나 그 후에는 이에 대해 신경을 쓰지 않는다는 점이었다. 용질의 질량을 측정하는 실험이 반복되면서 점차 양팔저울의 영점이 맞지 않게 되었는데도 매번 영점 조절 후에 용질의 질량을 측정하는 학생은 매우 드물었다.

또한 실험 마감 시간이 촉박해지자 학생들 중에는 마음이 급하여 분동을 집을 때 핀셋을 사용하지 않고 손으로 집어 저울 위에 올려놓는 경우도 종종 눈에 띄었다.

화학 실험 점검표의 둘째 항목은 메스실린더나 스포이드

의 사용에 관한 것이다. 이에 관한 실험 점검표의 평가 척도는 <표 6>과 같이 설정되었다.

<표 6> 화학 실험 점검표 둘째 항목의 평가 척도

-
- 메스실린더나 스포이드의 사용이 올바른가?(3점 배점)
 - 메스실린더나 스포이드의 사용이 모두 바른 경우(3점)
 - 1가지의 사용이 잘못된 경우(1점)
 - 2가지 모두의 사용이 잘못된 경우(0점)
-

메스실린더나 스포이드의 사용 역시 용액의 부피를 측정해야 하는 모든 실험에 기본이 되는 수공적 실험기능이다. 일반적으로 눈금이 있는 스포이드는 5 mL 이하의 용매를 측정할 때 사용하며, 그 이상의 용매를 측정할 때에는 메스실린더를 사용한다. 또한 스포이드 사용시 방울 수를 세는 것은 1 mL 이하의 작은 부피를 잴 때 쓰는 기법이다. 그러나 이번 실험에서 많은 학생들이 이러한 개념을 갖고 있지 않았다.

어떤 학생들은 10 mL 이상의 용매를 측정하면서도 메스실린더를 전혀 사용하지 않고 스포이드만을 사용하였다. 이들은 메스실린더를 단지 용매를 담는 용기로만 사용하는 것 같았다. 다른 학생들 중에는 메스실린더로 용매의 부피를 측정 후 다시 스포이드의 눈금으로 용매의 부피를 측정하면서 시험관에 옮기기도 하였다. 심지어 한 학생은 보고서의 자료 수집란에 용매의 구체적인 부피는 기록하지 않고 스포이드의 방울 수를 기록하였는데 70- 100 방울 정도씩을 기록한 경우도 있었다.

그들은 아마도 스포이드가 보다 정밀한 용액의 부피를 측정하는데 유리할 것이라고 생각하였을 것이다. 그러나 그러한 수고는 스포이드를 여러 번 사용하는 데서 초래되는 오차의 증가를 생각하지 못한 것이다. 아울러 제한된 시간 내에 실험을 마쳐야 하는 경우에 시간의 낭비를 초래하여 반복 실험 등 보다 정확한 자료 수집을 위하여 이루어져야 하는 실험 과정의 수행을 방해할 것이다. 실제로 많은 학생들이 단지 3가지 물질의 용해도도 측정하는 이번 화학실험을 주어진 50분의 시간 내에 마치지 못하였다.

이러한 실험의 정밀성에 대한 집착과는 대조적으로 어떤 학생들은 메스실린더로 측정된 물을 시험관이나 비이커에 옮겼다가 생각을 바꾸어 이리 저리 다른 용기에 옮기면서 용매를 흘리는 경우가 발생해도 용매의 줄어든 부피에는 신경을 쓰지 않았다. 이러한 학생들은 실험 점검표의 넷째 항목인 '증류수를 시험관에 정확하게 모두 옮기는가?(3점)'에

서 확인되었다.

또한 용매로 증류수가 제공되었음에도 불구하고 수도물을 용매로 사용하거나 메스실린더에 용질을 넣고 녹이는 등 기본적인 실험기능의 결여도 눈에 띄었다.

화학 실험 점검표의 셋째 항목은 고체시료의 취급에 관한 것이다. 이에 관한 실험 점검표의 평가 척도는 <표 7>과 같이 설정되었다.

<표 7> 화학 실험 점검표 셋째 항목의 평가 척도

고체시료를 주의하여 취급하는가?(3점 배점)

-
- 물질마다 시약스폰을 다르게 사용하고, 시료를 흘리지 않는 경우(3점)
 - 위의 내용 중 1가지가 잘못된 경우(1점)
 - 위의 내용 중에서 2가지 모두가 잘못된 경우(0점)
-

위에서 제시한 경우 이외에도 어떤 학생들은 한번 사용한 시료 용지를 다른 시료에 재사용하기도 하였다. 그리고 어떤 학생들은 시료를 비이커나 시험관에는 흘리지 않고 신중하게 옮겼으나, 시료를 부으면서 용기의 입구 쪽에 시료가 많이 붙게 되어도 이를 용매로 녹이려는 시도를 하지 않았다. 심지어 어떤 학생들은 유리막대 대신에 시약 수저를 것개로 사용하기도 하였다.

많은 학생들이 용매를 시험관이나 비이커에 따를 때에는 유리막대를 사용하지만, 정작 시료를 녹일 때에는 유리막대를 사용하지 않았다. 대부분이 손으로 시험관이나 비이커를 흔드는 정도에 그쳤다. 이번 실험이 용해도 측정 실험임에도 불구하고 용질과 용매를 넣은 후에 용질을 녹이려는 시도를 하지 않고 그냥 놔두는 경우도 있었다. 따라서 많은 학생들이 시료를 완전히 녹이지 않고 시험관 밑에 많이 가라앉은 상태에서 용해도 실험을 마쳤다.

화학 실험 점검표의 넷째 항목인 '증류수를 시험관에 정확하게 모두 옮기는가?' 와 다섯째 항목인 '실험이 끝난 다음 모든 기구와 시약들을 바르게 처리하는가?' 는 출제자가 제시한 평가 척도를 크게 벗어나는 학생들의 행동이 관찰되지 않았으므로 논의에서 제외하였다.

3) 실험 평가 방법

이번 대회에서는 국민학교와 중학교의 4가지 실험에 대한 평가 방법이 모두 달랐다.

국민학교 생물 실험의 경우에는 3명의 평가자가 모두 43명의 학생들을 점검표로 중복하여 평가하였으며, 따라서 한

학생 당 3종류의 점검표 평가 점수가 얻어졌다. 그리고 최종적인 점검표 평가 점수는 이를 평균하여 구하였다. 그러나 보고서는 3명의 평가자가 공동으로 채점하여 동일한 한 개의 평가 점수를 구하였다.

국민학교 지구과학 실험의 경우에는 3명의 평가자가 43명의 학생들을 3등분하여 각기 점검표와 보고서를 평가하였으므로 동일한 학생에 대한 중복 평가는 이루어지지 않았다.

중학교 물리 실험의 경우에는 국민학교 생물 실험과 마찬가지로 3명의 평가자가 모두 40명의 학생들을 중복하여 평가하였다. 그러나 이 경우에는 점검표뿐 아니라 보고서도 중복 평가하여 최종적으로 평균을 구하였다.

중학교 화학 실험의 경우에는 국민학교 지구과학 실험과 마찬가지로 3명의 평가자가 40명의 학생들을 3등분하여 점검표로 평가하였으며, 보고서는 3명의 평가자가 공동으로 채점하여 동일한 한 개의 평가 점수를 구하였다.

따라서 실험 점검표와 보고서의 평가 방법에 통일성이 없었다.

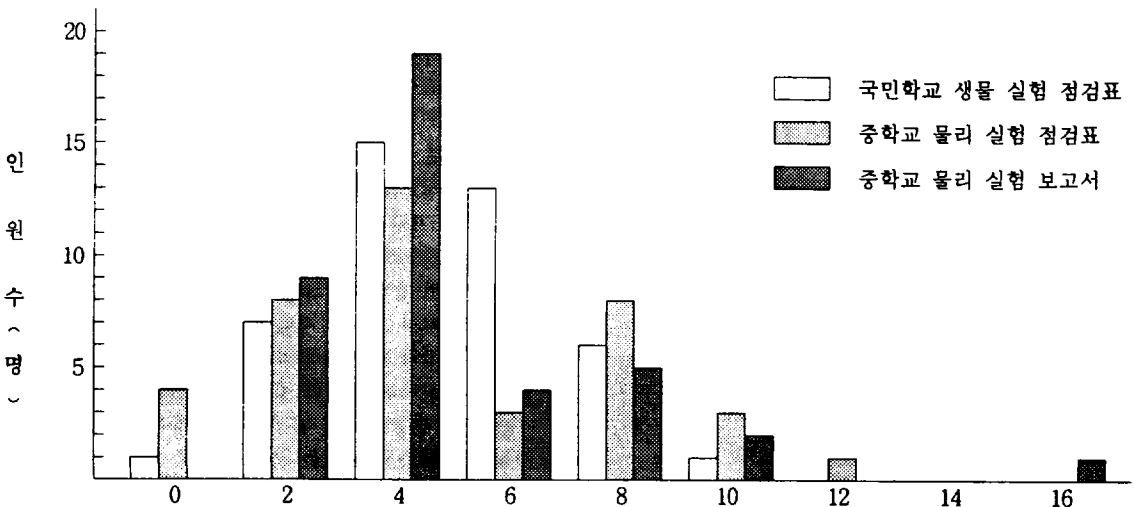
또한 실험 점검표와 보고서의 평가 기준의 명시는 평가의 객관성을 유지하기 위한 것임에도 불구하고 동일한 항목에 대해 평가자들 간의 견해 차이가 매우 큰 것으로 나타났다. 이러한 점은 같은 학생을 3명의 평가자가 중복하여 평가한 경우에 부여한 점수의 차이에서 확인할 수 있었다.

<그림 1>은 동일 학생의 실험 점검표와 보고서에 대해 평가자 3명이 부여한 점수의 차이(최대값에서 최소값을 뺀 것) 막대 그래프로 나타낸 것이다.

<그림 1>에서 3명의 평가자 중에 2명이 부여한 평가 점수의 차이가 국민학교 생물 실험 점검표의 경우에는 평균 4.5점이었으며 중학교 물리 실험 점검표의 경우에는 4.7점, 그리고 중학교 물리 실험 보고서의 경우에는 4.4점이었다. 그러나 평가자간에 평가 점수 차이가 평균의 대략 2배가 되는 9점 이상으로 벌어지는 경우도 국민학교 생물 점검표에서 한 명, 중학교 물리 점검표와 보고서에서 각각 세 명이나 있는 것으로 나타났다.

국민학교 생물 실험 점검표의 점수 배점이 20점, 중학교 물리 실험 점검표와 보고서는 각각 점수 배점이 25점인 것을 감안하면, 이러한 평가자간의 평가 점수 차이는 매우 크다고 할 수 있다. 이러한 점은 실험 평가를 실시하기 전과 후에 평가자들 사이에 평가 기준에 대한 의견의 조정이 거의 없었음을 의미한다고 할 수 있다.

평가자들의 평가 기준에 대한 사전 협의는 학생들의 실험 능력에 대한 객관적이고도 공정한 평가를 위해 필수적이다. 이번 실험의 경우, 점검표와 보고서의 채점 기준이 실제 학생들의 구체적인 실험 활동의 다양성에 비하여 모호하였기 때문에 평가자들 간의 사전 협의가 필수적이었다. 또한 평가 과정에서 평가자가 나름대로 설정한 구체적인 평가 기준



<그림 1> 동일 학생에 대한 평가자의 실험 평가 점수 차이

을 기록하고 학생들의 실험 능력을 평가한 후에 이를 대조함으로써 평가자들마다 다르게 인식될 가능성이 있는 평가 기준들을 일관성 있게 수정하는 작업이 이루어져야 하였다. 그러나 실제로 이러한 작업의 중요성이 평가자들 간에 인식되지 않아서 이러한 작업은 거의 이루어지지 않았다.

4) 실험 평가 항목간의 관계

실험 점검표의 경우에는 실험 시간에 세분화된 평가 항목을 가지고 각 평가자가 직접 평가를 실시하였으므로 실험에 참가한 개별 학생들의 구체적인 평가 자료가 얻어질 수 있었다. 그러나 보고서의 경우에는 세분화된 평가 항목이 제시되었음에도 불구하고 학생들이 어떠한 항목에서 어떠한 점수를 받았는지에 대한 구체적인 평가 자료가 수집되지 못하고 단지 총점만이 제시되었다.

예외적으로 화학 실험 보고서의 경우에는 분리된 평가 항목들에 따라 각 학생들의 구체적인 평가 자료를 구할 수 있었으므로, 화학 실험 보고서와 점검표의 각 평가 항목 간 관계를 알아보았다.

화학 실험의 경우에는 점검표에서 학생들의 '조작적 기능'만을 살펴보았으며, 보고서에서 '실험 설계', '변인통제', '자료수집', '자료처리', '결론'의 5가지 탐구과정의 영역들을 평가하였다. 따라서 이 6가지 평가 항목의 점수들 사이에 어떠한 관련성이 있는지 알아보기 위하여 상관계수를 구하였다.

<표 8>에서 평가 항목 중 '실험설계'의 점수는 '변인통제', '자료수집', '자료처리' 항목의 점수들과 높은 상관이 있는 것으로 나타났다. 그리고 '자료수집'과 '자료 처리' 점수들 간에도 상관이 높은 것으로 분석되었다. 그러나 올바른 '결론'을 내리는 능력과 '수공적 실험기능'의 점수는 다른 평가 항목의 점수들과 높은 상관 계수를 갖지 않았다.

<표 8> 화학 실험 평가 항목간의 상관관계

	실험 설계	변인 통제	자료 수집	자료 처리	결론	수공 기능	총점
실험설계	1.00						
변인통제	0.65*	1.00					
자료수집	0.61*	0.45	1.00				
자료처리	0.62*	0.36	0.56*	1.00			
결론	0.30	0.16	0.29	-0.02	1.00		
수공기능	0.16	0.14	0.43	0.20	-0.01	1.00	
총 점	0.82*	0.64*	0.83*	0.66*	0.34	0.61*	1.00

* p < .0001

화학 실험의 총점에 가장 높은 상관률 보이는 것은 '실험 설계'와 '자료수집' 평가 항목으로 0.8 정도의 상관계수를 나타냈으며, 그 다음에 '자료처리', '변인통제', '조작 기능'의 평가 항목으로 0.6 정도의 상관계수를 나타내었다. 그리고 '결론'이 총점과 가장 낮은 상관률을 보였다.

3. 실험대회 전체 평가 결과의 분석

과학탐구 실험대회의 전체 평가는 지필검사 객관식 문항과 주관식 문항, 그리고 두 종류 실험의 보고서와 점검표에 의해 이루어졌다. 국민학교와 중학교의 경우로 구분하여 각 평가 항목간의 상관관계를 알아본 결과를 <표 9>와 <표 10>에 제시하였다.

<표 9> 국민학교 실험대회 평가 항목간의 상관 관계

	객관식	주관식	생물 점검표	생물 보고서	지구과학 점검표	지구과학 보고서
객관식	1.00					
주관식	0.48	1.00				
생물점검	0.30	0.19	1.00			
생물보고	0.32	0.28	0.89*	1.00		
지학점검	-0.24	0.18	0.09	0.05	1.00	
지학보고	0.10	0.22	0.17	0.05	0.29	1.00
총 점	0.71*	0.71*	0.68*	0.69*	0.20	0.49

* p < .0001

<표 9>에서 객관식 문항 점수는 지구과학 실험 점검표의 점수와 음의 상관을 갖는 것으로 분석되었다. 이는 객관식 문항의 점수가 높을 수록 지구과학 실험 점검표의 점수가 낮다는 것을 의미한다. 객관식 문항 점수가 총점에 가장 큰 영향을 미치므로(상관계수 0.71) 결국 지구과학 실험 점수는 총점에 가장 적은 영향을 미치게 되었다.

총점은 지구과학 실험 점검표의 점수를 제외한 다른 모든 평가 항목의 점수들과 높은 상관관계를 나타내었다. 학생의 실험대회 전체 점수와 가장 높은 상관률 보이는 것은 지필검사 객관식과 주관식 문항 점수였으며, 그 다음으로는 생물실험 보고서와 점검표 점수로 나타났다.

각 평가 항목의 점수들 간에 상관계수가 낮은 양의 값을 가지며 총점과의 상관은 높은 경우에 고른 영역의 평가가 이루어진 것이라고 가정할 때, <표 10>에서 분석한 중학교 실험대회의 평가도 몇 가지 문제점이 나타났다. 즉, 실험대회의 물리 실험 보고서와 점검표 점수들 간에 매우 높

은 상관관계가 나타났으며, 주관식 점수는 화학 실험 보고서 점수 등과 음의 상관 관계를 보였다.

<표 10> 중학교 실험대회 평가 항목간의 상관 관계

	객관식	주관식	물리 점검표	물리 보고서	화학 점검표	화학 보고서
객관식	1.00					
주관식	0.38	1.00				
물리점검	0.20	-0.03	1.00			
물리보고	0.35	0.04	0.77*	1.00		
화학점검	0.35	-0.02	0.20	0.18	1.00	
화학보고	0.24	-0.12	0.30	0.18	0.27	1.00
총 점	0.79*	0.45	0.52*	0.55*	0.50*	0.61*

* p < .0001

그러나 총점은 지필평가와 실험 평가의 모든 항목의 점수들과 높은 양의 상관 관계를 갖는 것으로 분석되었으므로, 평가 항목 중 하나가 총점과 상관이 낮게 분석된 국민학교의 경우보다는 평가 항목의 기준 설정이 양호했다고 할 수 있다.

학생의 실험대회 전체 점수와 가장 높은 상관관을 보이는 것은 지필검사 객관식 문항의 점수였으며, 가장 낮은 상관관계를 보이는 것은 주관식 문항 점수였다.

V. 결론 및 제언

실험대회의 분석 결과를 토대로 다음과 같은 결론 및 제언을 하고자 한다.

첫째, 이번 실험대회의 실험 평가에서는 탐구과정에 비하여 학생의 수공적 실험기능에 대한 평가가 소홀히 취급되었다. 실제로 수공적 실험기능의 평가에 중점을 둔 화학 실험의 경우에는 저울 사용법, 부피 측정도구 사용법, 용매에 용질을 녹이는 과정 등 가장 기본적인 실험기능도 제대로 몸에 익히지 못한 학생들이 많이 관찰되었다. 따라서 앞으로는 평가에서 학생의 탐구과정에 관한 능력만을 강조하지 않고, 수공적 실험기능에 관한 평가도 강조되어야 할 것이다.

뿐만 아니라 Woolnough(1989), Bates(1985)등 많은 과학교육자들이 주장한 바와 같이 경향, 자신감, 직관력과 같은 정의적 능력과 형식적, 비형식적 학습을 통해 형성되는 개인적 통찰력, 그리고 과학의 성질에 대한 이해 등에 관한 평가 도구의 개발에도 노력을 기울여야 할 것이다.

둘째, 이번 실험대회에서는 실험 보고서의 평가 항목과

점검표의 평가 항목의 내용이 중복되는 경우가 있었다. 이러한 경우에는 실험 보고서와 점검표의 점수들 간에 상관관계가 매우 높게 분석되었다. 또한 보고서로 평가가 가능하여 실험 현장에서는 직접 평가할 필요가 없는 평가 항목이 실험 점검표에 제시되기도 하였다. 앞으로는 실험 현장에서 직접 평가가 이루어져야 하는 학생 활동들만을 실험 점검표로 평가하고, 기록하여 보존이 가능한 내용은 보고서에 기록하도록 하여 평가의 영역을 명확히 구분함으로써 평가의 효율성을 극대화시켜야 할 것이다.

셋째, 실험 점검표 및 보고서의 평가 항목에서 출제자가 제시한 평가 척도에 해당하지 않는 학생의 활동이 많이 관찰되었다. 실제로 출제자가 아무리 세밀하게 평가 척도를 설정한다 하더라도 현장에서 이루어지는 학생들의 실험 활동을 모두 포함하기는 어려울 것이다. 이러한 경우 평가자의 주관에 개입될 가능성이 있다. 실제로 이번 대회에서 세분화된 평가 척도가 제시되었음에도 불구하고 동일한 실험 내에서 동일한 학생에게 부여한 평가 점수는 평가자에 따라 큰 차이를 보였다.

이러한 문제점을 극복하기 위하여 평가자들은 평가를 실시하기 전에 평가 척도에 대한 구체적인 사전 협의를 거쳐야 할 것이다. 또한 실제 평가를 하면서 학생의 실험 행동 특징을 간략히 묘사함으로써 실험평가가 모두 끝난 후에 다른 평가자들과 구체적 평가 기준의 대조 작업을 벌여 평가의 객관성과 공정성을 유지하는 노력을 기울여야 할 것이다.

넷째, 실험 평가에서 점검표는 평가자가 실험에 직접 참여하여 평가 항목에 점수를 기재하였으므로 각 평가 항목별로 학생들의 실험 능력에 대한 구체적인 평가 자료를 얻을 수 있었으나, 보고서의 경우에는 실험대회가 끝난 후에 세 명의 평가자들이 평가한 방법이 네 종류의 실험마다 모두 달랐다. 또한 출제자가 구체적인 평가 척도를 제시하였음에도 불구하고, 점검표와 같은 구체적인 평가 자료를 얻을 수 있었던 것은 화학 실험뿐이었다.

그 이유 중의 하나로 출제자의 평가 척도를 근거로 각 평가 항목의 점수를 합산함으로써 보고서의 점수를 구하는 대신, 평가 항목들을 참고로 하여 평가자의 직관으로 보고서의 최종 점수를 결정한 경우를 들 수 있다. 따라서 화학 실험을 제외하고는 평가 항목과 총점과의 관계를 분석하여 학생들의 실험 점수에 영향을 미친 요인을 찾아내기가 어려웠다.

앞으로는 실험 점검표와 유사하게 평가자들이 실험 보고서의 평가 결과도 각 평가 항목별로 구분하여 기록할 수 있도록 보고서 평가 점수표를 작성할 필요가 있다.

실험대회가 원만히 이루어지기 위해 대회 운영에 고려할 필요가 있다고 여겨지는 사항들을 제안하고자 한다.

첫째, 사전 모의 실험의 실시와 실험 참가 조교의 훈련이
다. 실험대회를 시작하기 전에 출제자의 의도와 다른 해석
이 가능한지, 출제자가 고려하지 못한 상황이 일어날 가능
성이 있는지 등에 관해 미리 점검해 보기 위해서 사전 모의
실험의 실시는 반드시 필요하다고 본다. 또한 모의 실험을
통해 학생들이 자신의 실험 설계에 따라 자유롭게 실험 기
구를 선택하여 사용할 수 있도록 충분한 양의 실험 기구를
준비하는 일도 가능할 것이다.

실험에 참가하는 조교들을 사전 모의 실험으로 훈련시키
고, 실험의 내용 및 결과 등을 미리 개략적으로 파악하게 하
는 일도 매우 중요하다. 실험이 진행되면서 예상하지 못했
던 문제가 발생하였을 때에는 실험 조교의 적절한 조치가
필요하기 때문이다.

둘째 제안은 평가 문항을 개발하는 출제자들과 선정자들
이 고려하기를 바라는 사항이다.

실험 평가의 점검표의 평가는 매우 짧은 시간 내에 이루
어지기 때문에 평가자가 그 순간을 놓치면 평가가 불가능한
학생 활동에 관한 평가 항목은 가능한 한 배제하여야 한다.
실험대회의 운영상 한 명의 평가자가 여러 명의 학생들을
동시에 관찰하고 평가하므로 학생의 특정한 실험 활동을 관
찰하는 기회를 잃어버릴 수도 있기 때문이다. 이러한 경우
에 평가는 평가자의 임의적인 주관에 개입될 수 있기 때문
에 평가의 공정성에 문제가 생긴다. 실제로 이번 실험대회
에서 평가 척도가 세분화되어 명시되었음에도 불구하고 동
일 학생을 평가한 평가자들의 점수 차이는 매우 컸는데 여
기에는 이러한 요인도 작용하였다고 보여진다.

여러 학생들이 동일한 실험실 내에서 실험을 수행하므로
다른 학생들의 실험 과정과 기능을 모방함으로써 좋은 평가
점수를 받게 될 소지가 있는 평가 항목도 가능한 한 제외시
켜야 한다. 또한 학생들이 구체적인 실험 활동 없이 자신의
생각으로만 보고서를 작성하거나 문제를 해결할 수 있는 실
험도 역시 제외시켜야 한다. 이러한 평가 항목은 학생들의
실험 능력을 정확히 파악하는데 방해가 될 뿐이기 때문이
다.

참고문헌

김창식 외 4인, 과학학습 평가, 교육과학사, 1993.
박성혜, 박승재, 학생 과학탐구 실험대회의 비판적 분석, 과
학교육연구논총 19(1), 15-37, 1994.
하병권 외 4인, 과학탐구 실험대회 보고서, 전국 교육대학교
과학교육연구소 협의회, 1994.

황정규, 학교 학습과 교육평가, 교육과학사, 1985.
Bates, G.C.(1985). *The Role of the Laboratory in
Secondary School Science Programs*, Columbia
University, New York, N.Y. 10027.
Doran, R. L.(1980). *Basic Measurement and Evaluation of
Science Instruction*, NSTA, Washington, DC 20009.
Doran, R.L. & Boorman, J.(1992). Alfred Chan and Nicholas
Hejaily, *Successful Laboratory Assessment, The
Science Teacher*, pp. 22 - 27, April.
Fairbrother, B.(1989). Problems in the assessment of
scientific skills, edited by Wellington, J. *Skills and
Processes in Science Education - A Critical
Analysis*, Routledge, London & New York.
Kempa, R.(1983). *Assessment in Science*, Cambridge
University Press, London.
Leonard, W.H.(1983). An Experimental Study of a BSCS-
Style Laboratory Approach for University General
Biology, *JRST*, 20(9), pp.807-813.
Leonard, W.H.(1988). An Experimental Test of an
Extended Discretion Laboratory Approach for
University General Biology, *JRST*, 26(1), pp.79-91.
Shymansky, J.A., Kyle, W.C. & Alport, J.M.(1983). The
effects of new science curricula on student
performance, *JRST*, 20(5), pp.387-404.
Swain, J.L.(1988). Assessing Chemical Process Skills,
Education in Chemistry, September, pp. 142-144.
Tobin, K.(1990). Research on Science Laboratory
Activities, *School Science and Mathematics*, 90(5),
pp. 403-418.
Woolnough, B.E.(1989). Towards a holistic view of
processes in science education, edited by
Wellington, J. *Skills and Processes in Science
Education - A Critical Analysis*, Routledge, London
& New York.

ABSTRACT

Critical Analyses of '2nd Science Inquiry Experiment Contest'

Seoung Hey Paik

(Science Education Department of Seoul National University)

The purpose of this study was to analyse the problems of 'Science Inquiry Experiment Contest(SIEC)' which was one of 8 programs of 'The 2nd Student Science Inquiry Olympic Meet(SSIOM)'. The results and conclusions of this study were as follows:

1. It needs to reconsider the role of practical work within science experiment because practical work skills form one of the mainstays in current science. But the assessment of students' laboratory skills in the contest was made little account of. It is necessary to remind of what it means to be 'good at science'. There are two aspects: knowing and doing. Both are important and, in certain respects, quite distinct. Doing science is more of a craft activity, relying more on craft skill and tacit knowledge than on the conscious application of explicit knowledge. Doing science is also divided into two aspects, 'process' and 'skill' by many science educators.

2. The report's and checklist's assessment items were overlapped. Therefore it was suggested that the checklist assessment items were set limit to the students' acts which can't be found in reports. It is important to identify those activities which produce a permanent assessable product, and those which do not. Skills connected with recording and reporting are likely to produce permanent evidence which can be evaluated after the experiment. Those connected with manipulative skills involving processes are more ephemeral and need to be assessed as they occur. The division of student's experimental skills will contribute to the accurate assess of student's scientific inquiry experimental ability.

3. There was a wide difference among the scores of one participant recorded by three evaluators. This means that there was no concrete discussion among the evaluators before the contest. Despite the items of the checklists were set by preparers of the contest experiments, the concrete discussions before the contest were necessary because students' experimental acts were very diverse. There is a variety of scientific skills. So it is necessary to assess the performance of individual students in a range of skills. But the most of the difficulties in the assessment of skills arise from the interaction between measurement and the use. To overcome the difficulties, not only must the mark needed for each skill be recorded, something which all examination groups obviously need, but also a description of the work that the student did when the skill was assessed must also be given, and not all groups need this. Fuller details must also be available for the purposes of moderation. This is a requirement for all students that there must be provision for samples of any end-product or other tangible form of evidence of candidates' work to be submitted for inspection. This is rather important if one is to be as fair as possible to students because, not only can this work be made available to moderators if necessary, but also it can be used to help in arriving at common standards among several evaluators, and in ensuring consistent standards from one evaluator over the assessment period. This need arises because there are problems associated with assessing different students on the same skill in different activities.

4. Most of the students' reports were assessed intuitively by the evaluators despite the assessment items were established concretely by preparers of the experiment. This result means that the evaluators were new to grasp the essence of the established assessment items of the experiment report and that the students' assessment scores were short of objectivity.

Lastly, there are suggestions from the results and the conclusions. The students' experimental acts which were difficult to observe because they occur in a flash and which can be easily imitated should be excluded from the assessment items. Evaluators are likely to miss the time to observe the acts, and the students who are assessed later have more opportunity to practise the skill which is being assessed.

It is necessary to be aware of these problems and try to reduce their influence or remove them. The skills and processes analysis has made a very useful checklist for scientific inquiry experiment assessment. But in itself it is of little value. It must be seen alongside the other vital attributes needed in the making of a good scientist, the affective aspects of commitment and confidence, the personal insights which come both through formal and informal learning, and the tacit knowledge that comes through experience, both structured and acquired in play. These four aspects must be continually interacting, in a flexible and individualistic way, throughout the scientific education of students. An increasing ability to be good at science, to be good at doing investigational practical work, will be gained through continually, successively, but often unpredictably, developing more experience, developing more insights, developing more skills, and producing more confidence and commitment.