

초등 과학과 포트폴리오의 채점기준 개발과 신뢰도 검증

김찬종 · 최미애
(청주교육대학교) · (음성 대소초등학교)

Developing Scoring Rubric and the Reliability of Elementary Science Portfolio Assessment

Kim, Chan-Jong · Choi, Mi-Aee
(Chongju National University of Education) · (Eumseong Daeso Elementary School)

ABSTRACT

The purpose of the study is to develop major types of scoring rubrics of portfolio system, and estimate the reliability of the rubrics developed. The portfolio system was developed by Science Education Laboratory, Chongju National University of Education in summer, 2000. The portfolio is based on the Unit 2, The Layer and Fossil, and Unit 4, Heat and Change of Objects at fourth-grade level. Four types of scoring rubrics, holistic-general, holistic-specific, analytical-general, and analytical-specific, were developed. Students' portfolios were scored and inter-rater and intra-rater reliability were calculated. To estimate inter-rater reliability, 3 elementary teachers per each rubric (total 12) scored 12 students' portfolios. Teachers who used analytical-specific rubric scored only six portfolios because it took much more time than other rubrics. To estimate intra-rater reliability, second scoring was administered by two raters per rubric in two and half month. The results show that holistic-general rubric has high inter-rater and moderate intra-rater reliability. Holistic-specific rubric shows moderate inter- and intra-rater reliability. Analytical-general rubric has high inter-rater and moderate intra-rater reliability. Analytical-specific rubric shows high inter- and intra-rater reliability. The raters feel that general rubrics seems to be practical but not clear. Specific rubrics provide more clear guidelines for scoring but require more time and effort to develop the rubrics. Analytical-specific rubric requires more than two times of time to score each portfolio and is proved to be highly reliable but less practical.

Key words: science education, elementary, portfolio assessment, inter-rater reliability, intra-rater reliability, practicality

*2001.11.13(접수) 2001.12.19(최종 통과)

**이 논문은 2000년도 한국학술진흥재단의 지원에 의하여 연구되었음. (KRF-2000-041-D00370)

I. 서론

세계 주요 선진국에서는 지식 기반 사회로 일컬어지는 21세기에 활약할 2세 교육을 위해서 80년대 말부터 과학 교육의 개혁을 활발하게 시도하여 왔다. 이 개혁에서 가장 중요하고 핵심적인 역할을 하는 부분이 수행평가이다(Haertel, 1999; Kulm & Malcom, 1991). 수행평가는 과학교실 수준은 물론이고, 주(state) 수준, 국가 수준, 나아가서는 국제 수준의 대규모 과학 학업 성취도평가에까지 도입되고 있다(김찬중, 김진규, 임형, 2001; TIMSS, 1994). 세계 각국이 수행평가에 이처럼 많은 관심을 보이는 이유는 창의적인 지식의 생성 능력이 매우 중요시되는 21세기를 위한 인재 양성에 이 평가 방법이 크게 기여할 수 있을 것이라고 믿기 때문이다.

과학기술과 정보화에 기반을 둔 제품의 생산과 유통 뿐 아니라 창의적인 정보의 생산과 가공, 그리고 보급 등이 국제 시장에서 자국의 경쟁능력을 결정짓는 중요한 요소로 부각되고 있는 시점에서 과학교과에서 학생들의 창의성과 자기 주도성을 신장시키는 것은 국가적으로 매우 중요하고 시급한 과제가 아닐 수 없다.

우리 나라도 이러한 국제적인 추세에 발맞추어 90년대 중반부터 수행평가를 적극적으로 도입하고 있다(교육부, 1998; 국립교육평가원, 1996). 그럼에도 불구하고, 수행평가는 학교 현장에서 상반된 평가를 받고 있다. 일부에서는 미래지향적이고, 바람직한 평가 방법으로 인식하고 있는가 하면, 다른 일각에서는 어렵고 귀찮은 평가라는 인식을 하고 있다. 게다가 우리 나라의 열악한 교육 현실은 수행평가에 대한 거부감까지 불러일으키고 있으며, 수행평가의 근본 정신의 구현보다는 손쉽게 활용할 수 있는 수행평가 방법의 모색으로 흐르는 경향을 보인다. 학교 내신 성적이 상급학교 진학에 큰 영향을 주는 현 상황에서 학교 수준에서 수행평가의 도입은 교사에게는 큰 부담이, 학부모에게는 불만의 원인이 되고 있다. 수행평가의 성공적인 정착을 위해서는 수행평가의 타당성과 채점의 신뢰성을 검증하고, 이를 향상시키려는 방안을 모색해야 한다.

그 동안 국제적으로는 수행평가의 채점 기준과(Fairbrother, 1997; Feldman et al., 1998; Lederman & Niess, 1999; Shavelson et al., 1996; Shaw, 1997), 수행평가의 신뢰도(Cronbach et al., 1997; Swartz et al., 1999) 등에 대한 연구들이 활발하게 수행되고 있다. 국내에서는 수행평가에 대한 측정학적인 연구가 중학교 영어와 수학 교과(최연희, 권오남, 성태제, 1999) 등에서 수행되었으며, 수행평가가 비교적 만족할 만한 측정학적 특성을 가지고 있음이 보고되었다. 교육평가 분야에서는 수행평가의 타당도나 신뢰도 등과 같은 측정학적인 연구(김성숙, 남명호, 1995; 남현우, 1998; 성태제, 1994, 1999; 유선희, 1998)가 수행된 바 있다. 그러나 그 동안 과학과의 포트폴리오 평가에 대한 측정학적인 연구가 수행된 바는 없다.

본 연구의 주 목적은 과학과 수행평가의 방법 중 하나인 포트폴리오 평가의 다양한 채점 기준을 개발하고 신뢰도를 검증하는 것이다. 이 연구 목적을 성취하기 위해 다음과 같은 구체적인 연구 문제를 설정하였다.

- (1) 학생 포트폴리오 채점에 적절한 채점 기준 유형과 채점기준은 무엇인가?
- (2) 포트폴리오 채점 기준의 유형에 따른 채점자간 신뢰도는 어떠한가?
- (3) 포트폴리오 채점 기준의 유형에 따른 채점 시기별 신뢰도는 어떠한가?
- (4) 포트폴리오 채점 기준의 유형에 따른 실용도는 어떠한가?

II. 연구 절차 및 방법

본 연구의 내용은 크게 포트폴리오 평가 체제의 개발 및 적용, 채점 기준의 개발과 적용의 네 부분으로 나누어진다.

1. 포트폴리오 평가 체제의 개발

포트폴리오 평가 체제는 초등학교 4학년 2학기 지층과 화석, 열과 물체의 변화 단원을 대상으로 개발

하였다. 포트폴리오 평가 체제의 개발 방향은 Wiggins(1998)의 교육적 평가(educative assessment)의 개념을 바탕으로 하였다. 포트폴리오 평가 체제의 개발을 위한 구체적인 방향은 다음과 같다(김찬중, 1999).

- ① 학생들이 단원 내용에 흥미를 가질 수 있는 이야기나 만화를 초기에 제공한다.
- ② 학생들이 단원 내용과 관련하여 이미 알고 있는 것과, 배우고 싶은 것을 서술하도록 한다.
- ③ 학습 과정을 안내하고, 잘 기록할 수 있도록 학습일지 유형의 양식을 개발한다.
- ④ 수행 결과에 대한 동료, 교사 또는 부모와 상호작용을 주고받는 기회를 제공한다.
- ⑤ 피드백을 바탕으로 수행 결과에 대한 자기 조정(self-adjustment)을 권장한다.
- ⑥ 자기 평가를 실시할 기회를 제공한다.
- ⑦ 채점 기준을 바탕으로 교사가 학생들의 수행 과정이나 결과를 평가한다.
- ⑧ 평가 결과는 학생들의 장점을 더욱 신장시키고 부족한 부분을 보완하는 데 활용하도록 한다.

포트폴리오 평가 체제는 매 차시의 학습이 누적적으로 기록되는 학습일지(journal) 유형과 동료, 교사, 부모와의 상호작용을 위한 피드백을 기록할 양식과 기회를 포함하며, 이 피드백을 바탕으로 스스로 작성한 증거를 수정하고 발전시킬 수 있도록 유도한다(김찬중, 오영선, 2001).

포트폴리오 평가 체제는 2000년 7월과 8월에 걸쳐서 개발되었다. 개발에는 과학교육 전문가 1인, 초등교사 3인이 참여하였다. 개발 과정은 해당 학년의 교육과정 및 교과서 분석, 필수 학습 목표의 추출 및 분류 과정을 거쳤으며, 각 학습 목표를 위한 학습일지 양식 및 동료, 부모, 교사와 피드백을 위한 양식의 개발, 검토 및 수정 과정을 3회 이상 반복하였다.

2. 포트폴리오 평가 체제의 적용

포트폴리오 평가 체제는 경기도 평택에 소재한 A 초등학교에서 4학년 1개 반을 대상으로 적용하였다. 적용 시기는 2000년 2학기이며, 4학년 2학기 자연과

‘지층과 화석,’ ‘열과 물체의 변화’ 단원에 대하여 연속적으로 약 3개월 간에 걸쳐서 적용하였다.

연구에 참여한 학급의 담임 교사는 포트폴리오 체제 개발에 직접 참여하였으며, 그 과정에서 포트폴리오 체제의 특성과 구조 및 내용을 잘 이해하게 되었다. 포트폴리오를 적용한 학급의 학생 수는 36명이다. 참여 학급의 모든 학생이 학년말에 포트폴리오를 완성하여 제출하였다.

3. 포트폴리오 채점 기준 및 채점 양식 개발

지금까지는 포트폴리오 채점 기준에는 일반적으로 총괄적 채점기준과 분석적 채점기준이 주로 제시되어 왔다. 이 밖에도 일반적 채점 기준과 특수적 채점 기준의 문제가 발생한다. 즉, 모든 증거에 활용할 수 있는 기준(일반적)과 증거 각각에 한하여 활용할 수 있는 기준(특수적)이 그것이다. 본 연구에서는 채점 기준의 유형을 4 가지 유형으로 세분하였으며, 이는 총괄-일반(holistic-general), 총괄-특수(holistic-analytic), 그리고 분석-일반(analytic-general), 분석-특수(analytic-specific) 채점 기준이다(Fig. 1). 총괄-일반 채점기준은 소단원의 모든 목표를 채점할 수 있도록 일반적인 언어로 서술되어 있으며, 학생의 증거를 전체적으로 판단할 수 있도록 구성되어 있다. 총괄 특수 채점기준은 한 가지 학습 목표에 대한 증거만 채점할 수 있도록 특수한 언어로 서술되어 있으며, 학생의 증거를 전체적으로 판단할 수 있도록 구성되어 있다. 분석-일반 채점기준은 학생의 증거를 지식, 탐구, 태도, 실생활 등과 같은 하위 영역으로 구분하여 평가할 수 있도록 구성되어 있으며, 모든 학습 목표에 대한 증거를 채점할 수 있도록 일반적인 언어로 서술되어 있다. 분석-특수 채점기준은 학생의 증거를 지식, 탐구, 태도, 실생활 등과 같은 하위 영역으로 구분하여 채점할 수 있도록 구성되어 있으며, 한 학습 목표에 대한 증거만 채점할 수 있도록 특수한 언어로 서술되어 있다.

채점 기준 개발 과정은 채점기준 개발 방향 설정, 채점 기준 1차 개발, 1차 검토 및 수정, 채점기준 2차 개발, 2차 검토 및 수정, 예비 적용 및 수정과 같은

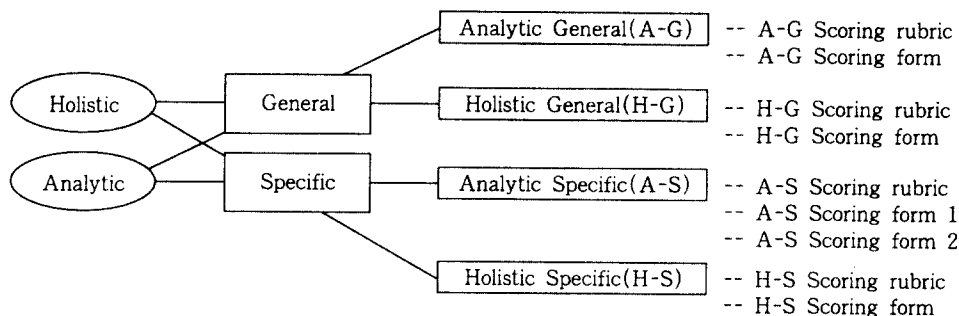


Fig. 1. Four types of scoring rubrics and scoring forms

과정을 거쳤다. 채점 기준 개발 방향의 설정은 과학 교육전문가 1인과 현직 초등교사 10명이 참여하였다. 설정된 개발 방향은 모든 채점 기준이 동일한 4단계 등급을 갖도록 하며, 네 등급은 상위에서부터 뛰어난 수행, 충실한 수행, 부족한 수행, 잘못된 수행으로 명명하였다. 학생 포트폴리오의 채점 단위는 초등학교 자연과 교사용지도서에 제시된 단원의 학습 목표를 기준으로 삼았다. 또한 채점 기준 이외에 채점과 채점 결과의 활용에 도움이 되도록 채점 양식과 채점 결과 기록 양식을 동시에 개발하기로 하였다.

채점 기준 1차 개발은 2001년 1월에 실시되었다. 개발에는 현직 초등 교사 1인이 약 2주 동안 기초 작업을 수행하였으며, 1차 검토 및 수정과정에서 과학교육전문가 1인과 현직 초등교사 5인이 기초 자료를 검토하고 수정하였다. 1차 수정 결과를 바탕으로 2001년 3월 중에 2주에 걸쳐서 현직 초등교사 1인이 2차 개발을 실시하였다. 2차 검토 및 수정 작업에서 과학교육전문가 1인과 현직 초등교사 10명이 1박 2일 동안 집중 작업을 통해서 2차 개발 결과에 대하여 토의하고 수정하는 과정을 거쳤다. 최종적으로 개발된 채점 기준은 다시 예비 적용 및 수정과정을 거쳤다. 2001년 5월 중 C교과대학교 과학교육과 4학년 학생들 8명을 2인 1조로 편성하여, 각 조가 4가지 채점 기준 중 한가지를 이용하여 실제 초등 학생들이 작성한 포트폴리오를 3개씩 채점하고 채점 기준의 문제점을 파악하며, 이를 보완하였다. 예비 채점에 참가한 교대 4학년들은 이미 예비 교사 교육을 3년 이상 받았을 뿐 아니라, 6개월 가량의 공동 연구 개발을 통하여 포

트폴리오 평가에 대한 기본적인 이론 및 양식 개발 경험을 가지고 있었다. 각 채점자들은 채점이 끝난 후에 채점하면서 힘든 점이나 해석하기 어려운 점 등을 자유롭게 토의하고 기록하는 시간을 가졌다. 이 과정을 통해 보다 올바른 채점을 할 수 있도록 채점 기준과 채점표를 수정·보완하였다. 이상의 과정을 거쳐서 최종적인 채점 기준이 완성되었다.

4. 학생 포트폴리오의 채점

완성된 채점 기준을 사용하여 학생 포트폴리오를 채점하였다. 채점은 2회 (1차와 2차) 실시하였다. 1차 채점은 2001년 5월 23일 오후 4시~밤 10시에 걸쳐서 실시되었다. 채점자는 교육 대학원에 재학 중인 초등 교사 12명이다. 채점자의 배경은 Table 1과 같이 성별과 경력, 4학년 경험, 포트폴리오 친숙도, 포트폴리오 실시경험 등에서 각 채점기준별로 유사하게 나타났다.

채점자들은 총괄-일반, 총괄-특수, 분석-일반, 분석-특수 채점기준별로 각각 3명씩 소집단을 이루어 각 소집단 내에서는 동일한 채점기준으로 동일한 학생 포트폴리오를 채점하였다(Table 2). '단원 2의 한 소단원 (1) 지층' 과, '단원 4의 (3) 열과 물체의 변화' 소단원의 10개의 소단원 목표에 대해서 상, 중, 하 수준의 포트폴리오 증거물을 12권씩 총 120여개의 목표를 채점하였다. 다만 분석-특수 채점기준의 경우에만 6권씩 총 60여개의 목표를 채점하였는데 그것은 분석-특수 채점기준에는 다른 채점기준에 없는 탐구영역 채

Table 1. The background of the scorers

| Rubric | No. of Scorers | Sex | | Teaching years | | 4th-gr. Experience | | Portfolio assessment | | | | | |
|--------|----------------|-----|---|----------------|----|--------------------|----|----------------------|---|-----|------------|----|---|
| | | m | f | ≥5 | <5 | yes | no | Familiarity | | | Experience | | |
| | | | | | | | | high | m | low | yes | no | |
| H-G | 3 | 1 | 2 | 2 | 1 | 1 | 2 | · | 3 | · | · | 3 | · |
| H-S | 3 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | · | · | 3 | · |
| A-G | 3 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | · | · | 1 | 2 |
| A-S | 3 | 1 | 2 | 1 | 2 | 1 | 2 | · | 3 | · | · | 2 | 1 |
| Total | 12 | 4 | 8 | 5 | 7 | 6 | 6 | 3 | 9 | · | · | 9 | 3 |

H-S: holistic-general, H-S: holistic-specific, A-G: analytic-general, A-S: analytic-specific

Table 2. The number of scorers and portfolios to be scored

| Rubric | Scorers | Objectives scored | Portfolios scored | Total no. of objectives scored |
|--------|---------|-------------------|-------------------|--------------------------------|
| H-G | 3 | 10 | 12 | 120 |
| H-S | 3 | 10 | 12 | 120 |
| A-G | 3 | 10 | 12 | 120 |
| A-S | 3 | 10 | 6 | 60 |

H-S: holistic-general, H-S: holistic-specific, A-G: analytic-general, A-S: analytic-specific

점 양식이 별도로 있어서 한 포트폴리오당 채점 분량이 더 많았기 때문이다. 따라서 채점자의 피로도를 고려하여 6권의 포트폴리오만을 채점하도록 하였다.

채점자들은 포트폴리오 채점에 대한 오리엔테이션을 통해 사전 훈련을 받은 뒤 채점에 참가하도록 하였다. 사전 훈련 내용은 연구에 대한 전반적인 소개, 채점 기준의 특성 및 내용 소개, 각 채점 기준을 이용한 채점 방법 등이며, 약 1시간 가량 실시되었다. 채점시 서로 토의나 의견 교환은 삼가도록 하여 채점자간 영향을 최소화하였다.

2차 채점은 채점시기별 신뢰도를 알아보기 위해 실시하였다. 2차 채점에서는 앞서 실시한 1차 채점에 참가한 채점자 중 각 채점기준별로 2명씩 총 8명의 채점자가 참여하였다. 2차 채점은 2001년 8월 8일 오후 4시~밤 9시 사이에 수행되었다. 채점 방법은 1차 채점 방법과 동일하게 하였고, 채점 대상 역시 동일한 포트폴리오를 사용하였다. 또한, 1차 채점과의 영향을 최소화하기 위해 약 두 달 반 가량 간격을 두고 실시

하였다.

실용도를 알아보기 위해서 각 채점자가 채점에 소요되는 시간을 기록하였다. 또한 연구자가 개발한 채점자용 설문지와 면담을 통하여 채점 과정에서 겪는 어려움을 파악하였다

5. 자료 입력 및 처리

각 채점기준을 이용한 채점 결과는 SPSSwin 통계 프로그램에 입력하였으며, 채점자간 및 채점자내 신뢰도를 위한 상관 계수를 구하였다. 또한 각 채점기준의 실용도를 분석하기 위하여 채점에 소요된 시간을 입력하여 평균 채점 소요 시간을 구하였다.

Ⅲ. 결과 및 논의

포트폴리오의 채점 기준과 신뢰도 및 실용도 검증은 목적으로 하는 이 연구 결과는 채점 기준, 채점자간

Table 3. Inter-rater reliability of each rubrics for learning objectives

| Rubric | Unit 2. (1) | | | | | Unit 4. (3) | | | | |
|--------|-------------|------|------|-------|------|-------------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| H-G | .587 | .596 | .835 | .744 | .800 | .506 | .501 | .414 | .787 | .619 |
| H-S | .538 | .616 | .126 | .709 | .710 | .475 | .696 | .283 | .285 | .601 |
| A-G | .665 | .547 | .194 | .790 | .769 | .389 | .667 | .801 | .612 | .695 |
| A-S | .966 | .675 | .635 | 1.000 | .465 | .768 | .881 | .637 | .708 | .804 |

H-S: holistic-general, H-S: holistic-specific, A-G: analytic-general, A-S: analytic-specific

및 채점 시기별 신뢰도, 실용도 순으로 제시하였다.

1. 포트폴리오의 채점 기준 개발

포트폴리오 채점 기준으로 총괄-일반, 총괄-특수, 분석-일반, 분석-특수의 4가지 유형을 개발하였다. 각 채점기준의 대표적인 예는 〈부록 1〉, 〈부록 2〉, 〈부록 3〉, 〈부록 4〉와 같다. 또한 채점에 도움을 줄 수 있는 채점양식을 각 채점기준별로 개발하였다(최미애, 2001). 모든 채점기준은 4가지 등급(뛰어난 수행, 충실한 수행, 부족한 수행, 잘못된 수행)으로 구성되어 있으며, 분석적 채점기준은 하위 영역을 지식·이해, 탐구, 태도, 실생활 관련의 4가지 영역으로 세분하고 있다. 이는 교육과정과 교과서를 분석을 바탕으로 설정한 것이다.

분석-특수 채점기준으로 채점을 할 때 탐구 영역에서는 탐구영역 채점표를 따로 사용하여 채점하도록 하였다. 채점 내용이 실험보고서와 같은 형태를 띠고 있어 실험과정을 분석할 수 있도록 탐구영역 채점표를 따로 개발하였다(최미애, 2001). 또한, 한 소단원의 채점이 끝나면 학생 개인별로 소단원에서 성취한 정도를 한 눈에 알아볼 수 있는 전체 채점 기록표를 개발하였다(최미애, 2001).

2. 채점자간 신뢰도

채점자간 신뢰도는 각 학습목표별로 제시하고, 이를 다시 종합하였다. 각 채점기준을 사용하여 한 소단원당 5개씩 총 10개의 학습 목표에 대한 증거물을 채점

한 결과를 이용하여 산출한 학습 목표별 채점자간 상관계수의 평균은 Table 3과 같다.

각 채점기준은 목표에 따라 .126에서 1.000 사이의 다양한 상관계수의 값을 나타낸다. 총괄-일반 채점기준은 신뢰도가 매우 높은 .80이상인 목표가 2개, 신뢰도가 높은 .6~.8 목표가 3개, 신뢰도가 있는 .4~.6에 5개로 나타났다. 총괄-특수 채점 기준은 신뢰도가 높은 목표가 5개, 신뢰도가 있는 목표 2개, 그리고 신뢰도가 낮은 목표가 3개로 나타났다. 분석-일반 채점기준은 신뢰도가 매우 높은 목표가 1개, 높은 목표가 6개, 신뢰도가 있는 목표가 1개, 그리고 신뢰도가 낮은 목표가 2개로 나타났다. 분석-특수 채점기준은 신뢰도가 매우 높은 목표가 4개, 높은 목표가 5개, 신뢰도가 있는 목표가 1개로 나타났다. 대체로 분석-특수 채점 기준의 채점자간 신뢰도가 높게 나타나며, 단원별 차이는 거의 없다.

각 채점기준별 채점자간 상관계수의 평균은 .525~.721로 채점자간 신뢰도가 높은 것으로 나타났다(Table 4). 분석-특수 채점기준의 채점자간 신뢰도가 가장 높고, 이어서 분석-일반, 총괄-일반, 총괄-특수의 순으로 상관계수가 높음을 알 수 있다.

Table 4. Inter-rater reliability for rubrics

| Rubric | M |
|--------|------|
| H-G | .636 |
| H-S | .525 |
| A-G | .643 |
| A-S | .721 |

H-S: holistic-general, H-S: holistic-specific, A-G: analytic-general, A-S: analytic-specific

분석적인 채점기준의 신뢰도를 영역별로 산출한 결과는 Table 5와 같다. 분석-일반 채점기준의 영역별 채점자간 상관계수의 평균은 .552~.670으로 신뢰도가 높은 것으로 나타났다. 태도 영역이 .670으로 가장 높았고, 지식·이해, 탐구, 실생활 관련 영역은 비슷한 수준을 보였다. 분석-일반 채점기준의 경우 채점자마다 목표에 관련된 하위 영역을 다르게 설정하는 경우가 많아서 신뢰도가 대체로 낮게 나타난 것으로 생각된다.

Table 5. Inter-rater reliability of sub-domains of analytic rubrics

| Sub-domain | A-G | A-S |
|------------------------|------|------|
| Memory · Comprehension | .553 | .618 |
| Inquiry | .552 | .685 |
| Attitude | .670 | .703 |
| Everyday life | .569 | .658 |

A-G: analytic-general, A-S: analytic-specific

분석-특수 채점기준의 영역별 신뢰도는 .618~.703으로 채점자간 신뢰도가 높은 것으로 나타났다. 지식·이해 영역이 .618로 가장 낮은 신뢰도를 보이며, 이어서 실생활 관련, 탐구, 태도 영역의 순으로 신뢰도가 높았다. 분석-특수 채점기준의 경우 학습 목표별로 하위 영역이 정해져 있었으며, 각 하위 영역별로 채점 기준이 제시되어 있어서 채점자들이 보다 명확한 판단을 할 수 있었던 것으로 보인다. 또한 탐구 과정에 대한 채점 양식이 신뢰도를 높이는데 기여한 것으로 보인다. 특히, 탐구영역의 채점자간 신뢰도가

높은 편에 속하며 분석-일반 채점기준의 탐구 영역 상관계수보다 약 .13이상 높은 상관계수를 나타내었는데 이것은 탐구영역 채점표가 채점자들의 일관된 채점을 도와주는 역할을 한 것으로 해석된다.

3. 채점시기별 신뢰도

채점 시기별 신뢰도를 학습 목표별로 Table 6에 제시하였다. 각 채점기준이 대부분의 목표에서 4이상의 상관계수를 보임으로써 채점자내 신뢰도가 있는 것으로 나타났다. 특히, '단원 2. (1)지층'의 목표 4)번과 5)번은 모든 채점기준에서 높은 상관계수를 나타내어 채점자내 신뢰도가 특히 높게 나타난 것을 볼 수 있다.

1차와 2차 채점 결과를 가지고 채점자들의 채점시기별 신뢰도를 알아보았다. 각 채점기준별로 2명의 채점자들이 같은 채점기준을 가지고 동일한 포트폴리오를 2회에 걸쳐 실시하였다. 각 채점기준의 채점시기별 신뢰도는 Table 7과 같다. 각 채점기준의 채점시기별 상관계수의 평균은 .557~.694로 채점자내 신뢰도가 있거나 높은 것으로 나타났다.

Table 7. Intra-rater reliability of rubrics

| Rubric | M |
|--------|------|
| H-G | .557 |
| H-S | .595 |
| A-G | .583 |
| A-S | .694 |

H-S: holistic-general, H-S: holistic-specific, A-G: analytic-general, A-S: analytic-specific

Table 6. Intra-rater reliability of rubrics for learning objectives

| Rubric | Unit 2. (1) | | | | | Unit 4. (3) | | | | |
|--------|-------------|------|------|------|------|-------------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| H-G | .645 | .558 | .203 | .808 | .540 | .616 | .583 | .434 | .580 | .601 |
| H-S | .420 | .654 | .594 | .772 | .609 | .430 | .717 | .531 | .438 | .610 |
| A-G | .377 | .328 | .342 | .768 | .655 | .494 | .733 | .840 | .620 | .563 |
| A-S | .839 | .718 | .524 | .917 | .599 | .812 | .829 | .554 | .502 | .550 |

H-S: holistic-general, H-S: holistic-specific, A-G: analytic-general, A-S: analytic-specific

채점자내 신뢰도 중에서 분석-특수 채점기준이 높은 값을 나타내었는데 그 이유로는 다음과 같은 점들을 들 수 있다. 첫째, 채점기준의 특성상 주관적인 판단을 많이 배제할 수 있었기 때문이다. 앞서 말한 바 있지만 분석-특수 채점기준의 경우 특수한 목표에 해당하는 내용을 분석적으로 구분하여 채점기준을 작성하였기 때문에 객관적이고 균형있는 채점을 할 수 있었다. 둘째, 채점에 사용한 채점표가 채점의 명료성을 더해 주었다. 채점기준자체가 세밀하므로 채점표 역시 세분화되어 작성되었으며 그뿐만 아니라 탐구영역 채점표를 따로 마련하여 채점하도록 하여 더욱 일치된 채점을 할 수 있었다. 교과와 특성상 실험활동이 많기 때문에 실험내용을 평가할 수 있는 탐구영역 채점표가 객관적인 채점에 도움이 되었다.

분석-일반 채점기준의 하위 영역별 채점시기별 신뢰도는 Table 8과 같다. 분석-일반 채점기준의 영역별 채점시기별 상관계수의 평균은 .541~.591로 채점자간 신뢰도가 있는 것으로 나타났다.

Table 8. Intra-rater reliability of sub-domains of analytic rubrics

| Sub-domain | A-G | A-S |
|------------------------|------|------|
| Memory · Comprehension | .562 | .668 |
| Inquiry | .549 | .767 |
| Attitudes | .591 | .698 |
| Everyday life | .541 | .526 |

A-G: analytic-general, A-S: analytic-specific

분석-특수 채점기준의 영역별 채점시기별 상관계수의 평균은 .526~.767로 채점자간 신뢰도가 있거나 높은 것으로 나타났다(Table 8). 탐구 영역의 상관계수의 평균은 .767로 채점자내 신뢰도가 높은 것으로 나타났다. 실생활 관련 영역이 .526으로 가장 낮았으며, 지식·이해와 태도 영역은 거의 비슷한 값을 보였다.

4. 채점기준의 실용도

채점기준의 실용도를 알아보기 위하여 채점하는데 걸리는 시간을 측정하였다(Table 9). 여기서 포트폴리오 한 권당 걸리는 시간은 본 연구에서 연구한 소단원 2개를 채점하는데 걸리는 시간을 뜻한다.

채점기준별로 채점시간에 약간의 차이가 있었는데 분석-특수 채점기준을 제외하고는 대부분 12~23분 정도가 소요되었다. 분석-특수 채점기준의 경우 채점표 이외에도 실험과정에 대한 탐구영역 채점표가 따로 있었기 때문에 포트폴리오 권수를 줄였음에도 불구하고 두 배 이상의 시간이 걸렸다. 그러나 그 외의 다른 채점기준들은 걸린 시간이 비슷하였다.

한편, 모든 채점기준에서 1차 채점보다 2차 채점을 할 때 걸린 시간이 짧았다. 그것은 1차 채점을 통해서 채점기준에 익숙해져 있으며, 또한 같은 포트폴리오를 가지고 채점했기 때문에 포트폴리오 양식도 익숙해졌기 때문이다. 따라서 더 잦은 채점을 실시할 경우 시간이 10분 이내로 단축될 수 있다고 본다. 따라

Table 9. Time required to score portfolios with each rubric

| Rubric | Trial | Total time for scoring (min.) (A) | No. of portfolios scored (B) | Scoring time per portfolio (min.) (A/B) | M (min.) |
|--------|-------|-----------------------------------|------------------------------|---|----------|
| H-G | 1st | 240 | 12 | 20 | 17 |
| | 2nd | 170 | 12 | 14 | |
| H-S | 1st | 270 | 12 | 23 | 18 |
| | 2nd | 145 | 12 | 12 | |
| A-G | 1st | 240 | 12 | 20 | 18 |
| | 2nd | 180 | 12 | 15 | |
| A-S | 1st | 300 | 6 | 50 | 43 |
| | 2nd | 210 | 6 | 35 | |

H-S: holistic-general, H-S: holistic-specific, A-G: analytic-general, A-S: analytic-specific

서 현장에서 포트폴리오를 평가할 때에도 채점기준만 준비되어 있다면 많은 노력과 시간을 들이지 않고도 채점할 수 있으리라 생각된다.

IV. 결론 및 토의

과학과 포트폴리오 평가나 수행평가는 채점자의 주관성이 많이 개입되어 객관성이 떨어지고 따라서 신뢰성이 낮다는 이유로 학부모와 학생은 불신감과 거부감을 가지고 교사는 부담을 느낀 측면이 있었다. 본 연구에서는 학생 과학 포트폴리오를 채점할 수 있는 다양하고 유용한 채점기준을 개발하고, 각 채점기준의 신뢰도를 검증하여 보았다. 이 연구 결과를 바탕으로 다음과 같은 결론을 내릴 수 있다.

첫째, 교사의 필요와 교실 상황에 따라 적절히 선택하여 활용할 수 있는 4가지 유형의 학생 과학 포트폴리오 채점기준을 개발하였다.

둘째, 총괄-일반 채점기준은 채점자간 신뢰도가 높은 편이고 채점자내 신뢰도가 있다.

셋째, 총괄-특수 채점기준은 채점자간 신뢰도와 채점자내 신뢰도가 있다.

넷째, 분석-일반 채점기준은 채점자간 신뢰도가 높고 채점자내 신뢰도가 있다.

다섯째, 분석-특수 채점기준은 채점자간 신뢰도와 채점자내 신뢰도가 높다.

여섯째, 총괄-일반 채점기준과 분석-일반적 채점기준의 경우, 개발된 기준으로 모든 포트폴리오 목표를 채점할 수 있으므로 매우 경제적이고 실용적이나, 여러 가지 다양한 목표를 충족시키기 위해서 일반적인 용어들을 사용했기 때문에 채점할 때 채점자들이 상당히 모호함을 느낀다.

일곱째, 총괄-특수 채점기준과 분석-특수 채점기준의 경우, 각각의 목표에 따른 채점기준이 있기 때문에 채점하기는 더 명확할 수 있으나, 목표별로 채점기준을 만들었기 때문에 양식이 다른 포트폴리오를 채점하기 위해서는 다시 채점기준을 만들어야 하는 번거로움이 있고 따라서 채점기준을 만드는데 시간이 많이 걸리는 단점이 있다.

여덟째, 채점기준의 실용도는 총괄-일반, 총괄-특수,

분석-일반 채점기준은 적절하였다. 그러나 분석-특수 채점기준은 다른 기준보다 2배 이상의 시간이 걸려 신뢰도는 높았으나 실용도는 낮았다.

본 연구 결과 포트폴리오 채점 기준의 신뢰도는 비교적 양호한 것으로 나타났다. 따라서 연구된 네 가지 채점 기준은 모두 활용이 가능하다. 활용하고자하는 목적과 상황에 따라서 적절한 채점 기준을 사용할 수 있을 것이다. 즉 대규모 학교에서 비교적 많은 학생들의 포트폴리오를 채점해야 하는 상황에서는 총괄-일반 채점 기준이 유용할 것이다. 교사가 시간이 비교적 많고, 채점 대상 학생의 수가 적거나, 학생들의 포트폴리오에 대해서 자세한 정보를 얻고 싶다면 분석-특수 채점 기준을 활용할 수 있다. 그 밖에도 총괄-특수, 분석-일반 채점 기준도 필요와 상황에 따라 선택하여 활용할 수 있다.

이 연구 결과로 나타난 신뢰도 지수는 비교적 양호한 값이기는 하지만 기존의 수행평가의 신뢰도 연구 결과와 비교할 때 상대적으로 낮은 값이다. 그 이유는 다음과 같은 점을 생각해 볼 수 있다. 첫째, 포트폴리오 평가가 가진 특성을 들 수 있다. 실험실습 평가 유형의 수행평가는 비교적 학생들이 수행하는 과정과 결과가 정형화되어 있으며, 따라서 교사들이 채점시 비교적 동질적인 판단이 가능하다. 여기에 비해서 포트폴리오에는 주어진 양식의 작성 이외에도 학생에 따라서는 다양한 증거를 포함시킬 수 있다. 즉 포트폴리오 평가는 실험실습 평가에 비해서 비정형화되어 있는 특성을 가진다.

둘째, 포트폴리오 평가는 수행 과제가 실험실습 평가와 비교할 때 분명하지 않다. 포트폴리오 평가는 교수 학습과 평가가 결합되어 있어서 학생들이 평가만을 목적으로 어떤 과제를 수행하는 경우와 상당한 차이를 보인다. 즉 학생들의 교실과 방과후 학습 활동의 종합적인 산물이 평가 대상이 되며, 따라서 수행 과제의 범위가 모호하다. 또한 학생들이 평가를 위한 수행보다는 학습을 위한 수행을 하기 때문에 평가를 위한 집중성이 낮다.

셋째, 채점자의 문제이다. 초등학교 교사는 시간적으로는 6개 학년의 전 교과를 지도하게 되어있다. 따라서 과학과 같은 특정 교과에 대한 전문성이 떨어진

다. 특정 학년 내용에 대한 심도있는 전문성 역시 낮다고 볼 수 있다. 또한 채점자 훈련 과정을 1회 밖에 제공하지 않은 것도 신뢰도 지수를 낮춘 요인으로 볼 수 있다.

신뢰도와 타당도는 서로 긴밀하게 관련되어 있으며, 신뢰도가 높아지면 타당도가 증진된다. 그 이유는 신뢰도가 증대하면 평가 대상의 실제 모습을 바르게 파악할 수 있는 가능성이 높아지기 때문이다. 따라서 포트폴리오 평가의 질을 높이기 위해서는 신뢰도와 타당도를 검증하고, 이를 향상시키려는 노력이 계속되어야 할 것이다.

적 요

본 연구의 목적은 초등학교 과학과 포트폴리오를 채점할 수 있는 다양한 채점기준을 개발하고, 개발된 각 채점기준의 신뢰도를 검증해 보고자 하는 것이다. 채점기준을 개발하기 위한 포트폴리오는 4학년 2학기 '단원 2. 지층과 화석', '단원 4. 열과 물체의 변화'를 중심으로 청주교대 과학교육 연구실에서 2000년 여름에 개발한 체제를 같은 해 가을, 경기도 중도시의 한 초등학교 4학년 한 학급에 적용하여 얻은 것이다.

총괄-일반, 총괄-특수, 분석-일반, 분석-특수의 4가지 채점기준을 개발하고, 각 채점기준에 근거하여 학생들이 작성한 포트폴리오 증거물을 채점하여 각 채점 기준별 채점자간 신뢰도와, 채점자내 신뢰도를 구하였다. 1차 채점에서는 총 12명의 채점자들이 각 채점기준별로 3명씩 그룹을 나누어 그룹당 12권의 포트폴리오 증거물을 채점하였다. 단, 분석-특수 채점기준의 경우 6권의 포트폴리오 증거물만을 채점하였다. 채점자내 신뢰도를 알아보기 위해 실시한 채점시기별 신뢰도에서는 1차 채점에 참가한 채점자 중 각 채점기준별로 2명씩 총 8명이 2차 채점에 참가하여 1차 채점과 동일한 방식으로 채점을 실시하였다.

채점결과를 SPSS 통계 프로그램에 입력하여 상관계수를 구한 결과, 총괄-일반 채점기준은 채점자간 신뢰도가 높고 채점자내 신뢰도가 있는 것으로 나타났고, 총괄-특수 채점기준은 채점자간 신뢰도와 채점자내 신뢰도가 있는 것으로 나타났다. 분석-일반 채

점기준은 채점자간 신뢰도가 높고 채점자내 신뢰도는 있는 것으로 나타났으며, 분석-특수 채점기준은 채점자간 신뢰도와 채점자내 신뢰도가 모두 높은 것으로 나타났다.

일반적인 채점기준들(총괄-일반, 분석-일반)의 경우, 하나의 채점 기준으로 모든 포트폴리오 목표를 채점할 수 있으므로 매우 경제적이고 실용적이나, 채점자들은 채점시 모호함을 느낀다고 하였다. 반면에, 특수적인 채점기준들(총괄-특수, 분석-특수)의 경우, 채점은 더 명확하게 할 수 있으나, 목표별로 채점기준을 개발해야 하므로 많은 시간과 노력이 필요하게 된다. 채점기준의 실용도 측면에서는 분석-특수 채점기준이 다른 기준보다 2배 이상의 시간이 걸려 실용도는 낮은 것으로 나타났다.

참 고 문 헌

- 교육부(1998). 수행평가의 이해. 서울: 교육부.
- 국립교육평가원(1996). 수행 평가의 이론과 실제. 국립교육평가원 : 서울.
- 김성숙(1995). 논술문항 채점의 변동요인 분석과 일반화가능도 계수의 최적화 조건. 교육평가연구, 8(1), 35-57.
- 김찬중(1999). 교과교육 학술세미나: 수행평가의 이론 및 현장 적용 사례-자연, 과학 교과-. 한국교육원대학교 부설 교과교육공동연구소.
- 김찬중, 김진규, 임형(2001). 과학과 수행평가의 이해와 활용. 서울: 경문사.
- 김찬중, 오영선(2001). 초등학교 자연과 상호작용 강화 학습일지의 학교 수준 적용 방안. 한국초등과학교육학회, 20(2), 187-196.
- 남명호(1995). 수행평가의 타당성 연구: 과학 실기 평가, 실험보고서 평가, 컴퓨터 시뮬레이션 평가의 비교. 고려대학교 대학원 박사학위 논문.
- 남현우(1998). 수행평가의 측정학적인 문제들에 관한 고찰. 교육평가연구, 11(2), 1-21.
- 성태제(1994). 논술형 고사와 예체능계 실기고사를 위한 채점자간 신뢰도 추정. 교육평가연구, 7(1), 43-56.

- 성태제(1999). 교육 평가 방법의 변화와 결과 타당도에 대한 고려. 교육학연구, 37(1), 197-218.
- 유선희(1998). 수행평가의 일반화가능도, 채점 결과의 신뢰도 및 지필검사와의 상관관계에 관한 연구. 교육평가연구, 11(2), 23-41.
- 최미애(2001). 초등학교 과학과 포트폴리오의 채점 기준개발과 신뢰도 검증. 청주교육대학교 교육대학원 석사학위 논문.
- 최연희, 권오남, 성태제(1998). 중학교 영어·수학 교과에서의 열린 교육을 위한 수행평가 적용 및 효과 분석 연구. 교육부 초등교육정책과 열린 교육 연구 과제 보고서.
- Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H.(1997). Generalizability analysis for performance assessment of student achievement for school effectiveness. *Educational and Psychological Measurement*, 57(3), 373-399.
- Fairbrother, R.(1997). Is this the right answer? *International Journal of Science Education*, 19(8), 887-894.
- Feldman, A., Kropf, A., Alibrandi, M.(1998). Grading with points: The determination of report card grades by high school science teachers. *School Science and Mathematics*, 98(3), 140-148.
- Haertel, E. H.(1999). Performance assessment and education reform. *Phi Delta Kappan*, 80(9), 662-666.
- Kulm, G. & Malcolm, S. M.(1992). *Science assessment in the service of reform*. Washington, D. C.: American Association for the Advancement of Science.
- Lederman, N. G., & Niess, M. L.(1999). Rubric's cube. *School Science and Mathematics*, 99(3), 113-115.
- Shavelson, R. J., Gao, X., & Baxter, G. P. (1996). On the content validity of performance assessments: Centrality of domain specification. In Birenbaum, M., & Dochy, F. J. R. C. (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge*, (pp. 131-141). Boston: Kluwer Academic.
- Shaw, J. M.(1997). Threats to the validity of science performance assessments for English language learner. *Journal of Research in Science Teaching*, 34(7), 721-743.
- Swartz, C. W., Hooer, S. R., Montgomery, H.J., & Wakelye, M. B.(1999). Using generalizability theory to estimate the reliability of writing scores derived from holistic and analytical scoring methods. *Educational and Psychological Measurement*, 59(3), 492-506.
- Third International Mathematics and Science Study(1994). *Performance assessment administration manual for the main survey*. Boston, Author.
- Wiggins, G.(1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.

〈부록 1〉 총괄 - 일반 채점기준

| 등 급 | 기 준 |
|---------------|------------------------------------|
| 4 (뛰어난 수행) | 증거들이 창의적이거나 기대수준 이상의 특성을 가짐 |
| 3 (충실한 수행) | 증거가 학습목표를 거의 대부분 또는 완전히 만족시킴 |
| 2 (부족한 수행) | 증거가 학습목표를 만족시키기에 불충분하지만 심각한 결함은 없음 |
| 1 (잘못된 수행) | 증거가 없거나 학습목표를 만족시키기에 심각한 결함이 있음 |

〈부록 2〉 총괄 - 특수 채점기준 [단원2-소단원 (1)-학습목표 1)]

| 단 원 | 2. 지층과 화석 | 소 단 원 | (1) 지층 |
|------|--|-------|--------|
| 학습목표 | 1) 여러 지층을 비교하여 각각 같은 점과 다른 점을 찾아 낼 수 있다. | 채점대상 | □ |

| 등 급 | 기 준 |
|---------------|--|
| 4 (뛰어난 수행) | 여러 지층을 비교하여 각각 같은 점과 다른 점을 찾아낸 내용들이 창의적이거나 기대수준 이상의 특성을 가짐 |
| 3 (충실한 수행) | 여러 지층을 비교하여 각각 같은 점과 다른 점을 찾아낸 내용들이 거의 대부분 또는 완전히 만족스러움 |
| 2 (부족한 수행) | 여러 지층을 비교하여 각각 같은 점과 다른 점을 찾아낸 내용들이 불충분하지만 심각한 결함은 없음 |
| 1 (잘못된 수행) | 여러 지층을 비교하여 각각 같은 점과 다른 점을 찾아낸 내용들이 없거나 심각한 결함이 있음 |

<부록 3> 분석 - 일반 채점기준

| 목표 : *목표에 해당되는 영역만 ○표하고 등급을 매기세요. | | | |
|--|---|---------------|-----------------------------------|
| 등급 \ 영역 | 지식·이해 () | 등급 \ 영역 | 탐구 () |
| 4 (뛰어난 수행) | 해당 지식을 충분히 이해하여 자신의 방식으로 표현함 | 4 (뛰어난 수행) | 해당 탐구요소의 사용이 바르고 숙달됨 |
| 3 (충실한 수행) | 해당 지식을 거의 대부분, 또는 전부 이해함 | 3 (충실한 수행) | 해당 탐구요소를 거의 대부분, 또는 전부 바르게 사용함 |
| 2 (부족한 수행) | 해당 지식을 어느 정도 알고는 있으나 완전하지 않음 | 2 (부족한 수행) | 해당 탐구요소를 어느 정도 사용하였으나 완전하지 않음 |
| 1 (잘못된 수행) | 해당 지식을 모르거나 잘못 이해함 | 1 (잘못된 수행) | 해당 탐구요소를 사용하지 않았거나 잘못 사용함 |
| 등급 \ 영역 | 태도 () | 등급 \ 영역 | 실생활 관련 () |
| 4 (뛰어난 수행) | 주어진 양식을 성실하게 해결할 뿐만 아니라 목표와 관련지어 더욱 자세히 조사함 | 4 (뛰어난 수행) | 학습목표를 해당 실생활 문제에 바르고 창의적으로 적용함 |
| 3 (충실한 수행) | 주어진 양식을 거의 대부분, 또는 전부 성실하게 해결함 | 3 (충실한 수행) | 학습목표를 해당 실생활 문제에 거의 대부분 바르게 적용함 |
| 2 (부족한 수행) | 주어진 양식의 절반 정도만 바르게 해결함 | 2 (부족한 수행) | 학습목표를 해당 실생활 문제에 적용하였으나 완전하지 않음 |
| 1 (잘못된 수행) | 주어진 양식을 안했거나 대부분 잘못 해결함 | 1 (잘못된 수행) | 학습목표를 해당 실생활 문제에 적용하지 않았거나 잘못 적용함 |

〈부록 4〉 분석 - 특수 채점기준 [단원4-소단원 (3)-학습목표 3]

| | | | |
|------|--------------------------------------|-------|------------------|
| 단 원 | 4. 열과 물체의 변화 | 소 단 원 | (3) 열에 의한 물체의 변화 |
| 학습목표 | 3) 열에 의한 기체의 부피변화를 실험을 통하여 확인할 수 있다. | 채점대상 | [4] |

| 영역 등급 | 지식·이해 | 영역 등급 | 탐구 |
|---------------|--|---------------|--|
| 4 (뛰어난 수행) | 열에 의한 기체의 부피 변화를 충분히 이해하여 자신의 방식으로 표현함 | 4 (뛰어난 수행) | 열에 의한 기체의 부피 변화에 대해서 바르고 창의적으로 실험함 |
| 3 (충실한 수행) | 열에 의한 기체의 부피 변화를 거의 대부분, 또는 전부 이해함 | 3 (충실한 수행) | 열에 의한 기체의 부피 변화에 대해서 바르게 실험함 |
| 2 (부족한 수행) | 열에 의한 기체의 부피 변화를 어느 정도 알고는 있으나 완전하지 않음 | 2 (부족한 수행) | 열에 의한 기체의 부피 변화에 대해서 실험하였으나 완전하지 않음 |
| 1 (잘못된 수행) | 열에 의한 기체의 부피 변화를 모르거나 잘못 이해함 | 1 (잘못된 수행) | 열에 의한 기체의 부피 변화에 대해서 실험하지 않았거나 잘못 실험함 |
| 영역 등급 | 태도 | 영역 등급 | 실생활 관련 |
| 4 (뛰어난 수행) | 포트폴리오를 성실하게 작성하고, 열에 의한 기체의 부피 변화와 관련지어 더욱 자세히 조사함 | 4 (뛰어난 수행) | 찌그러진 탁구공을 펴는 방법을 열에 의한 기체의 부피 변화와 관련지어 창의적으로 설명함 |
| 3 (충실한 수행) | 포트폴리오를 거의 대부분, 또는 전부 성실하게 작성함 | 3 (충실한 수행) | 찌그러진 탁구공을 펴는 방법을 열에 의한 기체의 부피 변화와 관련지어 바르게 설명함 |
| 2 (부족한 수행) | 포트폴리오를 절반 정도만 바르게 해결함 | 2 (부족한 수행) | 찌그러진 탁구공을 펴는 방법을 열에 의한 기체의 부피 변화와 관련지었으나 완전하지 않음 |
| 1 (잘못된 수행) | 포트폴리오를 안했거나 대부분 잘못 해결함 | 1 (잘못된 수행) | 찌그러진 탁구공을 펴는 방법을 열에 의한 기체의 부피 변화와 관련짓지 않았거나 잘못 설명함 |