

초등 과학 포트폴리오 평가와 다른 과학 평가 방법 간의 상관관계 분석

신현옥 · 이기영[†] · 김찬종[‡]

(신탄진초등학교) · (한성과학고등학교)[†] · (서울대학교)[‡]

An Analysis of Correlations Between Portfolio Assessment and Other Assessment Methods in Elementary School Science

Shin, Hyun-Ok · Lee, Ki-Young[†] · Kim, Chan-Jong[‡]

(Shintainjin Elementary School) · (Hansung Science High School)[†] · (Seoul National University)[‡]

ABSTRACT

The purpose of this study was to investigate the measuring range of portfolio assessment on students' abilities, using correlation coefficients of portfolio assessment with four other assessment methods. A portfolio system was developed based on four units of elementary fourth grade science, and applied to three fourth-grade science classes in Daejeon Metropolitan area and Gyunggi-do. Four different assessment methods, multiple-choice (short answer type), essay-type item, mind-mapping and laboratory assessment were also administered and scored by two elementary teachers attending graduate school. Correlation coefficients between portfolio assessment and four assessment methods were calculated with SPSS. Portfolio assessment showed moderate correlation with multiple-choice (short answer), essay-type, and mind-mapping, and low correlation with laboratory assessment. Portfolio assessment showed high correlation with multiple-choice assessment in 'understanding' and 'application' domains, but showed low correlation in 'recall' and 'inquiry' domains. In conclusion, portfolio assessment could measure various abilities measured by other assessment methods.

Key words : portfolio, science assessment, correlation, assessment method

I. 서 론

과학 지식 위주의 결과 중심의 전통적인 평가는 탐구 능력이나 과학 태도, 또는 창의성 등에 대한 개인별 수준이나 진전 상황에 관한 정보를 충분히 제공해주지 못한다는 비판을 받아왔다. 전통적인 평가 체제의 대안으로 대두되고 있는 것은 인간을 규정하는 것이 아닌 인간을 이해하는, 결과뿐만 아니라 과정까지 평가하며(Gitomer, 1993), 교수와 학습 그리고 평가를 통합적인 관점에서 접근하는 수행평가(performance assessment)이다. 수행평가는 1980년대 말부터 미국과 영국을 비롯한 많은 나라에서 학생들의 주요 특성을 평가하기 위한 노력으로 도입하고 있으며, 우리나라에서도 1990년대 후반부터 학교 현

장에 광범위하게 도입되어 실시되고 있다(김찬종 등, 2001).

과학 교과에서는 학생들의 구체적 경험(hands-on experience)을 통한 학습을 강조하며, 이러한 활동들은 학생의 탐구 능력 신장으로 이어지기 때문에 수행평가의 본질을 보다 더 잘 구현할 수 있다. 또한 수행평가 도입 이전부터 이미 지필 검사, 관찰, 보고서 평가, 실험 및 실습 등 지식, 탐구능력, 태도 평가를 위해 다양한 수행평가 기법을 활용해 왔다.

과학에서 사용할 수 있는 수행평가의 방법으로는 서술형 및 논술형 평가, 보고서 평가, 관찰 평가, 포트폴리오, 개념도, V도, 실기 평가, 면담, 그림법, 역할놀이를 이용한 평가 등이 있다(이양락, 1999; Doran *et al.*, 1998; Stiggins, 1994). 이 중 포트폴리

오(portfolio)는 학생의 노력, 발전, 성취를 보여주는 의미 있는 작품들의 모음집으로, 내용 선택에 있어서 학생들의 참여, 선택의 준거, 자기반성의 증거들이 포함된다(Paulson *et al.*, 1991). 포트폴리오는 단순히 자료들의 모음(folder)만을 의미하는 것은 아니며, 과학이나 수학 같은 과목에서 학생들의 성취를 보여주거나 모든 교과에서 범교과적인 능력을 평가하는 데 주로 이용된다. 포트폴리오를 이용하면 학생 중심 학습에서의 학습 결과뿐만 아니라 학습 과정까지도 파악할 수 있으며, 자기평가, 동료평가, 교사평가를 다양하게 이용할 수 있어 평가 측면 뿐 아니라 교육적인 측면에서도 효율적인 도구로 여겨진다. 또한 포트폴리오는 자신의 수준을 발견하게 하여 부족한 점을 찾고, 무엇을 해야 할 것인가를 결정하여 스스로 과제를 해 나가게 하는 도구이며(Wiggins, 1993), 교사들이 자신의 교수에 대한 반성의 한 수단으로 이용할 수 있는 역동적인 교수 전략 도구이기도 하다. 또한 포트폴리오가 교수 학습 측면에서 여러 긍정적인 효과가 있음은 많은 선행 연구들에 의해 밝혀진 바 있다(김혜정, 1998; 손수남, 1999; 이수정, 2002; 한세란, 1999). 또한 평가 측면에서도 타당도와 신뢰도가 높은 것으로 나타났다(윤선아, 2001; 최미애, 2001).

초등 과학 수업에서의 평가 방법에 관한 김찬중 등(2002)의 설문 결과에 의하면, 조사 대상의 20.9%에 해당되는 교사들이 실제로 포트폴리오를 사용하고 있었으며, 26.6%의 교사들이 포트폴리오가 바람직한 평가 방법인 것으로 생각하고 있었다. 그러나 실제로 교사들은 포트폴리오에 비해 덜 바람직하다고 생각하는 서술형이나 선택형 검사를 더 많이 사용하고 있는 것으로 나타났다. 이것은 소요 시간, 교사의 노력 등 포트폴리오가 지닌 현실적인 어려움 때문에 판단할 수도 있겠지만, 과학 평가 방법으로 포트폴리오의 필요성이나 유용성에 대해 확신하고 있지 못하기 때문일 수도 있다. 그러므로 이러한 현실적인 어려움에도 불구하고 교사들이 포트폴리오 평가를 유용한 도구로써 그 필요성을 인식하고 보편적으로 사용할 수 있도록 하기 위해서는 포트폴리오가 실제로 다른 여러 가지 평가 방법들이 측정하고 있는 다양한 평가 영역을 어느 정도 대체할 수 있는 대안적인(alternative) 평가 도구가 될 수 있는지에 대한 통계적인 검증이 있어야 할 것이다. 이러한 필요성을 토대로 본 연구에서는 초등 과학 포트폴리오 평가와

다른 평가 방법 간의 상관관계를 분석해봄으로써 포트폴리오 평가가 어떤 영역을 평가할 수 있는지 알아보고자 하였다.

본 연구에서는 다음과 같은 네 가지 구체적인 연구 문제를 설정하였다.

1. 초등 과학 포트폴리오와 다른 평가 방법 사이의 상관관계는 각각 어떠한가?
2. 초등 과학 포트폴리오와 다른 평가 방법 사이의 상관관계는 영역에 따라 어떤 차이가 있는가?
3. 초등 과학 포트폴리오와 다른 평가 방법 사이의 상관관계는 내용에 따라 어떤 차이가 있는가?
4. 초등 과학 포트폴리오는 어떤 영역을 평가할 수 있는가?

II. 연구 방법

연구 방법은 연구 대상과 연구 절차로 구분하여 서술하였다. 연구 절차는 다시 평가 도구 개발과 적용 및 채점에 대한 내용을 정리하였다.

1. 연구 대상

이 연구는 대전광역시 소재 S초등학교 4학년 2개 반과 경기도 용인시에 소재한 S초등학교 4학년 1개 반 총 120명의 학생을 대상으로 실시하였다. 포트폴리오 평가 체제 개발 및 적용 단원은 4학년 1학기 과학 교과서의 5단원 '혼합물의 분리', 6단원 '식물의 뿌리', 7단원 '강과 바다', 8단원 '별자리를 찾아서'와 같이 4개 단원을 선정하였다. 이 단원들을 선정한 이유는 실험 내용이 많이 포함되어 있을 뿐만 아니라 학생 위주의 활동들이 많아 학생들이 신출해내는 결과물들이 풍부할 것이라 판단했기 때문이다.

2. 연구 절차

가. 교과서 분석

단원 내에서 해당 차시의 중요도, 실험 관찰에서 포트폴리오 적용상의 중요도와 차시별 유용성 등을 분석하여 평가도구 개발을 위한 기초 자료를 제공하기 위해 교과서 분석을 실시하였다. 교과서 분석은 과학교육 전문가 1인 그리고 대학원 재학 중인 초등 교사 3인에 의해 2001년 11월부터 2002년 1월 까지 교육과정 해설서, 교과서, 실험 관찰 교재 그리고 교사용 지도서를 대상으로 이루어졌다. 분석은 표 1의 예시와 같이 김찬중(1999)의 필수 학습 목표 분류를

표 1. 교과서 필수 학습 목표 분류 예시(5단원)

| 차시 | 영역 | 기억 | 이해 | 적용 | 탐구 |
|-----|-----------------|----|----------------------------|---------------------------|------------------------------|
| 1 | 혼합물 | | 침전과 여과를 이용한 혼합물의 분리 | 수돗물 알갱이 크기를 이용한 혼합물 분리의 예 | 실험기능(거름장치로 분리하기) |
| 2 | 증발을 이용한 혼합물의 분리 | | | 염전 | 실험기능(증발장치로 분리하기) |
| 3 | | | 물에 녹는 고체와 녹지 않는 고체 혼합물의 분리 | | 실험기능(여러 가지 분리 방법으로 혼합물 분리하기) |
| 4 | | | 서로 섞이지 않는 액체 혼합물의 분리 | 바다에 유출된 기름 제거하기 | 관찰 실험기능(스포이트로 분리하기) |
| 5 | | | 색소 혼합물의 분리 | | 실험기능(색소 분리하기) |
| 6~7 | | | | 두부 만들기(우유에서 단백질 분리하기) | |

을 이용하여 영역과, 차시별로 학습 내용과 목표를 추출하였다.

나. 평가 도구 개발

이 연구에서는 표 2에서와 같이 4개 단원을 대상으로 총 5가지 평가 도구를 개발하였다. 평가 도구는 각 단원의 성격에 맞게 개발되었으며, 개발에는 대학원에 재학 중인 초등교사 2인과 과학교육 전문가 1인이 참여하였다. 실험 평가는 5단원 '혼합물의 분리'에서만 개발되었으며, 마인드맵은 6단원과 7단원, 서술형 평가는 5단원과 8단원에 대해서 개발되었다. 이

처럼 단원별로 평가 도구 유형을 달리한 것은 평가 도구의 유형을 다양화하면서 동시에 학생들의 부담을 줄이기 위해서였다.

포트폴리오 평가 도구는 2002년 1월부터 4월까지 4개 단원에 대하여 3~4차례의 수정 과정을 거쳐 개발되었다. 기존에 개발된 포트폴리오들을 분석하여 그 유형과 장·단점을 파악하였다. 기존의 포트폴리오는 학생이 해야 하는 과제가 너무 많다고 판단되어 단원별 수행과제는 2~3개로 제한하였다. 7차 교육과정에서 따라 개발된 실험 관찰¹⁾ 교재의 내용 중 많은 부분이 포트폴리오에 그대로 사용해도 무방할 것으로 보여서 이를 포트폴리오에 포함시켰다. 이렇게 완성된 포트폴리오 평가 도구는 2002년 5월 중순부터 7월까지 수업을 해나가면서 과학교육 전문가와 협의를 통해 부분적으로 수정하였다. 본 연구에서 개발된 포트폴리오의 수행과제는 학습 증거 자료로 수업 시간에 학습한 활동과 실생활과 관련 깊은 과제를 중심으로 구성하였으며, 실험 관찰은 수행과제와 함께 진행하였다. 표 3은 포트폴리오 수행과제 영역의 단원

표 2. 단원별 평가 도구 개발

| | 5단원 | 6단원 | 7단원 | 8단원 |
|-------------|--------|--------|-------|--------|
| 포트폴리오 평가 | ○ | ○ | ○ | ○ |
| 선다형(단답형) 평가 | ○ (12) | ○ (13) | ○ (8) | ○ (10) |
| 서술형 평가 | ○ (1) | | | ○ (1) |
| 마인드맵 평가 | | ○ | ○ | |
| 실험 평가 | ○ | | | |

*()는 문항의 수

표 3. 포트폴리오 수행과제의 단원별 구성

| | 5단원 | 6단원 | 7단원 | 8단원 |
|------------|--------------|--------------------------|-----------|--------------|
| 들어가며 | | 알고 있는 것 / 알고 싶은 것 적기 | | |
| 수행과제 1 | 소금장수 도와주기 | 눌부와 나무 | 땃목을 타고 | 복두칠성과 카시오페이아 |
| 수행과제 2 | 생활 속의 혼합물 분리 | 식물의 뿌리 공부 정리 | 어디로 눌러갈까요 | 내 별자리 |
| 되돌아보며 | | 내가 잘한 점 / 내가 부족한 점 찾기 | | |
| 친구 생각 | | 내 포트폴리오에 대한 친구들의 생각 알아보기 | | |
| 우리에게 주는 말씀 | | 선생님 말씀 / 부모님 말씀 | | |

¹⁾교과서와 별도로 만들어진 workbook 형식의 교재로, 교과서 단원의 내용과 같은 맥락의 여러 가지 다양한 활동들과 읽을거리로 구성된다.

표 4. 포트폴리오 실험관찰의 단원별 구성

| | 5단원 | 6단원 | 7단원 | 8단원 |
|------------|---|---|--|--|
| 활동 | 분리하기(흙탕물, 소금물, 소금과 모래, 물과 식용유, 사인펜 잉크 색소) 두 부 만들기 | 식물의 뿌리 모양 관찰하기/뿌리가 하는 일 알아보기/물방울의 식물어행/식물이 되어보기 | 강과 바다 조사계획 세우기/여러 강 모습 특징 살펴보기/물에 의한 땅 모양 변화 관찰하기/바다 밑의 땅모양 알아보기/강과 바다 조사결과 발표하기 | 밤하늘의 별 관찰하기/밤하늘의 별자리 관찰하기/별자리 이야기와 별자리 만들기/계절에 따른 별자리 알아보기/하룻동안 별자리 움직임 알아보기 |
| 이런 실험도 있어요 | 우유에서 단백질 분리하기 | 고구마 기르기 | | |
| 읽을거리 | 수돗물이 우리집에 오기까지/분유 만들기/하수의 처리 | 뿌리와 우리생활 | 강물은 조각가/바다 이야기 | 북두칠성과 카시오페이아, 견우와 직녀, 오리온자리에 얽힌 이야기 |

별 구성을 나타낸 것이고, 표 4는 실험관찰의 단원별 구성을 나타낸 것이다.

선다형(단답형) 평가 도구는 2002년 3월부터 해당 단원의 평가가 이루어지기 전까지 단원별로 3~5차례 수정 과정을 거쳐 개발되었다. 평가 문항은 표 5에서 보는 바와 같이 교과서 분석 시에 작성된 필수 학습 목표 분류들에 제시된 영역별, 차시별 내용으로 구성하였다. 4개 단원에서 총 43문항이 개발되었으며, 영역별로는 ‘이해’와 ‘기억’ 영역이 ‘적용’이나 ‘탐구’ 영역에 비해 많았다.

표 5. 평가 영역별 선다형 문항 수

| 단원 | 영역 | | | | | 계 |
|----|----|----|----|----|--|----|
| | 기억 | 이해 | 적용 | 탐구 | | |
| 5 | 2 | 5 | 4 | 1 | | 12 |
| 6 | 7 | 3 | 2 | 1 | | 13 |
| 7 | 3 | 3 | 1 | 1 | | 8 |
| 8 | 1 | 5 | 1 | 3 | | 10 |
| 계 | 13 | 16 | 8 | 6 | | 43 |

실험 평가는 ‘소금물+흙탕물의 분리’라는 과제를 선정하였으며, 4가지 평가 항목과 9가지 평가 기준을 설정하여 각 항목을 체크리스트를 이용하여 관찰 평가하였다(부록 참조). 또 개인별로 실험 기구를 주고 직접 실험하는 과정을 평가자가 관찰하여 채점하는 방식이므로 현실적인 여건을 고려하여 실험을 수행하는 학생들 중 절반만 평가하도록 하였다. 서술형 평가, 마인드맵 평가, 실험 평가 도구의 문항은 대학원 재학 중인 초등교사 3인과 과학 교육 전문가가 함께 개발하였으며 1~2차례의 수정 과정을 거쳤다.

다. 평가 도구 적용

포트폴리오를 적용한 수업은 2002년 5월 중순부터

7월까지 약 10주에 걸쳐 연구자와 연구에 참여한 3명의 교사에게 의해 이루어졌다. 포트폴리오 수행 과제 중 성격상 방과 후 과제에 해당되는 것은 일정한 시간을 주고 학생들이 원하는 시간에 포트폴리오를 작성할 수 있도록 자율성을 부여하였다. 또한 수행 과제와 실험 관찰을 병행하여 진행하면서 활동 내용을 한 눈에 알아볼 수 있도록 목차에 주제와 쪽수를 기록하도록 하였다.

선다형(단답형) 평가, 서술형 평가, 마인드맵 평가 그리고 실험 평가는 각 단원의 포트폴리오 적용 수업이 끝날 때마다 약 1주일 후에 실시하였다. 즉, 5단원 수업이 끝난 1주일 후에 선다형 평가와 서술형 평가 그리고 실험 평가를 실시하였다. 실험 평가는 사전에 평가 내용에 대해 미리 알도록 하였으며, 5회에 걸쳐 한 번에 8명의 학생을 입실시킨 후 2명의 평가자에 의해 각 2명씩 총 20명의 실험과정을 관찰하여 평가하였다. 그러나 실험 평가에 참여하는 학생들은 이 사실을 모르게 하였다. 40명의 학생이 실험을 하되 그 중 절반인 20명만을 관찰 평가한 것은 한 명의 평가자가 2명 이상의 학생의 전 실험 과정을 한꺼번에 관찰할 수 없다고 판단하였기 때문이다.

라. 평가자간 일치도

2명이 교차하여 채점한 평가 도구의 경우, 채점에 참여한 평가자간의 일치도를 알아보기 위하여 몇 개의 사례를 임의로 선정하여 채점하여 Cohen의 Kappa 계수를 산출하였다. Cohen(1960)의 Kappa 계수는 일치도 통계 중 우연에 의해 일치될 확률을 제거하여 두 채점자간 일치도를 추정한 것으로, Fleiss(1981)의 판단 기준에 의하면 .40 미만이면 평가자간 일치도가 낮고, .40 .60이면 일치도가 있고, .60 .75이면 일치도가 높고, .75 이상이면 일치도가

매우 높은 것으로 판단한다.

본 연구에서 개발된 평가 도구들을 적용한 결과 포트폴리오와 마인드맵 평가 도구의 평가자간 일치도는 표 6과 같았다. Fleiss의 판단 기준에 의하면, 평가자간 일치도는 높거나 매우 높은 것으로 나타났다.

표 6. 평가자간 일치도(Cohen의 Kappa 계수)

| | | 사례 수 | Kappa 계수 |
|-------------|-------|------|----------|
| 포트폴리오 평가 | 수행 과제 | 24 | .71 |
| | 실험 관찰 | 60 | .85 |
| 마인드맵 평가 | | 70 | .62 |

마. 상관계수 산출

자료의 처리에는 SPSSWIN 통계 프로그램을 사용하였으며, Pearson의 적률 상관을 이용하여 포트폴리오와 선다형 평가, 서술형 평가, 마인드맵 평가 그리고 실험평가 간의 상관계수를 산출하였다.

III. 연구 결과

4개 단원에 포트폴리오를 적용한 결과와 각 단원의 성격에 맞는 4가지 평가 방법을 시행한 결과와의 상관관계를 알아보고자 하였다. 우선, 4개 단원 전체를 대상으로 한 상관관계를 분석한 다음, 평가 영역(domain)별로 여러 평가 방법과 포트폴리오 평가와의 상관관계를 분석하였고, 마지막으로 평가 내용(unit)별로 포트폴리오와 다른 평가 방법들 간의 상관관계를 분석하였다. 상관계수에 따른 상관관계를 언어적으로 표현하는 절대적인 기준은 없으나 본 연구에서 상관관계의 판단은 성태제(2002)의 기준을 적용하였다. 상관계수 .00 이상 .20 미만은 ‘상관이 거의 없다’로, .20 이상 .40 미만은 ‘상관이 낮다’로, .40 이상 .60 미만은 ‘상관이 있다’로, .60 이상 .80 미만은 ‘상관이 높다’로, .80 이상 1.00 미만은 ‘상관이 매우 높다’로 판단하였다.

1. 포트폴리오와 다른 평가 방법 간의 전체 상관관계

표 7은 포트폴리오 평가와 다른 4가지 평가 방법 간의 전체 상관계수를 나타낸 것이다. 선다형(단답형), 서술형, 마인드맵 평가는 전체적으로 포트폴리오 평가와 상관이 있는 것으로 나타났으나, 실험 평가는 상관이 낮은 것으로 나타났다. 또한 4가지 평가 방법

중 포트폴리오 평가와 상관이 가장 높은 평가 방법은 선다형 평가인 것으로 판단할 수 있었으며, 모두 유의도 1% 수준에서 통계적으로 유의미한 상관을 보여주었다. 이러한 분석 결과로 판단할 때, 포트폴리오 평가는 기존의 다른 평가 방법들이 측정하고 있는 영역을 어느 정도 종합적으로 측정할 수 있는 평가 도구라고 볼 수 있다. 그러나 실험 평가와의 상관이 낮은 것은 포트폴리오 평가가 탐구 기능(skills)과 관련된 영역을 제대로 반영하지 못한 결과인 것으로 판단된다.

표 7. 포트폴리오와 다른 평가 방법 간 전체 상관계수

| | 선다형 (단답형) | 서술형 | 마인드맵 | 실험 |
|-------|--------------|-------|-------|-------|
| 포트폴리오 | .519* | .415* | .460* | .383* |

*p<.01

2. 포트폴리오와 다른 평가 방법 간의 영역별 상관관계

표 8은 선다형(단답형) 평가와 포트폴리오 평가와의 상관계수를 평가 영역별로 산출한 것이다. 포트폴리오 평가는 수행과제와 실험관찰 2개 영역으로 구분하였으며, 선다형 평가는 기억, 이해, 적용, 탐구 4개 영역으로 구분하였다. 포트폴리오 평가의 ‘수행과제’ 영역과 ‘실험관찰’ 영역 모두 선다형 평가와 높은 상관을 나타내었다. 선다형 평가 영역 중에서는 ‘기억’과 ‘탐구’ 영역이 포트폴리오와 상관이 낮은 것으로 나타났으며, ‘이해’와 ‘적용’ 영역은 상관이 있는 것으로 나타났다. 포트폴리오 평가와 가장 상관이 높게 나타난 것은 ‘적용’ 영역(.494)이었으며, ‘탐구’ 영역(.290)에서 가장 낮은 상관을 보였다. 영역별로 가장 높은 상관을 보인 것은 포트폴리오 평가의 ‘수행과제’ 영역과 선다형 평가의 ‘적용’ 영역(.589)이었으며, 가장 낮은 상관을 보인 것은 포트폴리오 평가의 ‘실험관찰’ 영역과 선다형 평가의 ‘탐구’ 영역(.371)이었다. 이러한 결과는 포트폴리오의 구성에서 수행과제가 주로 실생활 적용과 관련된 주제였기 때문인 것으로 판단된다. 하지만 ‘탐구’ 영역이 포트폴리오와 상관이 낮게 나온 것은 일반적인 예상을 벗어나는 결과였다. 특히 ‘실험관찰’과 ‘탐구’ 영역의 상관이 가장 낮은 것은 좀 더 심도 있게 논의해볼 문제라고 판단하여, 표 5의 실험관찰 구성과 선다형 문항의 탐구 영역을 비교하여 분석하여 보았다. 그

결과 전체 43개의 선다형 문항 중 탐구 영역에 해당되는 문항은 6개였으며, 이 중 4문항이 탐구 과정 중 '실험수행'이었으며, 나머지는 '실험설계'가 1문항, '결론도출'이 1문항이었다. 내용별로도 대부분이 실험, 관찰 등 탐구활동 위주로 구성되어 있는 실험관찰 내용을 토대로 개발되어 탐구 과정에서 별다른 차이를 발견할 수는 없었다. 하지만 단원별로는 5, 6, 7단원에서 1문항씩, 8단원에서 3문항이 출제되었는데, 전체 6문항 중 절반에 해당되는 3문항이 특정 단원에서 출제된 것은 문제가 있어 보였다. 다시 말해 전체적으로 탐구 영역 선다형 문항이 다른 영역에 비해 상대적으로 문항수가 적었고, 단원에 따라 고르게 출제되지 못한 것이 포트폴리오와 낮은 상관을 나타낸 하나의 원인이 될 수 있을 것이라 생각된다.

표 8. 포트폴리오와 선다형 평가 간 영역별 상관계수

| | 선다형(단답형) | | | | |
|------------|----------|-------|-------|-------|-------|
| | 기억 | 이해 | 적용 | 탐구 | 계 |
| 수행과제 | .584* | .583* | .589* | .355* | .709* |
| 포트폴리오 실험관찰 | .470* | .460* | .634* | .371* | .618* |
| 계 | .352* | .422* | .494* | .290* | .519* |

* $p < .01$

표 9와 표 10은 각각 6단원과 7단원 마인드맵 평가와 포트폴리오 평가와의 상관계수를 영역별로 산출한 것이다. 마인드맵은 이수정(2002)의 채점 기준을 적용하여 크게 중심원, 가지, 표현의 3개 영역으로 구분하고, 각 영역은 또 몇 개의 하위 영역으로 구분하여 채점하였다(부록 참조).

표 9와 표 10의 결과를 비교해보면 우선, 마인드맵 평가의 내용(단원)에 따라 포트폴리오 평가와의 상관성이 큰 차이가 있었다. 6단원 마인드맵 평가가 포트폴리오 평가와 상관성이 있는 반면, 7단원은 상관성이 낮은 것으로 나타났다. 또한 영역별로도 서로 상반된 결과가 나타났다. 6단원에서는 '중심원' 영역이 가장 상관성이 높고, '표현' 영역이 가장 상관성이 낮은 반면, 7단원에서는 정반대의 결과가 나타났다. 다시 말해, 마인드맵 평가는 포트폴리오 평가와 전체적으로는 상관성이 있지만 평가 내용이나 영역별로는 일관된 경향을 보이지 않았다. 하지만 전반적으로 포트폴리오의 수행과제보다는 실험관찰 영역이 마인드맵 평가 영역과 더 높은 상관성을 나타내었으며, 두 단원 모두 '중심원' 영역에서는 '대표성'이, '가지' 영역에서는 '위계성'이 가장 높은 상관성을 나타내었다. '대표성'과 상

관이 높다는 것은 포트폴리오가 주요 학습 개념의 파지 정도를 전체적으로 측정하는 평가 도구가 될 수 있으며, '위계성'과 상관성이 높다는 것은 포트폴리오가 파지 개념들의 체계성을 측정하는 평가 도구가 될 수 있음을 시사한다.

표 9. 6단원 마인드맵과 포트폴리오 평가 간 영역별 상관계수

| 영역 | 포트폴리오 | | | |
|-----|---------|--------|--------|--------|
| | 수행과제 | 실험관찰 | 계 | |
| 중심원 | 대표성 | .349** | .542** | .447** |
| | 독립성 | .227* | .236* | .201* |
| | 포괄성/균형성 | .254** | .396** | .352** |
| | 소계 | .383** | .564** | .478** |
| 가지 | 단계수 | .144 | .315** | .213* |
| | 잔가지수 | .156 | .248** | .124 |
| | 위계성 | .317** | .421** | .403** |
| | 소계 | .309** | .479** | .377** |
| 표현 | | -.048 | .210* | 0.99 |
| 계 | | .353** | .559** | .453** |

* $p < .05$, ** $p < .01$

표 10. 7단원 마인드맵과 포트폴리오 평가 간 영역별 상관계수

| 영역 | 포트폴리오 | | | |
|-----|---------|--------|--------|--------|
| | 수행과제 | 실험관찰 | 계 | |
| 중심원 | 대표성 | .146 | .117 | .138 |
| | 독립성 | .134 | .290* | .122 |
| | 포괄성/균형성 | -.003 | .144 | .059 |
| | 소계 | .127 | .188 | .138 |
| 가지 | 단계수 | .121 | .183 | .258** |
| | 잔가지수 | .254** | .134 | .202* |
| | 위계성 | .189 | .246* | .293** |
| | 소계 | .243* | .261* | .335** |
| 표현 | | .251** | .355** | .337** |
| 계 | | .217* | .267** | .278** |

* $p < .05$, ** $p < .01$

3. 포트폴리오와 다른 평가 방법 간의 내용별 상관관계

표 11은 평가 내용(단원)별로 포트폴리오 평가와 다른 평가 방법 간의 상관계수를 나타낸 것이다. 4개 단원에서 모두 실시한 선다형(단답형) 평가의 경우는, 포트폴리오 평가와 전체적으로는 상관(519)이 있는 것으로 나타났지만 내용별로는 8단원(429)을 제외하고는 상관계수 .40 미만으로 포트폴리오 평가와 상관성이 낮은 것으로 나타났다. 2개 단원에서 실시한 서술

형 평가는 포트폴리오 평가와 전체적으로는 상관(.415)이 있는 것으로 나타났으나, 단원 간에는 상당한 차이가 있었다. 5단원 서술형 평가의 경우는 포트폴리오 평가와 상관(.278)이 낮은 반면, 8단원의 경우는 상관(.422)이 있는 것으로 나타났다. 2개 단원에서 실시한 마인드맵 평가도 서술형 평가와 마찬가지로 단원 간에 상당한 차이를 보였다. 6단원 마인드맵 평가의 경우는 포트폴리오 평가와 상관이 있게 나타난 반면, 7단원은 상관이 낮은 것으로 나타났다. 1개 단원에서 실시한 실험 평가의 경우는 포트폴리오 평가와의 상관이 낮은 것으로 나타났다.

표 11. 포트폴리오와 다른 평가 방법 간 내용(단원)별 상관계수

| | 선다형 (단답형) | 서술형 | 마인드맵 | 실험 |
|-----------|--------------|--------|--------|--------|
| 5단원 | .339** | .278** | . | .383** |
| 6단원 | .398** | . | .453** | . |
| 포트폴리오 7단원 | .322** | . | .278** | . |
| 8단원 | .429** | .422** | . | . |
| 전체 | .519** | .415** | .460** | .383** |

* $p < .05$, ** $p < .01$

V. 결론 및 제언

본 연구에서는 초등 과학 포트폴리오 평가와 다른 평가 방법 사이의 상관관계를 다양한 측면에서 분석해봄으로써 포트폴리오 평가의 유용성에 대해 알아보고자 하였다. 이를 위해 4개 단원을 대상으로 포트폴리오 체제를 개발하고 수업에 적용한 후 단원의 성격에 따라 선다형(단답형) 평가, 서술형 평가, 마인드맵 평가, 실험 평가를 각각 실시하여 포트폴리오 평가와의 상관관계를 분석하였다.

본 연구 결과를 토대로 한 결론은 다음과 같다.

첫째, 선다형(단답형), 서술형, 마인드맵 평가는 포트폴리오 평가와 측정하는 영역이 상당 부분 중복된다고 볼 수 있으나, 실험 평가는 포트폴리오 평가와 측정 영역에서 중복되는 부분이 상대적으로 적다.

둘째, 평가 영역 중에서 포트폴리오 평가는 선다형(단답형) 평가와 '이해'와 '적용' 영역에서는 중복되는 부분이 많으나, '기억'이나 '탐구' 영역에서는 상대적으로 중복되는 부분이 적다.

셋째, 평가 내용에 따라서는 선다형(단답형) 평가는 포트폴리오와의 상관계수가 일관성을 보이지만, 서술

형과 마인드맵 평가는 일관성이 적어서 안정성(stability)이 낮은 것으로 판단된다.

본 연구에서 포트폴리오 평가와 실험 평가 사이의 상관계수가 낮게 나타났다. 그러나 실험 평가는 단지 1회만 실시되었기 때문에, 이 결과만을 가지고 이 두 가지 평가 방법이 서로 다른 영역을 측정한다고 결론을 내리기는 이르다. 보다 많은 실험 평가를 실시하고, 그 결과를 종합적으로 검토할 필요가 있다. 또한 포트폴리오 체제의 설계에서 실험과 탐구와 관련된 부분을 어떻게 반영하느냐에 따라서 실험 평가와의 상관관계를 높일 수 있을 것으로 기대된다.

본 연구를 통해서 포트폴리오 평가를 통해서 측정할 수 있는 영역이 기존의 다른 과학 평가 방법들이 측정하는 영역과 상당 부분 중복됨으로써 교환가능성(interchangeability)이 있음을 알 수 있었으며, 이를 통해서 포트폴리오 평가의 활용을 통해서 교수학습 및 평가의 통합을 기대할 수 있을 뿐 아니라, 기존의 여러 평가 방법들을 어느 정도 대체할 수 있음을 알 수 있었다.

참고문헌

- 김찬중(1999). 과학 수행평가를 위한 포트폴리오 평가의 적용. 이화여자대학교 교과교육연구소 주최 '수행평가의 방향과 과제' 학술 세미나 주제 발표 논문, 1999. 5. 19.
- 김찬중, 김진규, 임형(2001). 과학과 수행평가의 이해와 활용. 서울: 경문사.
- 김찬중, 이선경, 김수환, 고은성(2002). 충청지역 초등학교 과학교육 실태 조사. 청주교육대학교 과학교육연구소 논문집, 수학과 과학교육, 23, 55-72.
- 김혜정(1998). 초등학교 과학 평가에서 포트폴리오 평가의 적용이 과학 지식과 탐구능력 및 태도에 미치는 영향. 청주교육대학교 석사학위 논문.
- 성태제(2002). 타당도와 신뢰도. 학지사.
- 손수남(1999). 초등학교 수업에 포트폴리오 체제의 적용이 사회 심리학적 교실 환경에 미치는 영향. 청주교육대학교 석사학위 논문.
- 윤선아(2001). 초등학교 과학과 학생 포트폴리오의 내용타당도 검증 및 포트폴리오에서 얻을 수 있는 정보의 유형 분석. 청주교육대학교 석사학위 논문.
- 이양락(1999). 중학교 과학과 수행평가 시행 방안 및 자료 개발 연구. 한국교육과정평가원 연구보고 CRE 99-7.
- 이수정(2002). 포트폴리오 체제의 적용이 초등학교 학생의 개념의 양과 조직에 미치는 영향. 청주교육대학교 석사학위 논문.
- 최미애(2001). 초등학교 과학과 포트폴리오의 체점 기준개

발과 신뢰도 검증. 청주교육대학교 석사학위 논문.
 한세란(1999). 초등학교 자연과에서 포트폴리오 수업 활동이 학생들의 창의성과 탐구능력에 미치는 영향. 서울교육대학교 석사학위 논문.
 Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
 Doran, R., Chan, F., & Tamir, P. (1998). *Science educator's guide to assessment*. Virginia: National Science Teachers Association.
 Fleiss, J. L. (1981). *Statistical methods for rates and proportion*. New York: Wiley.
 Gitomer, D. H. (1993). Performance assessment and educa-

tional measurement. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 241-264). Hillsdale, NJ: Lawrence Erlbaum Associates.
 Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48(5), 60-63.
 Stiggins, R. J. (1994). *Student-Centered Classroom Assessment*. New York: Macmillan.
 Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey-Bass.

<부 록>

1. 실험 평가 문항

실험 평가 4학년 1학기 5. 혼합물의 분리

- 1) 평가 목표: 거름장치, 증발 장치를 바르게 꾸미고 혼합물을 분리할 수 있다.
- 2) 개인별 자료: 깔때기 1개, 깔때기 대 1개, 거름종이 1장, 유리막대 1개, 약수저, 성냥, 알코올램프, 비커 1개
- 3) 평가 방법: 거름장치, 증발장치를 꾸미고 소금물+흙탕물 혼합물을 분리하게 한다.
- 4) 평가 관점
 - ① 거름장치를 바르게 꾸미는가?
 - ② 증발장치를 바르게 꾸미는가?
 - ③ 혼합물을 분리하는 방법이 바른가?

5) 평가 항목 및 기준

| 평가 항목 | 평가 기준 | 채 점 |
|----------|--------------------------------------|-------|
| 거름장치 꾸미기 | 거름종이를 원추형으로 접는가? | 예/아니오 |
| | 거름종이를 물에 묻혀 깔때기에 접촉시키는가? | 예/아니오 |
| | 깔때기 대에 깔때기를 똑바로 놓는가? | 예/아니오 |
| | 깔때기와 비커의 위치를 바르게 잡는가? | 예/아니오 |
| 거르는 방법 | 여과 용액을 유리막대에 대고 거르는가? | 예/아니오 |
| | 여과 용액이 깔때기 밖으로 넘치지 않도록 조심스럽게 붓는가? | 예/아니오 |
| 증발장치 꾸미기 | 분리액을 증발접시에 조금 넣고 가열준비를 하는가? | 예/아니오 |
| | 준비가 끝난 후 알코올램프의 뚜껑을 열고 올바르게 불을 붙이는가? | 예/아니오 |
| | 알코올램프를 뚜껑을 이용하여 바르게 끄는가? | 예/아니오 |
| 정리하기 | 실험기구를 깨끗이 청소하고 정리하는가? | 예/아니오 |

2. 마인드맵 영역 및 채점 기준

| 영역 | | 채점 기준 | |
|---------|------|--|-------------------------------------|
| 대표성 | 상 | 중심주제에 관련되는 주요 개념들이 80%정도 제시되어 있음 | |
| | 중 | 중심주제에 관련되는 주요 개념들이 60%정도 제시되어 있음 | |
| | 하 | 중심주제에 관련되는 주요 개념들이 1~2개만 제시되어 있음 | |
| 중심원 | 독립성 | 상 | 주가지의 중심원상의 주요 개념들이 독립적으로 사용됨 |
| | 중 | 주가지의 중심원상의 주요 개념들이 부분적으로 중복되어 사용됨 | |
| | 하 | 주가지의 중심원상의 주요 개념들이 중복되어 사용됨 | |
| 포괄성/균형성 | 상 | 주가지의 중심원상에 배열된 개념들이 모두 포괄적이며 수준이나 비중이 같음 | |
| | 중 | 주가지의 중심원상에 배열된 개념들이 대부분 포괄적이며 수준이나 비중이 비슷함 | |
| | 하 | 주가지의 중심원상에 배열된 개념들이 덜 포괄적이며 수준이나 비중이 다름 | |
| 단계의 수 | 상 | 단계간에 관련이 있으며, 각 가지의 단계가 평균적으로 3단계 이상임 | |
| | 중 | 단계간에 관련이 있으며, 각 가지의 단계가 평균적으로 2단계 이하임 | |
| | 하 | 단계간에 한 단계가 뿐이거나 단계간에 관련이 없음 | |
| 가지 | 잔가지수 | 상 | 잔가지 사이에 관련이 있으며, 평균적으로 잔가지가 3갈래 이상임 |
| | 중 | 잔가지 사이에 관련이 있으며, 평균적으로 잔가지가 2갈래 이하임 | |
| | 하 | 잔가지가 하나이거나 잔가지 개념이 관련없는 내용임 | |
| 위계성 | 상 | 개념간에 관련이 있으며 위계성(일반적→구체적)이 높음 | |
| | 중 | 개념간에 관련이 있으며 위계성이 있음 | |
| | 하 | 개념간에 관련이 없는 경우가 많거나 위계성이 낮음 | |
| 표현 | 상 | 단어와 이미지를 적절히 표현하고 전체적으로 표현이 정확하고 참신함 | |
| | 중 | 단어와 이미지 등 기타 표현이 적절함 | |
| | 하 | 단어와 이미지 표현이 적절하지 않거나 관련없는 표현을 사용함 | |