

탐구적 일반화학실험 수행 평가 준거 개발

강순희 · 김양현 · 박종윤

(이화여자대학교)

The Development of Performance Scoring Rubrics for the Inquiry-Based General Chemistry Experiments

Kang, Soonhee · Kim, Yang-Hyun · Park, Jong-Yoon

(Ewha Womans University)

ABSTRACT

This study is to develop the performance scoring rubrics for the inquiry-based experiments of general chemistry course in the college of education. Two types of analytic scoring rubrics have been developed for nine different experiments. The first one is to assess scientific process skills from the written experimental reports. These analytic scoring rubrics include seven process skills selected from the Lawson's 'creative and critical thinking skills' and other known process skills. The second one is to assess the individual manipulative skills and experimental attitudes through direct observations by the teacher. The content validity of all scoring rubrics was testified by six science educators. Also the inter-scorer reliability of analytic scoring rubrics administered on the students' experimental reports was examined. The correlation coefficient between the scores obtained from the experiments and those of the written test for theoretical knowledges was found to be $r=.663(p < .01)$. From the variance($r^2=.440$), we would say indirectly that the 56% of this experimental assessment does not overlap with the theoretical knowledges test and assesses students' science process skills, manipulative skills, and attitudes.

Key Words : performance assessment, inquiry skill, chemistry experiment assessment, manipulative skills assessment, experimental attitude assessment, analytic scoring of process skill, analytic scoring of manipulative skill, analytic scoring of experimental attitude, inquiry-based experiments of general chemistry

I. 연구의 필요성 및 목적

최근 우리 나라는 학교 교육의 정상화를 위한 교육 개혁의 일환으로 '새 교육 문화 창조'를 교육 현장에 확산시키고 있으며 이를 위한 방안으로 초등학교 및 중등학교에 '새 교육, 새 물결 운동'을 적극적으로

추진하고 있다. '새 교육, 새 물결 운동'에는 여러 가지 과제들이 있는데, 그 중 한 가지가 결과 중심 평가에서 과정을 중시하는 평가로 일부 전환하는 것이다. 이에 따라 이미 초·중학교 현장에서는 수행 평가를 도입하여 실시하고 있으며, 1999학년도부터는 고등학교에서도 1학년부부터 지필 고사에서 배점의

*1999년 5월 18일 받음.

30% 이상을 서술형, 논술형으로 실시하도록 하고 있다(교육부, 1999). 이는 곧 지금까지 학교 교육에서 사용되어 왔던 전통적인 선택형 평가 방법에 대한 새 물결인 것이다.

이러한 학교 교육에서의 평가 방법에 대한 변화의 조짐은 1980년대 중반부터 선진 외국 특히, 미국(서혜애와 신성현, 1998, 1999a, 1999b; MSEBNRC, 1997; NCTM, 1989; NSES, 1996; O'Malley & Pierce, 1992, 1996; O'Neil, 1992; WPADP, 1996)을 위시하여 국내·외에서 대두되어 왔다.

일리노이 대학의 Harnish(1999)는 평가를 학생들이 알고 있는 것(know), 할 수 있는 것(can do), 그리고 성취한 것(accomplish)이 무엇인지에 대하여 증거를 수집하는 것(getting evidence)이라고 정의하고 평가의 기준으로 다음의 항목을 제시하고 있다. 첫째, 평가는 목적이 분명해야 한다. 둘째, 평가는 타당해야 하고, 학생, 학부모, 교사 모두가 신뢰할 수 있어야 한다. 셋째, 평가의 결과는 학생과 교사 모두에게 유용하게 활용되어야 한다. 마지막으로 효과적인 평가가 되기 위해서는 당면하는 많은 제약 조건들이 고려되어야 하며 여러 종류의 평가 방법들이 균형 있게 사용되어야 한다.

이러한 관점에서 생각해 볼 때 '새 교육, 새 물결 운동'에서 평가 방안 개선의 기본 취지는 여러 유형의 평가 방법들을 사용하고자 하는 것이며 지금까지의 전통적 평가 방법을 완전히 배제해야 한다는 것은 아니다. 즉, 지금까지 우리의 교육 현장에서는 전통적인 선택형 평가 방법에만 지나치게 의존해 왔으므로 이제는 그 대안적 평가 방법으로 수행 평가(performance assessment)의 사용을 병행할 것을 권장하게 된 것이다. 수행 평가란 특정한 과제에 대한 학생들의 수행 과정, 즉 학생들이 지식과 기능, 기술을 실제 생활이나 인위적인 평가 상황에서 얼마나 잘 수행하는지(doing, performing), 어떻게 수행하는지(how to do, how to performance)를 관찰이나 면접, 증거집 등을 통하여 종합적으로 판단하는 평가이다(최연희, 권오남, 성태제, 1998). 수행 평가는 지식이나 기능에 의한 정답 여부나 산출물만 평가하는 전통적인 선택형 평가와 구별된다는 점에서 대

안적 평가로, 실제 생활과 같은 자연스러운 상황에서 평가한다는 의미에서 참 평가로, 학생들의 능력과 기술에 대한 증거를 직접 관찰하여 평가한다는 점에서 직접 평가로, 학생들의 증거집을 통하여 평가한다는 점에서 포트폴리오 평가 등으로 그 개념이 확산되어 사용되고 있다.

특히 과학과에서는 다양한 방법으로 수행 평가를 적용할 수 있다. 학생들이 수행하는 실험 과정에서 탐구 기능들을 직접 관찰하면서 평가할 수도 있으며, 실험 후 제출한 실험 보고서를 통하여 학생들의 탐구 기능과 과학적 사고력을 평가할 수도 있다. 이와 더불어 실험이 아닌 교실 수업에서도 주어진 과제를 수행하는 동안 학생들이 사용하는 문제 해결력과 의사소통 기술 등에 대한 평가도 가능하고, 현장 답사, 수업 참관, 견학, 조사, 특별 활동 또는 방학 과제로서의 프로젝트나 탐구 과제 등의 수행에 대한 평가에도 사용할 수 있다.

학생들의 과학과 학업 성취도 평가를 위해서는 여러 가지 유형의 평가 방법들이 균형 있게 사용되는 것이 바람직하다. 전통적인 평가 방법으로 과학 지식과 이해를 평가하고, 실험 수행 평가를 통하여 과학적 사고력과 탐구 기능을, 실제 생활과 관련된 문제를 해결하도록 하고 과학의 가치를 인식하게 할 수 있는 수행 과제를 통하여 문제 해결력과 의사소통 기술 등을 평가한다면 학생들의 학업 성취도에 대한 지금까지의 전통적인 선택형 위주의 평가보다는 훨씬 더 타당하고 합리적인 평가가 될 것이다. 이러한 과학과 학업 성취도에 대한 여러 가지 유형의 평가 방법의 적용은 중등학교 뿐만 아니라 대학교 과정에서도 필요하며, 특히 과학 교사를 양성하는 사범대학 과학교육과에서는 더욱 필요한 것으로 생각된다. 왜냐하면 장차 교사가 될 학생들에게 과정 중심의 수행 평가를 경험하게 해 보으로써 학생들은 교사가 되었을 때 중등학교 현장에서 수행 평가를 보다 자신 있게 적용할 수 있기 때문이다.

그러나 현재까지는 수행 평가에 대한 구체적인 지침이나 절차를 제시해주지 않고 그냥 수행 평가를 실시하라고 했기 때문에 실제 학교 현장에서는 많은 혼란이 야기되고 있는 것이 현실이다. 그러므로 본 연

구에서는 과학 탐구 실험에 적용할 수 있는 수행 평가의 예를 제시하기 위하여 대학교 일반화학실험에 대한 수행 평가 적용의 지침과 준거를 개발하고자 하였다.

과학 실험에 대한 수행 평가의 적용을 위해서는 먼저 지금까지 과학과의 실험 활동에 대하여 실제로 어떻게 평가해 왔는가를 몇 가지 측면으로 재고해 볼 필요가 있다. 첫째, 일반적으로 몇 명이 한 조가 되어 조별로 실험 활동을 하게 하지만 실험 보고서는 개인별로 작성하여 제출하게 하고 있다. 다시 말하면, 조별로 협동학습을 하였음에도 불구하고 평가는 개인별로 하고 있다는 것이다. 그러므로 학생들의 조별 실험 활동이 활성화되기는 어렵다고 볼 수 있다. 둘째, 학생들이 실험 후에 제출하는 실험 보고서에 대한 평가의 기준이다. 일반적으로 실험 보고서 평가는 실험 결과가 얼마나 이론값에 근접하였는가, 또는 실험 보고서를 얼마나 성의 있게 작성했는가에 중점을 두고 있다. 즉 실험 과정에서 나타날 수 있는 학생들의 수행 과정, 예를 들면 탐구 기능이나 사고력 등을 평가하는 과정 중심의 평가이기보다는 결과 중심의 평가를 하고 있다.

따라서 본 연구에서 실시한 일반화학실험은 보다 과정 중심의 수행 평가를 하기 위하여 다음과 같이 진행하였다. 첫째, 세 명이 한 조가 되어 협동학습으로 실험을 진행하였다. 둘째, 각 실험마다 조원들이 보고서 작성자, 토론 진행자, 실험 결과 기록자 역할을 순번제로 담당하게 하였다. 셋째, 실험 후에는 개인별이 아닌 조별 보고서를 제출하게 하였다. 즉, 한 조가 협동하여 실험하였으므로 모든 조원들에게 같은 점수를 부여하는 것이다. 넷째, 학생들의 실험 과정에서 나타나는 수공적 기능이나 실험에 임하는 태도는 교수와 조교가 직접 관찰하여 점검하는 형태로 개인별 평가를 하였다. 다섯째, 학기말 실험 수행 과제 평가도 하나의 탐구 실험 수행 과제를 주어 조별로 실험하면서 해결하도록 하였다. 여섯째, 학생들에게 실험 보고서는 실험 과정에서 나타나는 탐구 기능을 요소별로 평가한다고 알려 주었다.

위와 같이 한 학기동안 진행한 일반화학실험에 대하여 학생들의 일반화학실험 수업에 대한 성취도를

평가하기 위한 채점 준거를 개발하는 것이 본 연구의 목적이다. 이를 위한 구체적인 연구 내용은 다음과 같다.

첫째, 각 탐구기능 요소에 대한 일반적 채점 준거를 개발한다.

둘째, 각 실험에 대하여 학생들이 실험 후 제출하는 조별 실험 보고서에 포함되어 있는 탐구 기능 요소를 평가하기 위한 분석적 채점 준거를 개발한다.

셋째, 각 실험에 대하여 실험이 진행되는 동안 교수가 직접적인 관찰에 의해 학생들의 개인별 수공적 기능 및 실험 태도를 평가하기 위한 점검 형태의 분석적 채점 준거를 개발한다.

넷째, 한 학기 동안 학생들의 일반화학실험 수업의 학업 성취도와 같은 기간 동안 학생들의 일반화학 이론 수업의 학업 성취도 사이의 상관 관계를 알아본다.

II. 연구 방법

사전 연구로서 본 연구실에서 개발한 탐구 과정 중심의 일반화학실험(강순희, 박중윤, 최병순, 1997)을 지난 3년 동안 현장에 투입하여 얻어진 학생들의 실험 보고서들을 토대로 하여 일차적으로 보고서 평가를 위한 일반적 채점 준거와 분석적 채점 준거들을 개발하였다. 연구자들은 이 사전 연구에서 개발한 채점 준거를 가지고 사범대학 과학교육과 1학년 2개반 60명의 학생들이 수강하는 일반화학실험 평가에 직접 사용하면서 수정·보완하여 최종적인 채점 준거를 만들었다.

학생들은 세 명이 한 조가 되어 전체로는 2개반 20조가 협동학습으로 실험을 하였으며, 실험 후에는 조별 실험 보고서를 제출하게 하였다. 학생들의 수공적 기능과 실험에 임하는 태도는 직접 관찰에 의한 점검 형태로 개인별 평가를 하였다. 모든 채점준거는 과학교육 전문가 세 명과 과학교육 전공 대학원생 세 명으로부터 타당도를 검증받았다. 개발한 실험 보고서에 대한 분석적 채점 준거를 가지고 학생들의 보고서를 세 명의 대학원생들이 채점하도록 하여 얻어진 결과로부터 채점자간 신뢰도를 검증하였다.

III. 연구 결과 및 논의

본 연구에 사용된 한 학기용 일반화학실험 주제는 다음의 10개이며, 연구 결과를 크게 일반적 채점 준거 개발과 각 실험에 대한 분석적 채점 준거 개발로 나누어 논의하도록 하겠다.

- 실험 1. 기본적인 실험조작 익히기
- 실험 2. 액체나 고체의 밀도는 어떻게 측정하나?
- 실험 3. 물질의 특성을 이용한 혼합물의 분리
- 실험 4. 여러 가지 크로마토그래피 방법에 의한 염록소, 검정 잉크, 식용 색소의 색소 분리
- 실험 5. 알코올의 농도는 어떻게 측정할까?
- 실험 6. 일상 생활에서의 산성 혹은 염기성 물질의 pH 측정 및 산-염기 적정법
- 실험 7. 염화은 침전 반응을 통한 알짜 이온 반응식의 정량적인 이해
- 실험 8. 화학 반응에서 열의 출입을 측정해 보자
- 실험 9. 동전에 도금을 해보자
- 실험 10. 물유리 수용액에서 아름다운 화학 정원을 만들어보자.

1. 일반적 채점 준거 개발

위 실험들을 수행하면서 나타나는 학생들의 탐구 능력을 평가하기 위한 탐구 기능 요소는 Lawson과 Birk(1994)가 제안한 '창조적이며 비판적인 탐구 사고력 기능'을 토대로 하여 우리나라의 실험 현실(남정희, 1996)에 맞게 수정하여 일곱 개(I-VII)의 범주로 정하였다. 확정된 일곱 개 범주의 탐구 기능 요소에 대한 일반적 채점 준거는 다음과 같다. 괄호 속에는 그 탐구 기능 요소를 포함하는 실험 번호를 나타내었다.

- I. 자연 현상을 정확하게 서술하는 기능 (실험 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
 - 1. 관찰 가능한 특징으로 사물을 서술, 배열, 분류하는가?
 - 2. 양, 길이, 면적, 무게, 부피 및 밀도와 같은

- 변수로 사물을 서술, 배열, 분류하는가?
- 3. 사물 그룹들의 다양하며, 일정한 특징들을 확인하는가?
- 4. 어떤 표본과 모집단 간의 차이점을 인식하고 무선 표집 방법을 확인하는가?

II. 논리적으로 예측하는 기능 (실험 2, 3, 10)

- 1. 정확한 관찰을 근거로 하는가?
- 2. 변인간의 관계를 제시하는가?
- 3. 논리적인가?
- 4. 검증 가능한가?

III. 실험을 계획하고 수행하는 기능 (실험 2, 3, 4, 5, 6, 7, 8, 9, 10)

- 1. 정확한 관찰과 측정을 위한 적절한 기구를 선택하는가?
- 2. 적절한 대조구를 포함시키는가?
- 3. 변인을 인식하고 통제하는가?
- 4. 예측을 검증할 수 있도록 실험을 계획하고 수행하는가?
- 5. 실험 절차는 적절한 순서로 배열되었는가?

IV. 관련된 실험 자료를 수집하는 기능 (실험 2, 3, 4, 5, 6, 7, 8, 9)

- 1. 측정이 정확한가?
- 2. 알맞은 단위나 척도를 사용하는가?
- 3. 예측한 것을 검증하는데 필요한 자료를 수집하는가?
- 4. 필요한 자료를 충분히, 정확하게 수집하여 기록하는가?

V. 자료를 변환하는 기능 (실험 2, 4, 6, 7, 8, 9)

- 1. 자료를 적절한 형태로 재조직하고 변환하는가?
- 2. 수집한 자료 내에서 경향성이나 공통 요소들을 확인할 수 있는가?
- 3. 변인들 사이의 관계를 인식하는가?

VI. 합리적인 결론을 도출해 내고 적용하는 기능

(실험 2, 4, 5, 6)

1. 관련 자료를 인식하고 이를 토대로 결론을 내리는가?
2. 예상 결과와 실제 결과를 비교해 합리적인 결론을 도출하는가?

Ⅶ. 예측과 결론을 비교·평가하는 기능 (실험 2, 3, 5, 6, 7, 8, 9, 10)

1. 예측과 결론이 맞지 않을 경우 문제점의 원인을 밝히기 위해 실험 설계, 자료 수집, 결과 처리 등을 검토하는가?
2. 구체적인 문제점을 밝히는가?
3. 해결 방안 및 개선점도 제시하는가?

2. 각 실험에 대한 분석적 채점 준거 개발

실험에 대한 분석적 채점 준거(analytic scoring rubric)는 크게 두 가지로 나누어 논의하겠다.

첫째는 학생들이 실험 후에 제출하는 조별 실험 보고서 채점 준거이다. 둘째는 학생들이 실험하는 과정에서 나타나는 수공적 기능과 실험에 임하는 태도를 교수가 직접 관찰하면서 개인별로 평가하는 점검 형태의 채점 준거이다.

1) 실험 보고서 채점 준거

앞서 개발한 일반적 채점 준거(I-VII)에 준하여 학생들이 실험하면서 작성한 실험 보고서를 평가할 수 있는 분석적 채점 준거를 각 실험별로 개발하였다. 본 연구에서는 조별(한 반 30명의 경우 10개조) 실험 활동을 하게 한 후에 보고서를 조별로 채점하였으나, 사정에 따라 개인별로 보고서를 채점하는데 사용할 수도 있다. 본 연구에서 조별 보고서 평가를 하는 이유는 학생들의 조별 협동학습에 의한 그룹의 협동성을 장려하기 위한 목적도 있으나, 또한 평가하는 교사의 일의 양을 줄일 수 있기 때문이기도 하다. 이러한 실험 보고서 평가는 학생들의 과학 지식을 평가하는 결과 중심 평가와 대응되는 평가로서, 학생들의 실험 수행 과정 즉 탐구 과정을 평가하는 수행 평가이다. 연구자들은 10개 실험에 대한 보고서 채점 준

거를 모두 개발하였으나(이화여대 과학교육과 화학교육 연구실 소장), 여기서는 그 중 '실험 5'에 대한 분석적 채점 준거만 제시하고자 한다.

“실험 5. 알코올의 농도는 어떻게 측정할까?”에는 탐구 기능 요소 Ⅲ, Ⅳ, Ⅵ, Ⅶ이 포함되어 있다. 각 탐구 기능 요소의 배점은 각각 5점, 3점, 3점, 3점이며, 그에 대한 분석적 채점 준거와 점수는 다음과 같다. 채점 준거에 해당하는 점수의 비중은 연구자들의 관점에 의하여 1점부터 3점까지 다음과 같이 다양하게 배당하였다.

Ⅲ. 실험을 계획하고 수행하는 기능(총 5점)

1. 제시된 에탄올을 희석시켜 다양한 %의 에탄올 용액을 제조한다(2점).
2. 다양한 %의 에탄올 용액들을 주류와 동시에 $\text{Na}_2\text{Cr}_2\text{O}_7 + \text{H}_2\text{SO}_4$ 용액과 반응시킨다(2점).
3. 각 용액의 색깔 변화를 10분 이하의 간격으로 비교·관찰한다(1점).

Ⅳ. 관련된 실험 자료를 수집하는 기능(총 3점)

1. 다양한 %의 에탄올의 색깔 변화를 기록한다(1점).
2. 주류의 색깔 변화를 기록한다(1점).
3. 다양한 %의 에탄올과 주류의 색깔 변화를 각기 10분 이하의 간격으로 기록한다(1점).

Ⅵ. 합리적인 결론을 도출해 내고 적용하는 기능(총 3점)

1. 주류의 색깔 변화와 다양한 %의 에탄올의 색깔 변화 결과를 비교한다(1.5점).
2. 색깔 변화를 관찰한 결과로부터 주류의 알코올 함량 범위를 제시한다(1.5점).

Ⅶ. 예측과 결론을 비교·평가하는 기능(총 3점)

1. 주류의 실제 함량과 결론이 맞지 않을 경우 문제점의 원인을 밝히기 위해 실험 설계, 실험 수행, 자료 수집, 결론 도출 단계 등을 검토한다(1점).

Table 1. Inter-Scorer Reliability of Total Score for Experiment 5

Subject	Inter-Scorer		
	A ↔ B	A ↔ C	B ↔ C
Ⅲ, Ⅳ, Ⅵ, Ⅶ	.889	.788	.838

Table 2. Inter-Scorer Reliability of Each Inquiry Skill for Experiment 5

Subject	Inter-Scorer		
	A ↔ B	A ↔ C	B ↔ C
Ⅲ. Skills in planning and conducting experiments	.937	.764	.858
Ⅳ. Skills in collecting and applying experimental data	.824	.977	.874
Ⅵ. Skills in drawing and applying reasonable conclusions	.941	.902	.959
Ⅶ. Skills in comparing and evaluating between prediction and conclusion	.730	.750	.730

2. 구체적인 문제점을 밝힌다(1점).

3. 해결 방안도 제시한다(1점).

*실제 함량과 결론이 일치하는 경우는 3점 부여.

'실험 5'의 실험 보고서에 나타난 탐구 기능에 대한 분석적 채점 준거에 준하여 채점한 채점자(A, B, C)간 신뢰도 결과는 Table 1, Table 2와 같다.

Table 1은 '실험 5'에 대한 보고서를 분석적 채점 준거에 의하여 채점한 '실험 5'의 보고서 총점(14점)에 대한 채점자간 신뢰도이다. 채점자간 신뢰도 중 가장 높은 것은 A 채점자와 B 채점자 사이의 신뢰도 .889이고, 가장 낮은 것은 A 채점자와 C 채점자 사이의 신뢰도로 .788이다. 총점에 의한 채점자 A, B, C간의 신뢰도가 모두 .788 이상으로 채점자간 신뢰도는 높은 것으로 나타났다.

Table 2에는 '실험 5'에 포함되어 있는 네 가지 탐구 기능 Ⅲ, Ⅳ, Ⅵ, Ⅶ에 대한 각각의 채점자간 신뢰도를 나타내었다. 네 가지 탐구 과정 기능 범주에 대하여 채점자간 신뢰도는 각각 .764, .824, .902, .730 이상으로 나타났다. 합리적인 결론을 도출해 내고 적용하는 기능(Ⅵ)에 대한 채점자간(B↔C) 신뢰도는 .959로 가장 높고, 예측과 결론을 비교·평가하는 기능(Ⅶ)에 대한 채점자간(A↔B, B↔C) 신뢰도가 .730으로 가장 낮다. 전체적으로 채점자간 신뢰도

는 모두 .730 이상으로 전반적으로 높은 신뢰도를 보여 주고 있다.

Table 1과 같은 방법으로 9개 실험 보고서에 대하여 각 실험의 총점에 대한 채점자간 신뢰도를 구한 결과는 Table 3에 나타내었다. 이 때 실험 1은 기본적인 실험 조작 익히기로서 안내된 실험 절차에 따라 학생들이 실험 전반에 필요로 하는 기기 또는 실험 수행 기능들을 연습하게 하는 것이므로 제외하였다.

실험 2, 3, 4, 5, 6, 7, 8, 9, 10에 대하여 신뢰도가 각각 .680, .771, .758, .788, .976, .639, .600, .730, .737 이상이므로 채점자간 신뢰도가 전반적으로 높다고 할 수 있다.

2) 수공적 기능과 태도에 대한 채점 준거

실험을 수행하는 과정에서 학생 개개인의 수공적 기능을 평가하는 채점 준거는 다음과 같다. 이 채점 준거는 교수와 조교가 직접 관찰에 의하여 점검하는 형태로 평가하도록 개발되었다. 수공적 기능 평가는 사용자의 기준에 따라 매 실험 시간에 개인별로 점검할 수도 있고, 모든 실험이 끝난 학기말에 대표적인 몇 가지 수공적 기능을 선별하여 점검할 수도 있다. 예를 들어, 이와 같은 수공적 기능의 평가는 개인별 평가이므로 학기말 또는 적당한 시기에 선별하여 학생 개개인의 평가 도구로 사용하면 된다. 본 연구에

Table 3. Inter-Scorer Reliability of Total Score for Nine Experiments

Subject	Inter-Scorer		
	A ↔ B	A ↔ C	B ↔ C
Experiment 2	.735	.680	.837
Experiment 3	.780	.771	.830
Experiment 4	.907	.758	.895
Experiment 5	.889	.788	.838
Experiment 6	.976	1.000	.976
Experiment 7	.639	.708	.995
Experiment 8	.600	.623	.898
Experiment 9	.730	.906	.806
Experiment 10	.737	.821	.850

서는 모든 실험에 대하여 평가가 가능하다고 생각되는 수공적 기능을 개인별 또는 조별로 사용할 수 있도록 모두 개발하였다(이화여대, 과학교육과 화학교육 연구실 소장). 그 중 하나인 "실험 4. 여러 가지 크로마토그래피 방법에 의한 염록소, 검정잉크, 식용색소의 분리"의 수공적 기능에 대한 채점 준거는 다음과 같다.

1. 시료를 찍은 점(spot)이 용매에 잠기지 않도록 한다(흡족 1점, 미흡 0점).
2. 용매선을 미리 표시하여 R_f값을 계산할 수 있도록 한다(흡족 1점, 미흡 0점).
3. spotting시 반점의 크기는 너무 크지 않게 여러 번 겹쳐 찍는다(흡족 1점, 미흡 0점).

실험 과정에서 학생들이 실험에 임하는 태도를 평가하는 채점 준거는 학생들이 실험 과정에서 토론에 참여하는 정도, 조원들이 상호 협동하는 정도, 주의 사항을 지키는 정도, 실험 중의 정리·정돈하는 태도 등으로 분류하여 교수와 조교가 직접 관찰에 의하여 점검하는 형태로 개발하였다.

1. 토론 시 참여도(총 3점)
 - 흥미를 가지고 토론에 적극적으로 참여한다(3점).
 - 참여하기는 하나 소극적이다(2점).

- 거의 참여하지 않는다(1점).
2. 실험 수행 시 참여도(총 3점)
 - 실험 수행 시 능동적이고, 다른 조원과 협조를 잘 한다(3점).
 - 실험 수행이나 다른 조원과의 협조에 소극적이다(2점).
 - 거의 참여하지 않는다(1점).
 3. 주의 사항 지키기(총 3점)
 - 실험 시 주의사항을 잘 지킨다(3점).
 - 실험 시 주의사항을 대체로 지킨다(2점).
 - 거의 지키지 않는다(1점).
 4. 정리 정돈(총 3점)
 - 실험 후 기구 정리 및 주변 정돈에 적극적으로 다(3점).
 - 실험 후 기구 정리 및 주변 정돈에 소극적이다(2점).
 - 거의 참여하지 않는다(1점).

지금까지 본 연구에서 일반화학실험 수업 성취도 평가를 위하여 개발한 분석적 채점 준거를 세 측면으로 나누어 제시하였다. 세 영역의 배점 비율은 평가자의 관점에 따라 각 영역에 가중치를 결정하여 조정할 수 있다. 본 연구에서는 실험 보고서 조별 평가를 50%, 수공적 기능의 개인별 평가를 10%, 실험에 임하는 태도의 개인별 평가를 20%, 그리고 학기말 실험 수행 과제 평가를 20%로 배점하였다. 학기말 실

험 수행 과제 평가는 실험 보고서 조별 평가만 실시하였다.

마지막으로, 학생들의 일반화학실험 수업 성취도와 일반화학 이론 수업 성취도와의 상관($r=.663, p<.01$)이 있는 것으로 나타났다. 그러나 두 학업 성취도에 대하여 공통 분산 영역이 44%($r^2=.440$)로 나타난 것으로 보아, 이러한 결과로부터 간접적으로나마 본 연구에서 개발한 실험 평가 도구는 이론 평가에서 측정하는 능력과 서로 다른 부분을 측정한다고 해석할 수 있으며, 전통적인 지식·이해 평가가 아닌 실험을 수행하면서 나타난 탐구 기능, 수공적 기능, 실험 태도를 평가하고 있다고 볼 수 있을 것이다. 따라서 앞으로 이에 대한 보다 구체적인 연구가 되어야 할 것이다.

IV. 결론 및 제언

본 연구에서는 사범대학 과학교육과 학생들이 수강하는 일반화학실험 수업 성취도 수행 평가 도구를 개발하였다. 이 수행 평가 도구에는 학생들의 실험 과정에서 나타나는 탐구 능력을 평가하는 실험 보고서에 대한 분석적 채점 준거와 학생들의 수공적 기능과 실험에 임하는 태도를 평가하는 분석적 채점 준거가 있다.

본 연구에서 개발한 이 평가 도구를 사용하여 얻어진 학생들의 일반화학실험 수업 성취도와 일반화학 이론 수업 성취도와의 44%($r^2=.440$)의 공통 분산 영역을 보여 주었다. 이러한 결과로부터, 간접적으로나마, 본 연구에서 개발한 실험 평가가 전통적인 지식·이해 평가가 아닌 실험을 수행하면서 나타난 탐구 기능, 수공적 기능, 실험 태도들을 평가하고 있다고 볼 수 있다. 개발된 채점 준거들의 타당도와 신뢰도는 검증받았으나, 앞으로 계속적으로 수정·보완하면 예비 과학 교사인 학생들이 수강하는 일반화학실험을 평가하는 데 보다 더 신뢰로운 평가 도구로 사용할 수 있을 것으로 생각된다.

이상의 연구 결과를 바탕으로 제시할 수 있는 몇 가지 제언은 다음과 같다.

첫째, 본 연구에서 개발한 탐구 기능 요소에 대한

일반적 채점 준거를 토대로 하면 다른 탐구 실험에 대해서도 실험 보고서를 통해 탐구 기능을 평가하는 분석적 채점 준거를 개발할 수 있을 것이다. 둘째, 본 연구에서 진행한 탐구 실험 수업을 예비 과학 교사인 학생들에게 적용한다면, 장차 그들이 중등학교 과학 교사가 되었을 때 시대적으로 요구하고 있는 실험 수행 평가를 할 수 있는 자질을 증진시키리라 사료된다. 마지막으로 지금까지 해오던 전통적 지식 이해 평가와 본 연구에서 개발한 실험 수행 평가뿐만 아니라, 현장을 고려한 과학 문제 해결력을 평가할 수 있는 수행 평가 또는 포트폴리오 평가 등의 다양한 평가 방법들이 개발되어 학생들의 과학과 학업 성취도 평가에 균형 있게 사용되어야 할 것이다.

적 요

본 연구에서는 과학 교사를 양성하는 사범대학 과학교육과 학생들의 일반화학실험 수행평가를 위한 분석적인 채점 준거를 개발하였다. 개발된 채점준거는 크게 두 가지로 나눌 수 있다. 첫 번째는 학생들이 실험을 수행하면서 나타나는 탐구 능력을 평가하기 위한 실험 보고서를 평가하는 분석적 채점 준거이다. 이 분석적 채점 준거는 일곱 가지 탐구 기능 요소로 나누어져 있으며, 한 학기용 실험 주제 9개에 대해서 모두 개발하였다. 두 번째는 학생들이 실험하는 동안에 교수와 조교가 직접적인 관찰에 의해 점검하는 형태로 평가하는 수공적 기능 채점 준거와 실험에 임하는 학생들의 태도를 평가하기 위한 분석적 채점 준거이다. 또한 학생들의 탐구 능력을 평가하는 실험 보고서에 대한 분석적 채점 준거의 기초가 되는 일곱 가지 탐구 기능 요소에 대한 일반적 채점 준거도 개발하였다. 모든 채점 준거에 대한 타당도는 과학교육 전문가 6인에 의하여 검증 받았으며, 채점 준거로 실험 보고서를 평가한 결과로부터 채점자간 신뢰도를 검증하였다. 본 연구에서 개발한 평가 도구를 사용하여 얻어진 학생들의 일반화학실험 수업 성취도와 일반화학 이론 수업 성취도는 상관이 있는 것($r=.663, p<.01$)으로 나타났으며, 44%($r^2=.440$)의 공통 분산 영역을 보여주었다. 따라서 본 연구에서 개발한

실험 평가 도구는 이론 수업 평가 도구에서 측정하지 못하는 다른 영역(56%)도 평가하고 있음을 알 수 있었다. 이러한 결과로부터 간접적으로나마 본 연구에서 개발한 실험 평가 도구가 전통적인 지식·이해만이 아닌, 실험을 수행하면서 나타난 탐구 기능, 수공적 기능, 실험 태도를 평가하는 도구라고 할 수 있다.

주요어: 수행 평가, 탐구 능력, 화학 실험 수업 평가, 수공적 기능 평가, 실험 태도 평가, 실험 탐구 능력 분석적 채점 준거, 수공적 기능 분석적 채점 준거, 실험 태도 분석적 채점 준거, 탐구적 일반화학실험

참 고 문 헌

- 강순희, 박종윤, 최병순(1997). 우수한 과학교사 양성을 위한 탐구적 일반화학실험서 개발. 교육부 보고, RR 95-IV-2.
- 교육부(1999). 교육소식 제1호.
- 남정희(1996). 화학실험에서 탐구과정기능의 평가도구 개발. 한국교원대학교 대학원 박사학위논문.
- 서혜애, 신성현(1998). 미국의 과학교육 기준개발. 과학교육, 411, 53-62.
- 서혜애, 신성현(1999a). 미국의 과학교육 기준개발 II. 과학교육, 412, 53-61.
- 서혜애, 신성현(1999b). 미국의 과학교육 기준개발 III. 과학교육, 413, 51-61.
- 최연희, 권오남, 성태제(1998). 중학교 영어·수학교과에서의 열린 교육을 위한 수행평가 적용 및 효과 분석 연구. 서울: 교육부.
- Harnish, D. L.(1999). Performance Based View of Assessment Design. 이화여자대학교 초청 세미나, 1999년 3월 17일. 서울: 이화여자대학교.
- Lawson, A. E., & Birk, J. P.(1994). *Chemistry-A Critical Thinking Approach*. Department of Chemistry, Arizona State Univ., U.S.A. Unpublished experiments textbook for students and teacher's guides.
- Mathematical Science Education Board of National Research Council(MSEBNRC)(1997). 초등 수학 수행평가 과제제작 및 분석. 권오남, 김경자 역. 서울: 양서원. 원전: *Measuring Up: Prototypes for Mathematics Assessment*. The National Academy Press, Washington, D.C., U.S.A.(1993).
- National Council of Teachers of Mathematics (NCTM)(1991). 수학 교육과정과 평가의 새로운 방향. 구광조, 오병승, 류희찬 역. 서울: 경문사. 원전: *Curriculum and Evaluation Standards for School Mathematics*, U.S.A.(1989).
- National Science Education Standards(NSES)(1996). National Research Council, U.S.A.
- O'Malley, J. M., & Pierce, L. V.(1992). *Performance and Portfolio Assessment for Language Minority Students*. NCBE Program Information Guide Series, No. 9.
- O'Malley, J. M., & Pierce, L. V.(1996). *Authentic Assessment for English Language Learners*. Reading, MA: Addison-Wesley.
- O'Neil, J.(1992). Putting Performance Assessment to the Test. *Educational Leadership*, 49(8), 14-19.
- Wisconsin Performance Assessment Development Project(WPADP)(1996). Wisconsin Student Assessment System: Performance Assessment Simpler. WI: University of Wisconsin-Madison.