

초등학생의 자유 탐구 활동 보고서의 평가 준거 개발 및 적용

김숙경 · 정진수[†] · 천재순[‡]

(서울등촌초등학교) · (대구대학교)[†] · (한국교원대학교)[‡]

Development and Application of Evaluation Criteria for Free Inquiry Activity Reports of Elementary School Students

Kim, Sook-kyoung · Jeoung, Jin-Su[†] · Chun, Jaesun[‡]

(Seoul Deungchon Elementary School) · (Daegu University)[†] · (Korea National University of Education)[‡]

ABSTRACT

The purpose of this study was to develop criteria for evaluating final reports of free inquiry activity and to apply developed evaluation criteria to free inquiry activity reports. 323 inquiry reports submitted to internal contests in the elementary schools as well as 189 award-winning ones from the inquiry competition for students in Seoul were collected. Nine categories and 18 sub-categories of evaluation criteria were derived from collecting reports analysis, literature reviews and interviews with 15 elementary school teachers. Criteria for each sub-category were organized into three steps and 12 elementary school teachers assigned scores for each sub-category, which were averaged. 132 including both award-winning and general reports were evaluated based on the developed evaluation criteria. The content validity and the reliability across scorers were significant. Furthermore, award-winning and general reports were significantly distinguished by the developed criteria in terms of all sub-category scores. The developed evaluation criteria will be an effective tool to assess student's free inquiry activity.

Key words : inquiry reports, free inquiry, evaluation criteria, elementary school student, elementary school teacher

I. 서 론

오늘날 학교 현장에서는 다양한 탐구 요소를 경험할 수 있는 진정한 과학 탐구 활동보다는 하나의 탐구 요소를 중심으로 한 간단한 실험이나 단혀 있는 지시적 탐구 활동을 제공하고 있다(윤혜경과 박승재, 2000). 그리고 PISA 2006 결과를 분석하면 우리나라 학생들의 과학 성취도는 대폭 하락하고 있으며, 과학에 대한 태도에 있어서도 과학과 관련된 자아 효능감이나 흥미 등이 낮은 것으로 나타났다(이양락, 2008). 또한 2007년 개정 과학과 교육과정이 추구하는 인간상 중의 하나가 창의적인 능력을 발휘하고, 새로운 가치를 창출하는 사람이다(교육

인적자원부, 2007). 이렇게 종합적인 탐구 과정을 경험하고, 학생들의 과학에 대한 흥미를 높이며, 창의적인 인간을 육성하기 위하여 2007년 개정 과학과 교육과정에서는 자유 탐구 활동을 도입하게 되었다.

자유 탐구 활동은 학생들이 탐구하고 싶은 주제를 스스로 선정하여 자기 주도적으로 심화된 탐구를 하는 것으로 3~6학년에서 1년 동안 최소 6시간을 수업하도록 되어 있다. 자유 탐구 활동과 관련된 선행 연구들을 살펴보면, 자유 탐구 활동의 긍정적인 교육적 효과들을 찾아볼 수 있다. 학생들이 자유 탐구 활동을 한 후 통합 탐구 능력이 높아졌으며(박종호 등, 2001), 학생의 탐구 능력을 기르기 위해서는 개

이 논문은 한국교원대학교 2009학년도 KNUE 학술연구비 지원을 받아 수행하였음.

2010.1.9(접수), 2010.2.10(1심 통과), 2010.2.20(2심 통과), 2010.2.21(최종 통과)

E-mail: chunjs@knu.ac.kr(천재순)

방향을 띤 탐구 활동을 하는 것이 효과적이라는 연구 결과가 있다(홍준의 등, 2007). 외국의 경우 자유 탐구 활동을 실시한 결과, 학생들이 지식을 더 이상 단편적으로 습득하지 않고 스스로 종합적인 탐구 과정을 하게 되었다는 연구 결과가 있다(Backus, 2005).

자유 탐구 활동이 학교 현장에서 긍정적인 효과를 얻기 위해서는 지도 방법과 평가 방안에 대한 안내가 반드시 필요하다(진순희와 장신희, 2007). 2007년 개정 과학과 교육과정에 자유 탐구 활동에 대한 학교에서의 운영 방안과 평가 방법이 제시되어 있으나, 자유 탐구 활동을 처음으로 지도해야 하는 교사들에게 도움이 될 수 있는 방안들이 구체적으로 나타나 있지 않다. 또한 학교 현장에서 과학 탐구 활동을 하기 어려운 근본적인 원인에는 교사에게 학생들의 탐구 활동을 가르칠 수 있는 구체적인 전략이 마련되어 있지 않으며(권용주 등, 2003a), 타당하고 객관적으로 평가하는 것 역시 쉽지 않은 것으로 여러 연구에서 나타나고 있다(주정은과 차희영, 2007; 김찬중과 최미애, 2002).

2010년부터 초등학교에 도입될 자유 탐구 활동을 지도하고 평가하는 초등학교 교사들에게 도움을 줄 수 있는 방법의 하나가 바로 평가 준거의 개발이다. 평가 준거는 교사가 학생들이 학습 활동을 하는 과정과 결과에서의 잘잘못을 파악하는데 도움을 주게 된다. 따라서 학생들이 학습 활동에 몰입하고 깊게 생각할 수 있도록 도와주고, 교사로서 하여금 학생에게 학습 활동의 결과에 대한 구체적인 내용의 피드백을 가능하게 해준다. 또한 평가 준거의 상위 수준은 교사가 학생에게 기대하는 결과를 나타내므로 교사에게 학습 활동의 목표와 지도 전략을 세우는데 도움을 주며, 학생에게는 학습 활동에 대한 이해를 높여 학습의 질을 향상시킬 수 있다. 그리고 준거가 마련되어 있으므로 객관적인 평가가 용이하고, 평가하는데 걸리는 시간을 단축시킬 수 있다는 장점이 있다(Andrade, 2000). 따라서 자유 탐구 활동의 최종 산물이자 타인과 의사소통의 역할을 담당하는 보고서를 타당하고 객관적으로 평가할 수 있는 준거에 관한 연구는 의미가 크다고 할 수 있다.

이 연구에서는 문헌 연구와 자유 탐구 활동의 성격을 지닌 서울학생탐구대회에서 입상한 보고서, 교내학생탐구대회에 제출된 일반 보고서를 수집, 분석하여 초등학생의 자유 탐구 활동 보고서의 평가

준거를 개발하고, 개발된 평가 준거를 학생들이 작성한 보고서에 적용하여 현장 적용 가능성을 알아보고자 한다.

II. 연구 방법

1. 평가 준거 개발 과정

1) 보고서 선정 및 수집

서울학생탐구발표대회의 목적은 3~6학년 학생들이 개개인의 흥미와 능력에 맞는 주제를 스스로 찾아 탐구하고, 이를 정리하여 발표하는 기회를 제공함으로써 탐구의 기쁨과 문제 해결 방법, 창의성 및 자기 주도적 학습력을 기르는데 있다(서울특별시과학전시관, 2007). 자유 탐구 활동 역시 학생이 스스로 주제를 찾고, 탐구하면서 과학에 대한 흥미를 높이고, 종합적인 탐구 활동을 경험하게 하고자 하는 것이 목적이므로 이 대회와 그 성격이 유사하다고 볼 수 있다. 따라서 서울학생탐구발표대회에서 2006~2008년 동안 입상한 탐구 보고서 189편과 이 대회에 참여하기 위해 서울에 있는 초등학교 두 곳에서 실시한 교내 학생탐구대회에 참여한 일반 탐구 보고서 323편을 수집하였다.

2) 자유 탐구 활동 보고서 평가 준거에 대한 교사 면담

초등학교 교사들이 자유 탐구 활동 보고서를 평가할 때 어떤 점에 중점을 두고 평가할 것인지에 대한 인식을 알아보기 위해 면담을 실시하였다. 면담 대상은 초등학교 교사들 중 한 번 이상 대회 진출의 경험이 있거나 탐구 보고서를 평가한 경험이 있고 연구 협조에 동의한 초등학교 교사 15명으로 교사의 배경은 구체적으로 표 1과 같다.

면담은 직접 대면 또는 전화 통화를 통하여 실시하였으며, “2010년에 처음 도입되는 자유 탐구 활동을 평가하는데 있어서 학생들의 최종 보고서를 평가하시게 됩니다. 선생님께서는 탐구 보고서를 평가하실 때, 어떤 점에 중점을 두시고 평가하시는 지 말씀해 주십시오.”라는 질문에 대한 응답을 듣고 전사하여 분석하였다.

3) 보고서의 탐구 유형 분류

대범주 및 하위 영역을 추출하기에 앞서 수집한

표 1. 평가 주안점에 대한 면담에 응답한 교사의 배경

단위 : 명(%)

	근무지			교육경력			심화전공		합계
	서울	경기	기타	1~5년	6~10년	11년 이상	과학	비과학	
남	1	1	0	0	1	1	2	0	2(13.3)
여	6	3	4	2	6	5	8	5	13(86.7)
합계	7(46.7)	4(26.7)	4(26.7)	2(13.3)	7(46.7)	6(40)	10(66.7)	5(33.3)	15(100)

탐구 보고서를 탐구 유형별로 분류하였다. 자유 탐구 활동의 경우, 주제와 더불어 탐구 방법 역시 학생들이 스스로 고안할 수 있기 때문에 다양하다. 따라서 탐구 방법에 따라 평가 준거가 달라질 수 있다는 판단 하에, 수집한 탐구 보고서 512편의 탐구 방법을 분석하여 발견-귀납형과 가설-연역형, 조사형, 혼합형의 네 종류의 탐구 유형으로 분류하였다. 발견-귀납형은 인위적으로 변인을 통제하지 않고 주어진 현상이나 상황에 따라 관찰, 측정 활동 등을 통하여 정보를 수집하는 유형을 의미한다(권용주 등, 2003b). 가설-연역형은 가설 또는 예상한 것을 검증하기 위하여 인위적으로 변인을 통제하고 조작하여 정보를 수집하는 유형을 의미한다(권용주 등, 2003b). 조사형은 어떤 사건이나 현상에 대하여 견학, 면담 및 설문 활동, 문헌 혹은 인터넷 조사를 통하여 정보를 수집하는 유형을 의미한다(교육인적자원부, 2005; Crawford, 2000). 마지막으로 혼합형은 위의 세 가지 탐구 유형 중 두 가지 방법을 이용하여 정보를 수집하는 유형을 의미한다.

었다. 보고서는 512편의 탐구 보고서 차례를 중심으로 분석하였다. 예를 들어 탐구 보고서가 주제와 결과로만 되어 있는 경우를 R1이라고 했을 때, 주제와 결과 이외에 탐구 동기가 첨가되어 있을 경우를 R2라고 명명하였다. 차례가 하나씩 늘어난 보고서를 발견할 때마다 유형을 하나씩 늘려가는 방법으로 보고서 유형을 분석하였다. 차례에 가장 많은 항목이 있는 보고서 유형을 토대로 대범주를 추출하였다.

문헌 분석은 국내에서 열리는 과학 관련 대회에서의 보고서 평가 관점과 탐구 및 실험 보고서 평가 준거를 개발한 연구 논문 및 문헌들을 살펴보았다.

추출된 하위 영역들에 대한 조작적 정의는 평가 준거를 평가자들에게 제시할 때 해당 하위 영역에서 어떤 점을 평가해야 하는지에 대한 평가 관점을 정확하게 이해할 수 있도록 도와주는 역할을 하게 되므로, 관련 문헌들을 중심으로 정의하였다.

표 2는 대범주 추출과 하위 영역 추출 및 조작적 정의를 내리기 위해 고찰한 문헌들을 나타내었다.

4) 대범주 추출, 하위 영역 추출 및 조작적 정의

대범주와 하위 영역을 추출하기 위해서 교사 면담 결과 외에 보고서, 문헌의 내용 분석을 실시하

5) 하위 영역의 수준을 평가하는 척도 기술

각 하위 영역에 해당하는 탐구 보고서의 수준에 대한 척도를 기술하기 위하여 조작적 정의를 내릴

표 2. 대범주 추출, 하위 영역 추출 및 조작적 정의를 위해 고찰한 문헌

평가 준거	문헌
국내 과학 관련 대회의 보고서	서울특별시과학전시관(2007), 한국과학교육단체총연합회(2008)
탐구 및 실험 보고서 평가 연구	강순희 등(1999), 전영석과 박종찬(2006), 홍준의 등(2007), Butler & McMunn(2005), Lawson(1995)
탐구 과정	2007년 개정 초등학교 과학과 교육과정 해설서(교육인적자원부, 2007), 과학 지식의 생성과 평가(권용주 등, 2008), 탐구과정평가표(허명, 1984), Klopfer의 과학 교육 목표 분류(김상달 등, 2005), SAPA II(AAAS, 1990)
창의성 개념 및 구성 요소	한국교육개발원(1991), Lubart(1994)
기타	강진명(1997), 교육과학기술부(2008), 교육인적자원부(2002, 2005), 권용주 등(2000), 권재술과 김범기(1994), 김선자와 최병순(2005), 김용신(2000), 박국태(2008), 백성혜와 박승재(1992), 성태제(2006), 신동훈(2007), 신미영과 최승언(2008), 우종욱 등(1998), Bell et al.(2003), King(1994), Scardamalia & Bereiter(1992), Wu & Krajcik(2006)

때 고찰한 선행 연구들을 토대로 수준에 대한 척도를 기술하였으며, 일부분은 보고서 내용을 정성적 분석법 중 분석적 귀납법(analytic induction)과 동료 심문법(peer debriefing)으로 분석하여 척도를 기술하였다(Stainback & Stainback, 1988). 또한 정량적 분석의 경우는 평균을 산출하여 그 결과를 토대로 과학 교육 전문가 및 동료 연구자와 함께 논의하여 기술하였다.

6) 점수 배점 의뢰 및 평가 준거표 구성

평가 준거의 점수 배점을 위하여 연구 협조에 동의하고 과학 교육을 전공한 초등학교 교사 12명에게 의뢰하였다. 의뢰에 응답한 교사의 배경은 구체적으로 표 3과 같다.

100점을 만점으로 대범주 및 하위 영역의 점수를 배점하도록 하였으며, 그 결과를 수합한 후 12명의 점수 배점을 각 영역별로 평균을 계산하였다. 평균의 소수 첫째 자리에서 반올림하여 각 영역의 가장 높은 수준에 배점하였으며, 아래 수준은 과학 교육 전문가 및 동료 연구자들과 논의하여 결정하였다. 이렇게 한 결과, 교사들이 중요하게 생각하는 영역에 대하여 가중치를 부여할 수 있었다.

보고서의 일반적 형식에 맞춰서 대범주를 나열하고, 그 대범주에 속하는 하위 영역과 각각의 영역에 대한 점수 및 그 수준 척도를 점검표 형식으로 일차 개발하였다.

7) 수정 및 보완, 어문 검토

과학 교육 전문가 및 동료 연구자와 함께 여러

차례의 수정과 보완의 과정을 거친 후에, 국어 교육 전문가 1인에게 평가 준거표의 문법과 맞춤법 등의 어문에 대하여 검토를 받아 완성하였다.

8) 내용 타당도와 채점자간 신뢰도, 현장 적용 가능성 분석

내용 타당도는 자유 탐구 활동 보고서를 평가하기 위한 대범주 및 하위 영역들의 선정이 타당한지, 각 하위 영역에 대한 조작적 정의 및 수준 척도, 점수 배점 등이 타당한지에 관한 내용으로 검증받았다. 내용타당도를 검증 받기 위하여 과학 교육 박사학위 소지자 2인과 과학 교육 석사학위 이상 소지자 7인에게 내용타당도를 검증받았다.

채점자간 신뢰도를 알아보기 위하여 일반 보고서 10편을 무작위로 추출한 후, 자유 탐구 활동 보고서의 평가 경험이 없고 연구 협조에 동의한 초등학교 교사 10명에게 의뢰하였다. 채점자간 신뢰도 검증에 참여한 교사의 구체적인 배경은 표 4와 같다.

연구자는 신뢰도 검증에 참여한 교사들에게 1시간 동안 연구의 목적과 평가 준거표의 구성, 하위 영역의 평가 관점, 평가 준거표 사용 방법을 설명하였다. 각 보고서를 채점할 때, 토의나 의견 교환을 하지 않도록 안내하였고, 독립적으로 평가하여 채점자간의 영향을 최소화하였다. 평가한 결과를 SPSS 12.0 통계 프로그램을 이용하여 총점을 비롯한 각 하위 영역별로 채점자간 신뢰도를 알아보았다.

현장 적용 가능성은 3가지 측면에서 살펴보았다. 우선 개발한 평가 준거가 현장에서 활용되기 위해서는 평가 시간이 적게 소요되어야 하므로, 채점자

표 3. 평가 준거의 점수 배점에 응답한 교사의 배경

단위 : 명(%)

	근무지			교육경력			과학 교육 학위			합계
	서울	경기	기타	1~5년	6~10년	11년 이상	학사	석사과정	박사과정	
남	1	1	0	0	1	1	0	2	0	2(16.7)
여	4	2	4	3	5	2	3	3	4	10(83.3)
합계	5(41.7)	3(25.0)	4(33.3)	3(25.0)	6(50.0)	3(25.0)	3(25.0)	5(41.7)	4(33.3)	12(100)

표 4. 채점자간 신뢰도 검증에 참여한 교사의 배경

단위 : 명(%)

	근무지			교육경력			심화전공		합계
	서울	경기	기타	1~5년	6~10년	11년 이상	과학	비과학	
남	1	0	0	0	0	1	0	1	1(10.0)
여	6	1	2	1	5	3	3	6	9(90.0)
합계	7(70.0)	1(10.0)	2(20.0)	1(10.0)	5(50.0)	4(40.0)	3(30.0)	7(70.0)	10(100)

간 신뢰도 검증에 참여한 평가자 10명이 보고서 10편을 평가하는데 소요된 시간을 측정하였다. 둘째, 평가를 마친 후에 평가 준거표의 장점과 단점, 평가하는데 어려웠던 점 등에 대하여 면담을 하였다. 면담은 평가가 끝난 직후에 수행하였으며, “평가하시면서 느끼셨던 어려움이라든지 장점이나 단점에 대해서 말씀해 주십시오.”라는 질문에 대한 응답을 듣고 전사하였다. 전사한 면담 내용은 계통도 분석법을 이용하여 분석하였다. 계통도 분석법이란 수집된 정성적 자료들이 갖는 특징을 범주화하고, 범주 사이의 관계를 기술한 후, 각 범주마다 나타난 자료의 빈도수를 나타내는 것으로 자료를 전체적으로 파악할 수 있도록 도와주며, 정성적 자료들을 정량적으로 분석할 수 있으므로 자료의 객관성을 갖게 된다(백성혜와 박승재, 1992). 셋째, 개발한 평가 준거표를 적용하여 입상 보고서와 일반 보고서가 각 하위 영역에서 실제로 차이가 나는지 알아보았다. 연구 대상은 2008년 서울학생탐구발표대회에서 입상한 보고서 4, 5, 6학년 각 22편과 학년별로 22편을 무작위 추출한 일반 보고서이다. 추출한 보고서를 평가한 후에 SPSS 12.0 통계 프로그램을 이용하여 두 집단 간 평균의 차이가 통계적으로 유의한지를 파악하기 위하여 두 독립표본 *t* 검증을 실시하였다. 만약 유의한 차이가 있다면 실제로 두 집단 간에 얼마나 차이가 있는지 알아보기 위하여 집단 간 차이의 크기(effect size)인 Cohen의 *d*값을 구하였다.

III. 연구 결과 및 논의

1. 대범주 추출과 하위 영역 추출 및 조작적 정의

대범주는 교사 면담, 수집한 보고서 분석, 국내 과학 관련 대회의 보고서 평가 관점, 탐구 및 실험 보고서 평가와 관련된 문헌 분석을 통해 추출하였다. 교사 면담 결과 주제, 탐구 방법, 결과, 결론, 노력성, 보고서 형식의 6가지 대범주가 추출되었으며, 수집한 탐구 보고서의 차례와 내용을 중심으로 분석한 결과, 주제, 탐구 동기, 탐구 문제, 이론적 배경, 탐구 방법, 가설, 결과, 결론, 발전적 사고, 참고 문헌의 10가지 대범주가 추출되었다. 국내에서 개최되는 과학과 관련된 대회의 보고서 평가 관점을 분석한 결과, 주제와 탐구 방법에 대한 평가 관점

은 모든 대회에서 제시되었으며, 그 밖에 가설, 결과, 결론, 이론적 배경, 발전적 사고, 노력성이 추출되었다. 교사 면담과 비교해 보면, 주제, 탐구 방법, 결과, 결론, 노력성 범주가 공통된 범주로 추출되었으나, 발전적 사고가 결론의 하위 영역이 아닌 하나의 대범주로 나타났다. 마지막으로 탐구 및 실험 보고서 평가와 관련된 문헌(강순희 등, 1999; 전영석과 박종찬, 2006; 홍준의 등, 2007; Butler & McMunn, 2005; Lawson, 1995)을 분석한 결과, 주제, 탐구 문제, 탐구 방법, 가설, 결과, 결론, 발전적 사고, 이론적 배경의 8가지의 범주가 나타났다. 일부 문헌들 중에 결론 안에 발전적 사고가 하위 영역으로 포함되어 있거나, 탐구 방법 안에 가설이 하위 영역으로 제시되어 있었다. 따라서 본 연구에서는 분석한 결과를 종합하여 평가 준거의 대범주로 주제 및 탐구 동기, 탐구 문제, 이론적 배경, 탐구 방법, 결과, 결론, 참고문헌, 노력성, 보고서 형식의 대범주 9가지를 추출하였다.

대범주에 따르는 하위 영역은 교사 면담 결과와 관련 문헌 분석을 통하여 추출하였다. 추출한 하위 영역은 다음과 같다.

1) 주제 및 탐구 동기에 대한 하위 영역

교사 면담과 국내 과학 관련 대회 평가 관점을 분석한 결과, 주제의 독창성과 목적의 명확성이 추출되었다. 또한 전영석과 박종찬(2006)은 주제 및 탐구 동기를 평가하는데 있어서 독창성, 명확성, 과학성의 3가지를 제시하였으며, 이 중에서 과학성은 탐구할 만한 가치가 있는 실생활에 유용한 주제인지를 평가하는 것으로 주제의 실용성이라고 할 수 있다. 그런데 창의성의 정의를 살펴보면 새롭고 실생활에 유용한 것을 생산하는 능력이므로(Lubart, 1994), 주제의 실용성은 독창성과 함께 평가할 수 있다. 따라서 주제 및 탐구 동기에 대한 하위 영역으로 명확성과 독창성 및 실용성을 추출하였다.

2) 탐구 문제에 대한 하위 영역

Butler와 McMunn(2005)은 탐구 문제의 주제에 대한 적합성과 형식을 제시하였다. 형식은 탐구 문제를 의문문의 형태로 제시하는 것을 뜻하는데, 반드시 의문문일 필요가 없이 ‘~가 되는지 혹은 ~인지 알고 싶다.’ 등과 같은 문장의 형식도 충분히 학생의 호기심을 나타내는 것이므로 탐구 문제라고

볼 수 있다. 또한 신동훈(2007)은 의문을 많이 생성할수록 탐구 과정에서 풍부한 결과와 결론을 얻을 수 있다고 진술하였다. 따라서 탐구 문제의 하위 영역으로 적합성과 유창성을 추출하였다.

3) 이론적 배경에 대한 하위 영역

박국태(2008)는 주제와 관련이 있는 기존 연구 결과에 관해서 논리적으로 진술하는 것으로 이론적 배경을 정의하고 있다. 이 정의로부터 이론적 배경의 하위 영역으로 적합성과 논리성을 추출할 수 있다. Butler와 McMunn(2005)은 주제에 대한 적합성을 하위 영역으로 제시하였다. 따라서 이론적 배경에 대한 하위 영역으로 이론적 배경의 논리적 진술을 포함한 주제에 대한 적합성을 하위 영역으로 추출하였다.

4) 탐구 방법에 대한 하위 영역

교사 면담 결과, 탐구 방법의 하위 영역으로 과학성, 적합성을 추출하였다. Klopfer(1971)의 과학 교육 목표에서는 탐구 방법의 적합한 선택, 적합한 과정 설계가 제시되었으며, 권용주 등(2008)의 과학 지식 생성 과정과 평가에서는 탐구 방법에 대한 하위 영역으로 설계의 다양성, 정교성, 탐구 방법의 유형 수준을 제시하였다. 강순희 등(1999)은 적합성, 탐구 과정의 적합한 배열을 탐구 방법의 하위 영역으로 제시하였고, 전영석과 박종찬(2006)은 적합성, 과학성, 안전 대책을 제시하였다. Butler & McMunn(2005)은 탐구 과정의 하위 영역으로 기구 및 재료의 제시, 재현성, 체계성과 안전 대책을 기술하였다. 여기서 유창성의 경우 얼마나 많은 탐구 방법을 설계하는지를 평가하는 것이다. 그러나 자유 탐구 활동의 최종 보고서의 경우, 설계한 것들 중에서 탐구 문제를 해결할 수 있는 가장 적합한 것을 선택하여 탐구를 수행하게 되어 있으므로(Klopfer, 1971), 최종 보고서 안에는 학생이 최종 선택한 설계만이 보고서에 진술하기 때문에 제외시켰다. 탐구 방법의 유형은 실험에 있어서 여섯 가지 단계로 나누어 단순한 관찰에서부터 비례비교법까지 실험의 수준을 나누어 점수를 부여한 것으로, 발견-귀납형에 속하는 관찰이 가장 낮은 수준으로 제시되어 있기 때문에 관찰 활동이 중요한 초등학교 실정(교육인적자원부, 2005)에 맞지 않다. 또한 체계성이나 안전 대책, 재현성은 탐구 과정을 체계적으로 자제

하게 기록하는 정교성에 포함시킬 수 있다. 따라서 이를 종합하여, 탐구 과정에서의 하위 영역으로 탐구 기구와 재료의 충족성, 탐구 방법의 독창성, 탐구 과정의 적합성, 정교성, 과학성의 5가지를 추출하였다.

또한 탐구 과정을 합리적이고 논리적으로 설계하였는지 평가하기 위하여 탐구 유형별로 하위 영역을 추출하였다. 발견-귀납형의 경우, 여러 문헌에서는 관찰의 다양성, 관찰 기구의 적합한 선택, 단위 사용과 정량화를 제시하였다(강순희 등, 1999; 교육과학기술부, 2008; 교육인적자원부, 2005; 전영석과 박종찬, 2006; 홍준의 등, 2007; AAAS, 1990; Butler & McMunn, 2005; Klopfer, 1971; Lawson, 1995). 또한 Bell 등(2003)은 인위적인 조작을 할 수 없는 유형이므로 주변 환경이나 상황을 자세하게 언급해야 한다고 하였다. 이 중에서 측정 단위는 올바른 측정 단위를 사용하였는지에 관하여 평가하는 것으로, 이는 탐구 과정이 아닌 결과를 제시할 때 나타나는 영역이므로 결과의 하위 영역으로 제시하였다. 또한 다양성의 경우 모든 감각기관을 사용해서 다양한 관찰을 하는 것을 의미한다. 그러나 세포 단위의 현미경 관찰의 경우, 시각적 관찰만이 가능하므로 모든 발견-귀납형의 탐구 과정에 해당하지 않으므로 제외시켰다. 따라서 관찰 및 측정 도구 선택의 적합성과 주변 환경의 제시를 발견-귀납형의 하위 영역으로 추출하였다.

가설-연역형의 경우, 김선자와 최병순(2005)은 변인 통제가 실험 설계에 중요한 역할을 한다고 보았다. 또한 국내 과학 관련 대회 평가 관점 및 여러 문헌에서는 가설 설정, 변인 통제의 2가지 하위 영역을 추출할 수 있었다(강순희 등, 1999; 교육과학기술부, 2008; 교육인적자원부, 2005; 전영석과 박종찬, 2006; 허명, 1984; 홍준의 등, 2007; AAAS, 1990; Butler & McMunn, 2005; Klopfer, 1971; Lawson, 1995). 따라서 가설 설정과 변인 통제를 가설-연역형의 하위 영역으로 추출하였다.

조사형의 경우, 성태제(2006)는 설문 연구를 위해서 연구 목적에 적합한 대상이나 장소를 선정하고 사전 연구를 실시하여 설문지를 완성한 후, 설문을 하도록 하고 있다. 그리고 설문 연구는 모든 연구 대상에게 동일한 도구를 사용하는 현상 자체에 대한 연구로, 현상과 원인 사이의 인과관계를 해석할 수 없는 단점이 있으므로 면담과 같은 질적 연

구 방법을 함께 사용할 것을 권장하고 있다. 김용신(2000)은 견학 활동에서 주제에 맞는 장소를 섭외하고, 학습이 될 수 있도록 사전 준비가 철저해야 한다고 주장하고 있다. 교육인적자원부(2005)는 면담을 하기 전에 면담의 주제에 적합한 대상을 선정하고, 질문이나 약속 장소 등의 사전 준비가 필요하다고 하였다. 사전 준비는 견학, 면담 및 설문 연구에 모두 해당되지만, 탐구 과정의 정교성에 포함될 수 있으므로 제외시켰다. 따라서 조사형의 하위 영역으로 대상, 장소의 적합성과 조사 방법의 다양성을 추출하였다.

5) 결과에 대한 하위 영역

교사 면담 결과에서는 결과의 정확성과 자료 변환을 추출하였으며, 국내 과학 관련 대회 평가 관점에서는 자료의 정확성, 자료 해석의 타당성을 추출하였다. 여러 문헌에서는 결과에 대한 하위 영역으로 자료의 변환과 결론을 얻을 수 있는 타당한 자료 해석 과정을 제시하였다(교육인적자원부, 2007; 전영석과 박종찬, 2006; 홍준의 등, 2007; AAAS, 1990; Klopfer, 1971; Wu & Krajcik, 2006). 허명(1984)은 자료의 정확한 기록, 자료의 변형, 해석 및 분석을 제시하였으며, 강순희 등(1999)은 결과의 정확성 외에 단위나 척도 사용의 정확성, 자료의 변환 여부, 자료의 해석을 제시하였다. Butler & McMunn (2005)은 정확한 자료 기록, 자료 변환, 정확한 단위, 그래프의 제목, 축, 간격과 같은 자료의 형식을 제시하였고, Lawson(1995)은 자료 변환, 그래프의 형식과 자료의 해석을 결과의 하위 영역으로 두었다. 이를 종합하여 기록의 정확성, 자료 변환의 제시와 형식, 자료 해석의 타당성을 결과에 대한 하위 영역으로 추출하였다.

6) 결론에 대한 하위 영역

교사 면담 결과로부터 발전적 사고의 유무, 결론의 타당성을 추출하였으며, 국내 과학 관련 대회 평가 관점으로부터는 실제적 응용의 발전적 사고, 결론의 타당성, 결론의 제시 여부를 추출하였다. 우리나라 교육과정(교육인적자원부, 2007)과 SAPA II (AAAS, 1990)에서 결론은 해석된 자료를 바탕으로 도출한다고 하였으며, 강순희 등(1991)은 결론의 타당성, 탐구의 문제점과 개선 방안을 제시하였다. 전영석과 박종찬(2006)과 홍준의 등(2007)은 결론의

타당성, 탐구 과정에 대한 한계 인식, 결론의 가치를 하위 영역으로 두고 있다. Butler와 McMunn(2005)은 결론의 타당성, 탐구의 약점을 인식하고 그에 대한 적절한 대안을 제안할 것을 제시하고 있다. Lawson (1995)은 결론의 타당성, 탐구의 개선 방안, 새로운 탐구의 제안을 제시하였다. 발전적 사고의 종류를 종합해 보면, 응용 방안, 탐구의 문제점과 개선점, 새로운 연구의 제안으로 볼 수 있다. 따라서 발전적 사고의 유무를 확장하여 다양한 발전적 사고를 할 수 있는가에 초점을 맞추어 발전적 사고의 다양성을 하위 영역으로 추출하였다. 이와 함께 결론의 타당성을 결론의 하위 영역으로 추출하였다.

7) 참고문헌에 대한 하위 영역

APA 작성법에서는 참고문헌이란 연구를 할 때 인용한 문헌들에 대한 정보를 제공하기 위하여 논문 말미에 일정한 양식으로 정리해 놓은 것으로 정의하였으며(강진령, 1997), 박국태(2008) 역시 참고문헌에 대하여 같은 정의를 내리고 있으며, 정리양식은 연구자마다 다르지만 체계성을 갖고 있다고 하였다. 따라서 참고문헌의 하위 영역으로 체계성을 추출하였다.

8) 노력성에 대한 하위 영역

교사 면담 결과로부터 자립성을 추출하였다. 국내 과학 관련 대회 평가 관점에서는 노력성에 대한 하위 영역으로 학생이 얼마나 노력을 기울였는가와 학생이 얼마나 주도적으로 탐구에 참여했는가, 탐구를 지속적으로 수행했느냐를 제시하였다. 탐구의 지속성은 학생이 얼마나 노력을 기울였는가에 포함될 수 있으므로, 노력성의 하위 영역으로 학생의 탐구에 대한 기여도와 자립성을 추출하였다.

9) 보고서 형식에 대한 하위 영역

교사 면담 결과 보고서의 형태를 갖춘 체계적 정리와 탐구 절차의 충실한 실행 여부를 추출하였다. 국내 과학 관련 대회 평가 관점에서는 탐구 과정에 따른 체계적인 보고서 작성을 추출하였다. 전영석과 박종찬(2006)과 홍준의 등(2007)은 보고서의 체계적이고 논리적인 구성을 제시하였다. 이를 종합하여 보고서 형식에 대한 하위 영역으로 탐구 절차에 따른 논리적 구성을 포함한 체계성을 추출하였다.

이와 같이 추출한 대범주에 따른 하위 영역은 모두 18가지로 각 하위 영역의 조작적 정의는 표 5와 같다.

2. 자유 탐구 활동 보고서의 평가 준거표

각 하위 영역의 수준별 척도 기술과 점수 배점 의뢰 결과를 바탕으로 점검표 형식으로 개발하였

표 5. 대범주에 따른 하위 영역의 조작적 정의

대범주	하위 영역	조작적 정의	
주제 및 동기	명확성	탐구 주제 및 동기에 탐구의 목적과 탐구하고자 하는 핵심적인 내용을 구체적으로 표현하는 것을 의미한다.	
	독창성 및 실용성	탐구 주제 및 동기 부분에 기존의 지식과 일치하지 않는 새로운 상황을 인지하거나 실생활에서 겪는 문제를 찾아 해결하고 탐구하고자 하는 것을 의미한다.	
탐구 문제	적합성 및 유창성	정해진 주제 영역 안에서 해결되어야 할 문제를 가능한 많이 발견하고 자신의 언어로 재구성하는 것을 의미한다.	
이론적 배경	적합성	탐구를 수행하는 데에 필요한 원리, 법칙, 이론 및 선행 연구의 결과 등을 조사한 후, 재구성해서 제시하는 것을 의미한다.	
	탐구 기구와 재료의 충족성	탐구에 필요한 기구와 재료들을 하나도 빠짐없이 모두 나열하는 것을 의미한다.	
탐구 방법	독창성	탐구 문제를 해결하기 위하여 기존의 탐구 방법에서 탈피하여 참신하고 독특한 탐구 방법을 설계하는 것을 의미한다.	
	적합성	탐구 문제를 해결하기 위하여 적합한 탐구 과정을 선택하는 것을 의미한다.	
	정교성	다른 사람이 탐구를 되풀이할 수 있을 정도로 탐구 과정이 체계적이고 상세하게 제시하는 것을 의미한다.	
	과학성	문제를 해결하는데 필요한 자료를 얻을 수 있도록 탐구 과정을 합리적이고 논리적으로 설계하는 것을 의미한다.	
	발견-귀납형	주변 환경의 제시	인위적으로 변인을 통제할 수 없는 대신 관찰 및 측정 대상의 채집 장소나 주변 환경을 구체적으로 기술하는 것을 의미한다.
		도구 선택의 적합성	정확한 관찰 및 측정을 위해 인간의 감각을 확장시키는 도구를 탐구의 목적에 맞게 선택하는 것을 의미한다.
	가설-연역형	가설 설정	탐구 문제에 대한 잠정적인 답을 설정하는 것으로 변인이 포함된 검증 가능한 가설을 세울 수 있는 능력을 의미한다.
		변인 통제	탐구에 영향을 주는 여러 조건을 확인하고 독립 변인과 종속 변인 이외의 다른 변인을 일정하게 고정하는 것을 의미한다.
	조사형	대상, 장소의 적합성	주제와 조사하는 목적에 적절한 대상 혹은 장소를 선정하는 것을 의미한다.
		조사 방법의 다양성	견학, 전문가와의 면담 및 설문과 같은 여러 가지 조사 방법을 이용하여 탐구하는 것을 의미한다.
결과	기록의 정확성	탐구를 통해 얻은 자료를 사실대로 정확하게 기록하는 것을 의미한다.	
	자료변환의 제시와 형식	탐구를 통해 얻은 자료를 탐구의 목적에 맞고, 이해하기 쉽도록 표나 그래프 등으로 변환하고, 변환된 자료를 올바른 형식으로 제시하는 것을 의미한다.	
	자료해석의 타당성	탐구를 통해 얻은 자료들을 바탕으로 의미 있는 결론을 얻을 수 있도록 자료들 사이의 관계성을 찾아내는 것을 의미한다.	
결론	타당성	탐구 결과에 근거하여 문제에 대한 해답을 얻거나 가설에 대한 옳고 그름의 판단을 내리는 것을 의미한다.	
	발전적 사고의 다양성	탐구에 대한 반성 및 대안이나 실생활에 대한 응용방안, 탐구와 관련된 새로운 탐구와 같은 현재의 탐구를 발전시킬 수 있는 방안을 다양하게 제시하는 것을 의미한다.	
참고 문헌	체계성	탐구하는 과정에서 인용하였거나 참고한 문헌을 다른 사람들이 손쉽게 찾을 수 있도록 일정한 양식으로 정리하여 제시하는 것을 의미한다.	
	탐구에 대한 기여도	탐구 과정 전반에 걸쳐 평가자가 인지 가능한 학생의 노력 정도를 의미한다.	
노력성	자립성	탐구를 할 때 다른 사람의 도움을 받지 않고 학생의 주도적인 탐구가 실행된 것을 의미한다.	
	체계성	탐구 과정에 따라 체계적으로 정리하고 작성하여 보고서의 형태를 갖춘 것을 의미한다.	

다. 완성된 평가 준거표는 부록에 제시하였다.

우선, 2007년 개정 과학과 교육과정을 토대로 직접 탐구 확인란을 마련하였다. 이는 다른 사람이 대신 탐구를 하거나 다른 사람이 한 탐구를 베껴서 제출하는 것을 방지하기 위하여 증빙 자료나 질문을 통해서 확인하는 것으로(교육인적자원부, 2007), 학생이 직접 탐구하지 않은 경우 0점 처리하도록 하였다.

각 하위 영역에 대한 수준은 모두 세 수준으로 기술하였으며, 점수 배점을 의뢰한 결과를 토대로 대범주와 하위 영역에 배점을 하였다. 가장 높은 점수 배점을 받은 대범주는 탐구 방법으로 가장 많은 하위 영역이 추출되었으며, 자유 탐구 활동이 결과보다는 과정을 강조하기 때문으로 보인다. 가장 낮은 배점을 보인 하위 영역은 탐구 기구와 재료의 충족성으로 지시적 탐구가 대부분인 학교 현장에서는 탐구 기구와 재료가 모두 일괄적으로 제공되기 때문에 탐구 기구와 재료를 나열하는 것에 대한 인식이 부족하기 때문인 것으로 보인다. 대범주에 따른 하위 영역의 배점표는 표 6과 같다.

평가자가 학생의 보고서를 평가한 후, 부족한 범

표 6. 대범주와 하위 영역의 배점표 단위: 점

대범주	배점	하위 영역	배점
주제 및 동기	11	명확성	5
		독창성 및 실용성	6
탐구 문제	6	적합성 및 유창성	6
이론적 배경	5	적합성	5
탐구 방법	36	탐구 기구와 재료의 충족성	3
		독창성	6
		탐구 과정	7
		과학성	14
결과	15	기록의 정확성	5
		자료 변환의 제시와 체계성	5
		자료 해석의 타당성	5
결론	10	타당성	6
		발전적 사고의 유창성	4
참고문헌	4	체계성	4
노력성	8	탐구에 대한 기여도	4
		자립성	4
보고서 형식	5	체계성	5
합계	100	합계	100

주에 대하여 피드백을 용이하게 할 수 있도록 각 범주별로 소계를 기록하는 공간을 마련하였다. 총평을 기록하는 공간은 교사의 전문적이고 융통적인 평가를 위하여 마련한 것으로 학생의 보고서에 나타난 특이한 점 등 중요한 평가 내용을 적도록 하였다(김은진 등, 2003).

탐구 과정의 과학성의 경우, 점수 칸이 둘로 나누어져 있는데, 혼합형의 탐구 보고서가 한 가지 유형으로 탐구한 것과 동일한 점수를 받을 수 있도록 과학 교육 전문가의 조언을 받아 혼합형의 유형별로 평가한 후 평균 점수를 내도록 고안하였다. 예를 들어 조사형으로 탐구한 후, 가설-연역형 탐구를 했을 경우, 각 유형별로 평가하여 점수를 합산한 후, 둘로 나누어 평균 점수를 구하도록 하였다.

3. 평가 준거의 타당도, 신뢰도 및 현장 적용 가능성 검증

1) 내용 타당도

각 하위 영역의 조작적 정의와 평가 준거에 대하여 과학 교육 전문가 9인에게 내용 타당도를 의뢰한 결과 조작적 정의의 내용 타당도는 4.807, 평가 준거는 4.686이었다. 이는 내용 타당도 지수의 일반적인 수락기준인 4.0(송영욱과 김범기, 2007) 이상이므로 높다고 할 수 있다.

2) 채점자간 신뢰도

평가 준거표 전체의 채점자간 신뢰도 평균은 0.662로 신뢰도가 있다는 기준에 속하므로(성태제, 2002), 개발한 평가 준거표는 채점자간 신뢰도가 있다고 볼 수 있다. 채점자간 신뢰도를 구한 결과는 표 7과 같다. 채점자간 신뢰도가 가장 높은 항목은 총점으로 0.894였으며, 가장 낮은 하위 영역은 '주제 및 동기' 범주 중 독창성 영역으로 0.423이었다. 주제 및 동기의 독창성의 채점자간 신뢰도를 높이기 위한 방안의 하나로 교사 면담 중에 어떤 주제들이 이전에 탐구되었는지 보충 자료가 필요하다는 의견이 있었다. 따라서 일반 보고서 323편에서 두 명 이상의 학생이 동일한 주제와 탐구 방법으로 탐구하여 유사한 내용을 지닌 보고서 제목을 표 8과 같이 나타냈다. 따라서 보충 자료의 활용도를 높이기 위해서는 매년 수정 및 보완해야 할 것이다.

표 7. 채점시간 신뢰도 결과

항목		채점시간 신뢰도 평균
총점		0.894
주제 및 동기	명확성	0.683
	독창성 및 실용성	0.423
탐구 문제	적합성 및 유창성	0.661
이론적 배경	적합성	0.831
	탐구 기구와 재료의 충족성	0.622
탐구 방법	독창성	0.538
	적합성	0.517
	정교성	0.621
	과학성	0.618
결과	기록의 정확성	0.647
	자료변환의 제시와 체계성	0.877
	자료해석의 타당성	0.601
결론	타당성	0.614
	발전적 사고의 유창성	0.828
참고문헌	체계성	0.890
노력성	탐구에 대한 기여도	0.539
	자립성	0.513
보고서 형식	체계성	0.616
	평균	0.662

3) 현장 적용 가능성

현장 적용 가능성을 알아보기 위하여 3측면에서 살펴보았다. 우선 본 연구에서 개발한 평가 준거표를 사용하여 탐구 보고서 10편을 평가하는데 걸리는 시간을 측정하여 표 9에 나타내었다. 초등학교 교사 10명이 10편의 보고서를 평가하기 위해 소요되는 평균 시간은 보고서 한 편당 5분 1초가 소요되었다. 한 학급을 30명이라고 가정했을 때, 평가에 소요되는 시간은 2시간 30분 30초가 된다. 이는 자유 탐구 활동 보고서가 1년의 결과물이라고 보았을 때 낮은 효율이라고 보기 어려우며, 30명 이상의 다인수 학급에서 일관성 있는 평가가 이루어질 수 있으므로 개발한 평가 준거표는 효과적이라고 볼 수 있다.

둘째, 평가가 끝난 직후 교사 10명의 면담을 실시한 후 전사하여 그 내용을 계통도 분석법을 이용하여 그림 1로 나타내었다. 교사 응답의 예는 다음과 같다.

“우선 이런 평가에 대한 경험이 부족한 사람들은 어려울 것 같아요. 이 평가 준거를 활용하는 방법을 익힐 수 있는 기회가 있으면 좋겠어요. 사실 저도 처음 평가할 때는 익숙하지 않아 힘들었거든요. 그렇지만 평가할 수

표 8. 일반 보고서 중 2명 이상이 동일하게 탐구한 보고서

단위 : 편

보고서 탐구 내용	보고서 수
환경오염의 종류에 따른 원인 및 결과를 인터넷으로 조사한 탐구	20
생선이나 닭의 뼈를 이용한 탄산음료가 우리 몸에 미치는 영향 탐구	8
여러 용액을 산성과 염기성을 구분하는 탐구, 매미의 종류와 생활 방식을 인터넷으로 조사한 탐구	7
주변에서 구할 수 있는 용액으로 얼렸다 녹는 시간을 측정하는 탐구	6
실온에서 껍질을 벗긴 사과와 갈변을 관찰하는 탐구, 인체의 생김새와 기능을 인터넷으로 조사한 탐구, 식빵을 이용하여 곰팡이가 자라는 조건을 알아보는 탐구	5
우유를 캔과 유리병에 넣어 캔 우유의 가능성을 확인하는 탐구, 견학을 통해 천연염색을 조사한 탐구, 알약을 여러 액체에 넣어 녹는 시간을 알아보는 탐구, 별자리 종류와 유래를 인터넷으로 조사한 탐구	4
꽃을 여러 음료수에 넣어 관찰하는 탐구, 여러 종이를 물에 넣어 흡수율을 측정하는 탐구, 여러 용액들이 어는 속도를 측정하는 탐구, 초란을 만들어 관찰하는 탐구, 진자 운동을 확인하는 탐구, 인터넷으로 조사한 탐구-지구 온난화의 원인과 결과, 장수풍뎅이의 생태, 휴연이 인체에 미치는 영향	3
여러 색의 수성 사인펜 번짐 현상 탐구, 종이를 여러 용액에 넣어 어린 후 관찰하는 탐구, 고구마 생장을 관찰하는 탐구, 잘 알려진 방법으로 음식물 쓰레기 냄새를 줄이는 방법을 확인하는 탐구	2
인터넷으로 조사한 탐구-개미의 종류와 생장, 중국산 매미의 생장과 없애는 방법, 모기한테 덜 물리는 방법, 태양계의 종류와 특징, 머리카락의 특징, 솟의 좋은 점, 일기 예보 절차와 우리 생활에 미치는 영향, 공룡의 종류와 특징, 수돗물 이리수의 정화 과정	

표 9. 평가 소요 시간

단위: 분.초

	교사 A	교사 B	교사 C	교사 D	교사 E	교사 F	교사 G	교사 H	교사 J	교사 J
총시간	38.12	40.05	40.57	49.42	40.34	47.02	48.36	59.41	69.42	67.34
평균	3.49	4.00	4.05	4.58	4.03	4.42	4.51	5.58	6.58	6.45

를 익숙해져서 빠르게 평가할 수 있었던 것 같아요. 또 객관적인 준거가 있어서 평가하는데 오히려 편했어요. 보고서에 어떠한 내용들이 들어가야 하는지 오늘 배워서 학생들을 가르칠 때 도움이 많이 되겠네요. 자유 탐구 활동이 도입되기 전에 이런 평가 준거가 있다는 것을 초등학교 교사들에게 알려주면 잘 활용될 수 있을 것 같아요.”

위의 경우, 긍정적인 면과 부정적인 면으로 범주화한 후, 다시 평가 준거로 인한 평가의 용이함, 평가 시간 단축, 경험이 부족한 교사들을 위한 연수 필요, 지도 방안 안내로 분류하였다. 교사 1명씩 차례로 내용을 분류하면서 새로운 내용이 나올 때마다 첨가하였으며, 같은 내용으로 응답한 교사 빈도수를 괄호 안에 나타내었다.

면담 결과, 평가 준거가 마련되어 있어 평가가 용이하였으며, 절대 평가가 가능하고, 평가하는 것과 함께 지도 방안을 안내받을 수 있으며, 학생의 탐구 보고서 수준을 높일 수 있다는 응답이 있었다. 평가 준거에 익숙해진다면 평가에 소요되는 시간이 단축될 수 있고, 평가의 근거가 마련되어 있어 객관적 평가가 가능하여 평가의 수준이 상승될 것이라고 응답하였다. 이는 Andrade(2000)가 제시한 평가 준거의 장점들과 일치하는 결과이다. 또한 진순희와 장신호(2007)의 연구에서 제안한 과학 탐구를 위한 교사 연수 프로그램이나 자유 탐구 활동의 지도 자료에 대한 지침서가 될 수 있을 것이다. 그러나 긍정적인 면과 함께 평가 준거의 수준이 어렵다는 지적이 있었으며, 학교에서 활용되기 위해서는

보고서의 평가 경험이 부족한 교사들을 위해 연수나 보충 자료가 필요하며, 평가 훈련이 필요할 것이라고 응답했다. 이는 평가 준거의 수준이 어렵지만 교사 연수 프로그램 자료로 활용할 경우 학교 현장에 적용이 가능함을 시사한다.

셋째, 개발한 평가 준거표를 적용하였을 때 입상 보고서와 일반 보고서의 차이가 실제로 있는지 알아보기 위하여 t 검증을 실시하였으며, 집단 간 차이의 크기를 보기 위하여 Cohen의 d값을 구하여 표 10에 나타내었다.

18가지 하위 영역과 총점에서 수준이 다른 입상 보고서와 일반 보고서 사이에서 유의한 차이를 보였으며, 집단 간 차이의 크기인 Cohen의 d값 역시 모든 영역에서 차이를 보였다. 이는 서울학생탐구 발표대회에서 보고서를 심사하는 과학 교육 전문가들에 의해 높은 평가 점수를 받은 입상 보고서가 본 연구에서 개발한 평가 준거표를 활용하여 일반 초등학교 교사들에 의해서도 높은 평가를 받은 것을 의미한다. 따라서 본 연구에서 개발한 평가 준거표는 학교 현장에서 초등학교 교사들이 초등학교생들의 자유 탐구 활동 보고서를 평가하는데 있어서 타당하고 유용한 도구가 될 수 있음을 시사한다.

IV. 결론 및 제언

이 연구는 2010년부터 초등학교에 처음으로 도입되는 자유 탐구 활동의 최종 보고서를 타당하고 객관적으로 평가하기 위한 준거를 개발하고, 개발한 평가 준거를 실제로 탐구 보고서 평가에 적용하여 현장 가능성을 분석하는 것을 목적으로 하였다.

교사 면담 결과, 문헌, 보고서 분석을 통하여 평가 준거를 구성할 대범주 9가지와 대범주에 따르는 하위 영역 18가지가 추출되었다. 이를 토대로 개발한 평가 준거의 내용 타당도, 채점자간 신뢰도를 분석한 결과, 타당도와 신뢰도가 있는 평가 준거가 개발되었다.

현장 적용 가능성 측면에서는 우선 평가하는데 소요된 시간은 자유 탐구 활동이 이루어지는 기간에 비례해서 생각하면 효율이 낮다고 하기는 어렵다. 또한 다인수 학급에서 일관성 있는 평가를 할 수 있다는 장점이 있으므로 개발한 자유 탐구 활동 보고서의 평가 준거를 활용한다면 교사들이 자유 탐구 학습을 평가하는데 도움을 줄 수 있을 것이다.

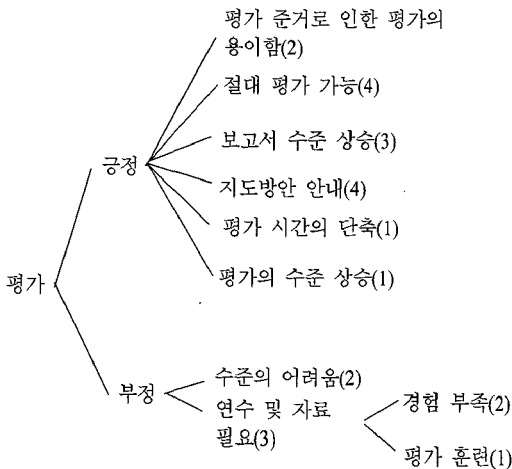


그림 1. 평가 후 교사 응답 분석.

표 10. 입상 보고서와 일반 보고서 평가의 t 검증 결과

n=132(각 66편)

대범주	하위 영역		평균	표준편차	t	p	Cohen's d	
주제 및 동기	명확성	입상	4.39	0.93	3.74	0.000	0.66	
		일반	3.58	1.52				
	독창성 및 실용성	입상	3.41	2.35	7.37	0.000	1.30	
		일반	0.82	1.63				
탐구 문제	적합성 및 유창성	입상	4.18	2.16	4.39	0.000	0.76	
		일반	2.55	2.12				
이론적 배경	적합성	입상	3.44	2.15	4.95	0.000	0.86	
		일반	1.56	2.21				
탐구 방법	탐구 기구와 재료의 충족성	입상	2.17	1.18	2.36	0.020	0.41	
		일반	1.67	1.24				
	독창성	입상	2.09	1.97	4.60	0.000	0.81	
		일반	0.73	1.40				
	적합성	입상	5.32	1.27	4.82	0.000	0.85	
		일반	4.09	1.63				
	정교성	입상	5.26	2.08	5.97	0.000	1.05	
		일반	3.27	1.72				
	과학성	입상	7.86	3.80	5.01	0.000	0.88	
		일반	4.83	3.10				
	결과	기록의 정확성	입상	4.24	1.10	2.87	0.005	0.50
			일반	3.64	1.32			
자료 변환의 제시와 형식		입상	3.97	1.21	7.85	0.000	1.42	
		일반	1.59	2.14				
자료 해석의 타당성		입상	4.58	1.07	5.29	0.000	0.98	
		일반	2.98	2.19				
결론	타당성	입상	5.09	1.39	6.67	0.000	1.17	
		일반	3.27	1.72				
	발전적 사고의 유창성	입상	1.45	1.14	6.00	0.000	1.05	
		일반	0.39	0.87				
참고 문헌	체계성	입상	2.42	1.34	8.51	0.000	1.48	
		일반	0.61	1.11				
노력성	탐구에 대한 기여도	입상	2.85	1.00	6.19	0.000	1.13	
		일반	1.98	0.54				
	자립성	입상	1.67	0.88	10.17	0.000	1.77	
		일반	3.33	1.00				
보고서 형식	체계성	입상	4.76	0.66	11.08	0.000	1.95	
		일반	3.24	0.90				
	총점	입상	69.36	10.44	12.30	0.000	2.15	
		일반	44.17	12.95				

그리고 평가 후 면담을 실시한 결과, 평가 준거가 마련되어 있기 때문에 평가가 용이하고, 객관적 평가 및 절대 평가가 가능하며, 지도 방안을 안내받

을 수 있다고 교사들은 응답했다. 또한 평가 준거를 근거로 보고서 작성을 지도하면 보고서의 수준이 높아질 수 있다고 하였다. 그러나 평가 준거의 수

준이 어렵고, 보고서 평가 경험이 적은 교사를 위해서, 그리고 학교 현장에 활용되기 위해서는 이 평가 준거에 대한 교사 연수가 필요하다는 지적이 있었다. 즉, 개발한 평가 준거의 긍정적인 측면이 부각되기 위해서는 평가 준거를 활용하기 위한 교사 연수가 수반되어야함을 시사한다. 마지막으로 개발한 평가 준거표를 입상 보고서와 일반 보고서에 적용하여 *t* 검증과 Cohen의 *d*값을 구한 결과 모든 하위 영역과 총점에서 유의한 차이를 보였다. 이는 개발한 평가 준거가 초등학생의 자유 탐구 활동 보고서를 평가 하는데 유용한 도구가 될 수 있음을 보여준다.

개발한 초등학생의 자유 탐구 활동의 보고서 평가 준거는 자유 탐구 활동 보고서에 대한 이상을 제시하여 주므로, 평가 준거에 제시된 항목들을 모두 만족시키는 것은 매우 어려울 것이다. 그러나 학교 현장에서 자유 탐구 활동을 지도하거나 평가 할 때 활용한다면 다음과 같은 교육적 시사점을 가질 수 있다.

첫째, 자유 탐구 활동에 대하여 초등학교 교사들에게 지도 방법과 평가에 관한 지침서로서의 역할을 할 수 있을 것이다.

둘째, 학생들이 자유 탐구 활동을 하는 과정에서 습득해야 할 능력을 제시함으로써 탐구 활동의 질을 높이는 역할을 할 수 있다.

셋째, 학생들과 학부모에게 자유 탐구 활동의 평가에 대한 타당성과 신뢰성 있는 준거를 제공할 수 있다.

넷째, 과학 교과와 다른 수행 평가와 자유 탐구에 관한 교사 연수 자료, 자유 탐구 활동을 연구하고자 하는 연구자들에게 기초 연구 자료로 활용될 수 있을 것이다.

참고문헌

강순희, 김양현, 박종윤(1999). 탐구적 일반화학실험 수행 평가 준거 개발. 한국과학교육학회지, 19(4), 507-515.
 강진령 편저(1997). APA 논문작성법. 서울: 양서원.
 교육과학기술부(2008). 과학 3-1 교사용 지도서(실험본). 교육인적자원부(2007). 2007년 개정 초등학교 과학과 교육과정 해설서.
 교육인적자원부(2005). 초등학교 교사용 지도서 과학 6-2.
 교육인적자원부(2002). 국어 6-2 초등학교 교사용 지도서.

권용주, 고경태, 정진수(2003a). 생물학 가설의 검증에서 연역적 과학지식의 구조와 생성 과정. 한국생물교육학회지, 31(3), 236-245.
 권용주, 양일호, 정원우(2000). 예비 과학교사들의 가설 창안 과정에 대한 탐색적 분석. 한국과학교육학회지, 20(1), 29-42.
 권용주, 정진수, 박윤복, 강민정(2003b). 선언적 과학 지식이 생성 과정에 대한 과학 철학적 연구 - 귀납적, 귀추적, 연역적 과정을 중심으로-. 한국과학교육학회지, 23(3), 215-228.
 권용주, 정진수, 이준기, 이일선(2008). 과학적 탐구사고력 향상을 위한 과학지식의 생성과 평가. 서울: 메이드.
 권재술, 김범기(1994). 초·중학생들의 과학탐구능력 측정도구의 개발. 한국과학교육학회지, 14(3), 251-264.
 김병하 역(1992). 질적 연구의 이해와 실천. 서울: 특수교육.[원전: Stainback, S., & Stainback, W. (1988). *Understanding & conducting qualitative research*. Virginia: The Council for Exceptional Children.]
 김선자, 최병순(2005). 변인 통제 문제해결 과정에서 나타난 초등학생의 실험설계 및 증거제시 특성. 한국과학교육학회지, 25(2), 111-121.
 김용신(2000). 경험과 구성 사회과 현장학습론. 서울: 문음사.
 김은진, 박현주, 강호감, 노석구(2003). 과학 수행 평가 문항의 선정 및 제작을 위한 평가 준거의 개발. 한국교육학회지, 23(1), 175-189.
 김찬중, 최미애(2002). 초등 과학과 포트폴리오의 채점 기준 개발과 신뢰도 검증. 한국과학교육학회지, 22(1), 176-280.
 박국태(2008). 과학교육논문작성법. 충북 청원: 한국교원대학교 출판부.
 박종호, 김재영, 배진호(2001). 자유탐구활동이 초등학생의 과학탐구능력과 과학적 태도에 미치는 영향. 초등과학교육, 20(2), 271-280.
 백성혜, 박승재(1992). 계통도 분석법을 이용한 실험보고서 평가의 효과 분석. 한국과학교육학회지, 12(1), 93-101.
 서울특별시과학전시관(2007). 선생님들을 위한 과학행사 운영 길잡이. 서울: 유일인쇄사.
 성태제(2002). 타당도와 신뢰도. 서울: 학지사.
 성태제(2006). 교육연구방법의 이해. 서울: 학지사.
 송영욱, 김범기(2007). 활동적인 문제해결 모형(SPPE) 개발 및 중학생들의 문제해결 활동에 대한 인식. 한국과학교육학회지, 27(4), 309-317.
 신동훈(2007). 과학 현상에 대한 초등학생들의 의문 유형과 초등 교사들의 설명유형과의 관계. 초등과학교육, 26(2), 149-160.
 신미영, 최승언(2008). 8학년 학생들의 탐구 보고서에 나

- 타난 과학방법의 특징. 한국지구과학회지, 29(4), 341-351.
- 우종욱, 김범기, 한안진, 허명(1998). 국가 수준의 과학탐구능력 평가체제 개발. 한국과학교육학회지, 18(4), 617-626.
- 윤혜경, 박승재(2000). 확장적 과학 탐구 활동을 통한 중학생의 탐구 동기 변화. 한국과학교육학회지, 20(1), 137-154.
- 이양락(2008). PISA 2006에서의 과학 성취도 조사 및 문항분석 결과. 현장과학교육, 2(1), 15-23.
- 전영석, 박종찬(2006). 고학 고등학교 학생의 물리 분야 개방적 탐구과제 수행 준비도 분석. 새물리, 52(4), 345-355.
- 주정은, 차희영(2007). 관찰에 의한 분류하기 탐구 능력 평가 준거 개발. 초등과학교육, 26(4), 407-417.
- 진순희, 장신호(2007). 과학 탐구에 대한 초등 교사들의 지도 경험. 초등과학교육, 26(2), 181-191.
- 한국과학교육단체총연합회(2008). 제16회 한국학생과학 탐구올림픽 종합보고서 <권 1>. 서울: 거목문화사.
- 한국교육개발원(1991). 사고력 신장을 위한 프로그램 개발 연구(V) (연구보고 RR 91-18). 서울: 한국교육개발원.
- 허명(1984). 과학 탐구 평가표의 개발. 한국과학교육학회지, 4(2), 57-63.
- 홍준의, 이인호, 전영석(2007). 초등학교 과학 영재 학생의 탐구 수행 능력 분석. 초등과학교육, 26(3), 267-275.
- AAAS (American Association for the Advancement of Science). (1990). *Science a process approach II*(pp. 26- 29). N.H.: Delta Education.
- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.
- Backus, L. (2005). A year without procedures: removing procedures from chemistry labs creates opportunities for student inquiry. *Science Teacher*, 72(7), 54-58.
- Bell, R. L., Blair, L. M., Crawford, B. A. & Lederman, N. G. (2003). Just do it? Impact of a science apprenticeship program on high school students' understandings of the nature of science and scientific inquiry. *Journal of Research in Science Teaching*, 40(5), 487-509.
- Butler, S. M. & McMunn, D. N. (2005). Gathering evidence in a variety of ways: Meaningful feedback. In C. Casbon & D. Nalley (Eds.), *How to assess student performance in science: Using classroom assessments to enhance learning* (pp. 48-50). NC: SERVE Center. (Eric Document Reproduction Service No. ED 498 182).
- Crawford, B. A. (2000). Embracing the essence of inquiry: new roles for science teachers. *Journal of Research in Science Teaching*, 37(9), 916-937.
- King, A. (1994). Guiding knowledge construction in the classroom: effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31(2), 338-368.
- Klopfer, L. E. (1971). Evaluation of learning in science. In B. S. Bloom, J. T. Hasting & G. F. Madaus (Eds.), *Handbook of formative and summative evaluation of student learning*. NY: McGraw-Hill. [김삼달, 이용섭, 최성봉 (2005). Klopfer의 교육목표 분류에 따른 제7차 교육과정의 중학교 과학 교육목표 분석 -7학년을 중심으로-. *Journal of Korean Earth Science Society*, 26(7), 640-651.]에서 재인용.
- Lawson, A. E. (1995). Student Assessment. In S. Horne (Ed.), *Science teaching and the development of thinking* (pp. 288-289). Belmont: Wadsworth Publishing Company.
- Lubart, T. I. (1994). Thinking and problem solving. In R. J. Sternberg (Ed.), *Handbook of perception and cognition* (2nd ed.), (pp. 289-323). San Diego, CA: Academic Press, Inc.
- Scardamalia, M. & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. *Cognition and Instruction*, 9(3), 177-199.
- Wu, H. & Krajcik, J. S. (2006). Inscriptional practices in two inquiry-based classrooms: A case study of seventh grader's use of data tables and graphs. *Journal of Research in Science Teaching*, 43(1), 63-95.

〈부록〉 초등학생 자유 탐구 활동 보고서 평가 준거표

- * 점수 부여 방법 : 각 하위 영역별로 평가한 후 해당되는 점수에 ○표 하시오.
- * 탐구 과정의 과학성은 탐구 유형을 먼저 파악하고, 혼합형인 경우 반복 평가한 후, 평균을 계산하여 옆 칸에 적으시오.
- * 총평 : 보고서의 특이사항이나 평가에 도움이 될 수 있는 의견이 있을 경우 적으시오.
- * 직접 탐구 확인 : 다른 사람이 대신 탐구를 하고 보고서를 썼거나 인터넷 자료를 배긴 경우인지 확인하여 직접 탐구를 했으면 ○표, 아니면 ×표를 한 후, ○표에 한해서 평가하시오(활동 사진, 방문 승인증, 관람표, 면담 내용 기록지, 탐구일지 등의 증거 자료, 탐구에 대한 질문 등을 이용하여 확인).

보고서 제목		직접 탐구 확인		
대범주 (배점)	하위 영역 (배점)	평가 준거	점수	소계
주제 및 동기 (11)	명확성 (5)	제시한 주제와 동기에 탐구의 목적과 내용이 명확하게 나타나 있다.	5	
		제시한 주제와 동기에 탐구의 목적과 내용이 나타나 있으나 명확하지 않다.	3	
		제시한 주제와 동기에 탐구의 목적과 내용이 나타나 있지 않다.	1	
	독창성 및 유용성 (6)	이전에 탐구되지 않은 새로운 주제이며, 유용성이 높은 주제이다.	6	
		이전에 탐구되지 않은 새로운 주제이나, 유용성이 낮은 주제이다.	3	
		이전에 탐구되었던 주제이며, 이전 탐구를 재확인하는 주제이다.	0	
탐구 문제 (6)	적합성 및 유창성 (6)	주제에 적합한 탐구문제를 3개 이상 제시하였다.	6	
		주제에 적합한 탐구문제를 1~2개 제시하였다.	3	
		주제에 적합한 탐구문제를 전혀 제시하지 못하였다.	0	
이론적 배경 (5)	적합성 (5)	제시된 이론적 자료가 탐구를 수행하는데 적합하다.	5	
		제시된 이론적 자료의 일부가 탐구를 수행하는데 적합하지 않거나, 내용이 부족하다.	3	
		탐구를 수행하는데 필요한 이론적 자료가 전혀 제시되어 있지 않다.	0	

대범주 (배점)	하위 영역 (배점)	평가 준거	점수	소계		
탐구 방법 (36)	탐구 기구와 재료의 충족성 (3)	탐구에 사용된 기구와 재료가 모두 제시되었다.	3			
		탐구에 사용된 기구와 재료가 일부만 제시되었다.	1			
		탐구에 사용된 기구와 재료가 전혀 제시되어 있지 않다.	0			
	독창성 (6)	탐구에 사용된 기구 및 재료, 탐구 과정이 이전 탐구에서 사용하지 않은 새로운 것이다.	6			
		탐구에 사용된 기구 및 재료나 탐구 과정 중 한 가지가 이전 탐구에서 사용하지 않은 새로운 것이다.	3			
		탐구에 사용된 기구 및 재료, 탐구 방법이 이전 탐구에서 사용했던 것으로 새롭지 않다.	0			
	적합성 (6)	탐구 과정이 문제를 해결하기에 매우 적합하다.	탐구 과정이 문제를 해결하기에 매우 적합하다.		6	
			탐구 과정이 문제를 해결하기에 일부 적합하지 않다.		3	
			탐구 과정이 문제를 해결하기에 전혀 적합하지 않다.		0	
		정교성 (7)	탐구 과정이 매우 상세하고 체계적으로 제시되어있다.		7	
			탐구 과정이 제시되어 있으나, 상세하지 못하거나 체계적이지 못하다.		4	
			탐구 과정이 전혀 제시되어 있지 않다.		0	
	발견 귀납 형	주변 환경의 제시	주변 환경을 매우 자세하게 제시하였다. (채집한 장소나 탐구 장소의 환경, 측정 위치 등)		7	
			주변 환경을 간단하게 제시하였다.		4	
			주변 환경을 전혀 제시하지 않았다.		0	
		도구 선택의 적합성	탐구를 위해 선택한 관찰이나 측정 도구가 적합하다.		7	
			탐구를 위해 선택한 관찰이나 측정 도구가 일부만 제시되었거나 적합하지 않다.		4	
			관찰이나 측정도구를 전혀 제시하지 못하였다.		0	
		가설 연역 형	가설 설정		가설을 제시하였다.	7
					예상을 제시하였다.	4
					가설 또는 예상을 전혀 제시하지 못하였다.	0
			변인 통제		변인통제를 매우 잘 하였다.	7
					변인통제를 일부만 하였다.	4
					변인통제를 전혀 하지 못하였다.	0
	조사 형	대상 장소의 적합성	조사 대상이나 장소가 탐구 목적에 매우 적합하다.		7	
			조사 대상이나 장소가 탐구 목적에 조금 부적합하다.		4	
			조사 대상이나 장소가 탐구 목적에 매우 부적합하다.		0	
		조사 방법의 다양성	2 가지 이상의 조사 방법을 사용하였다. (조사 방법 : 설문, 면담, 견학)		7	
			1 가지의 조사 방법을 사용하였다.		4	
			문헌 조사 방법만을 사용하였다.		1	
	탐구 과정 (14)	과학성 (14)				

대범주 (배점)	하위 영역 (배점)	평가 준거	점수	소계
결과 (15)	기록의 정확성 (5)	탐구한 결과를 정확하게 기록하였다.	5	
		탐구한 결과를 기록하였으나, 기록된 일부가 부정확하다.	3	
		탐구한 결과를 기록하였으나, 일부 사실이 아닌 주관적인 느낌을 기록하였다.	1	
	자료 변환의 제시와 형식 (5)	탐구를 통해 수집된 자료가 표나 그래프 등으로 제시되었으며, 제시된 자료의 형식이 모두 바르다.	5	
		탐구를 통해 수집된 자료가 표나 그래프 등으로 제시되었으나, 제시된 자료의 형식이 일부 바르지 못하다. 혹은 자료의 일부가 표나 그래프 등으로 제시되지 못하였다.	3	
		탐구를 통해 수집된 자료를 표나 그래프 등으로 전혀 제시하지 않았다.	0	
	자료 해석의 타당성 (5)	탐구를 통해 수집된 자료에 근거하여 자료를 해석하였다.	5	
		탐구를 통해 수집된 자료에 근거하지 않고 자료를 해석하였거나, 자료의 일부를 해석하지 못하였다.	3	
		탐구를 통해 수집된 자료에 대한 해석이 전혀 없다.	0	
결론 (10)	타당성 (6)	결과 내적인 근거에 의해 결론을 내렸다.	6	
		결론의 일부를 근거 없이 혹은 결과 외적인 근거에 의해 내렸거나 내리지 못하였다.	3	
		전혀 결론을 내리지 못하였다.	0	
	발전적 사고의 다양성 (4)	2종류 이상의 발전적 사고가 있다. (발전적 사고 : 응용방안, 문제점과 개선점, 새로운 연구의 제안)	4	
		1종류의 발전적 사고가 있다. 발전적 사고가 전혀 없다.	2 0	
참고 문헌 (4)	체계성 (4)	탐구에 이용된 참고문헌이 체계적으로 정리되었다.	4	
		탐구에 이용된 참고문헌이 제시되어 있으나 체계적으로 정리되지 않았다.	2	
		탐구에 이용된 참고문헌이 전혀 제시되어 있지 않다.	0	
노력성 (8)	탐구에 대한 기여도 (4)	탐구하는데 노력을 많이 기울였다.	4	
		탐구하는데 노력을 기울였다.	2	
		탐구하는데 노력을 아주 조금 기울였다.	1	
	자립성 (4)	다른 사람의 도움을 받지 않았다.	4	
		다른 사람의 도움을 조금 받았다.	2	
다른 사람의 도움을 많이 받았다.		1		
보고서 형식 (5)	체계성 (5)	탐구 보고서의 형태를 갖추었으며, 탐구 절차에 따라 체계적으로 정리하였다.	5	
		탐구 보고서의 형태를 갖추었으나, 탐구 절차의 일부가 없거나 체계적이지 못하다.	3	
		탐구 보고서의 형태를 전혀 갖추지 못하였다.	1	
총점				점
총평				