

중학교 3학년 학생들의 개방적 탐구에서 과학적 탐구과정기능에 대한 자기평가 수행 분석

황성원 · 김희경 · 유준희¹ · 박승재
(서울대학교) · ¹(한국교육과정평가원)

Ninth Graders' Self-Assessment of Scientific Process Skills in Open Investigation

Hwang, Sung-Won · Kim, Hee-Kyong · Yoo, June-Hee¹
· Pak, Sung-Jae

(Seoul National University) · ¹(Korea Institute of Curriculum and Evaluation)

ABSTRACT

Understanding the students' ability and tendency of self assessment is one of the important factors in helping their learning and in giving the appropriate feedbacks. The purposes of this study is to analyse the coincidency of the students' self-assessment with the teacher assessment, and the tendency of over and under estimation. The students performed open investigations and assessed their process skills by checking their investigation reports with a given checklist. The total mean of the 'Self-Assessment Capability(SAC)' was 0.73 by 1. The SAC was related to the achievement, not to the gender. The correlation between SAC and achievement was significantly positive($r=0.7$, $p<.01$), and for the girls, the tendency was remarkable. Among the five areas of process skills of open investigation, students got the highest SAC score in the managing and interpreting data area, and the lowest SAC in the drawing conclusion area. The mean of 'Self-Assessment Tendency(SAT)' for all items showed that students overestimated their process skills. While the high achievement students underestimated their process skills, the middle and low showed the tendency of overestimation.

Key words: Self-Assessment Capability, Self-Assesment Tendency, open investigation, scientific process skills

I. 연구의 배경과 목적

과학적 탐구과정기능을 평가하는 목적으로는 학생의 성취를 판단하여 외적으로 드러내기 위한 총괄적 목적과 학습의 어려움을 파악하여 되먹임(feedback)

을 통해 학생 자신의 학습을 돕는 형성적 목적이 있다. 특히 형성적 목적으로 이루어지는 평가는 과학적 탐구과정기능에 대한 학생의 이해 정도를 드러내어 교사에게는 교수 방법의 개선을 통한 적절한 되먹임의 기회를 제공하고, 학생에게는 효과적인 학습 전략

*2000년 12월 21일 받음.

을 가지게 하는 역할을 한다. 그러나 현재 학교에서 이루어지는 탐구과정기능의 평가가 형성적인 목적에 얼마나 기여하고 있는지는 의문이다. 최근 1급 정교사 연수에 참가한 교사들을 대상으로 한 김호진 등(2000)의 연구에 의하면, 과학적 탐구능력의 평가는 주로 실험보고서 평가에 의해 이루어지고 있으며(61%), 가장 많이 평가하는 탐구과정 기능이 자료해석 능력(86%)인 것으로 나타났다. 이는 교사 1인이 담당하는 학생의 수가 많고, 중등학교의 교과서에 제시된 실험의 목적과 방법이 대개 주어져 있기 때문이다. 이러한 평가가 가진 문제점은 첫째, 보고서의 실험 결과 해석 부분만 평가하기 때문에 탐구과정 기능 전반에 대한 학생의 이해 정도를 평가하기가 어렵고, 둘째, 평가의 기준을 알고 주도적으로 평가하는 사람이 교사이기 때문에 효과적인 학습을 위해 평가의 결과를 사용하는 책임을 교사가 지게 되며, 따라서 셋째, 학생들은 평가의 결과를 받아들이는 수동적인 존재가 되어 탐구과정기능에 대한 학습을 돕는 형성적 목적의 평가가 이루어지기 힘들다는 점이다.

탐구과정기능을 평가하기 위한 기존의 연구들은 평가의 영역을 대체로 문제인식 및 가설설정, 탐구설계, 탐구수행, 자료해석, 결론도출 및 평가로 구분한다(김미경 외, 1996; 우종욱 외, 1997; 우종욱 외, 1998). 과학적 탐구능력을 학습하는 것이 탐구의 전 영역에 걸쳐 탐구과정기능을 이해하고 수행하는 것이라면 이를 돕기 위한 형성평가의 대상도 탐구의 전 과정이 되어야 할 것이다. 이러한 의미에서 학생이 문제인식에서 결론도출의 전 과정을 직접 수행하는 개방적 탐구가 형성평가의 대상이 되는 것은 매우 중요하다.

형성평가가 효과적인 학습에 기여하려면 학생이 평가의 기준을 알고 평가의 결과를 능동적으로 활용할 것이 기대된다. 학생이 자신의 학습을 평가하는 자기평가는 이러한 목적에서 행해지는 평가이다. 최근 과학교육 연구자들 사이에서 자기평가의 중요성과 역할에 대한 관심이 증가하고 있는데, 블랙(Black, 1998)은 '학생들이 자신의 학습에 책임을 지는 것이 효과적인 학습을 의미한다면, 학생들은 자신의 평가에 포함되어야 한다'고 주장하면서 자기평가가 형성평가의 핵심적인 요소임을 강조하였다(Black & William,

1998). 자기 평가는 학생들이 자신의 약점을 드러내는데 두렵지 않은 환경을 제공하고, 독립심을 기르고, 교사의 평가 기준과 비교할 기회를 제공하며, 평가 자체가 학습의 자세한 기록이 되는 이점을 가지고 있다(Pottenger, 1990). 페어브라더 등(Fairbrother, et al., 1995)은 중등학생을 대상으로 한 연구에서 학생들에게 평가의 기준을 알게 하는 것이 학습 전략이 될 수 있다는 것을 보였으며, 따라서 학생들에게 자기를 평가하는 방법을 가르쳐야 한다고 지적하였다. 또한 타미르(Tamir, 1999), 슈테인과 밀러(Stein & Miller, 1998)는 자기평가를 돕는 도구를 개발하였다.

자기평가가 형성평가로서 효과적인 학습을 돕는 역할을 하려면 학생 스스로가 자기평가에 사용할 평가의 기준을 이해하고, 이에 비추어 믿음만하게 자기평가를 수행할 수 있어야 한다. 그리고 학생들에게 되먹임을 주어야 할 교사는 학생들이 스스로를 어떻게 평가하는지 알아야 한다. 일례로 대학생들을 대상으로 한 연구에 의하면 학생들은 교사에 비해 자신의 성취를 높게 평가하는 경향이 있었고, 특히 낮은 수준의 인지기능을 요하는 평가 문항에서는 학생과 교사의 상관도가 높았으며, 높은 수준의 인지기능을 요하는 문항에서는 상관도가 낮은 경향이 있었다(Zoller, et al., 1997; Zoller, 1999). 국내에서는 전통적으로 교사가 학생의 실험을 평가해왔기 때문에 탐구과정기능에 대한 자기평가에 대한 연구가 미흡한 실정이다.

본 연구의 목적은 중학교 3학년생들이 과학적 탐구과정기능에 대한 자기 평가를 얼마나 신뢰롭게 수행할 수 있는지 알아보는 것이다. 구체적인 연구 문제는 첫째, 학생들의 탐구과정기능에 대한 자기평가가 교사평가와 얼마나 일치하는지, 둘째, 학생들이 교사에 비해 자신의 탐구과정기능을 얼마나 높거나 낮게 평가하는지 성별, 성취도별, 탐구과정기능의 영역과 항목별로 알아보는 것이다.

II. 연구의 내용 및 방법

연구 대상은 서울에 소재한 중학교 3학년 남녀학생 63명(남학생 33명, 여학생 30명)이었다. 연구는 1999

년도 1학기 4차시에 걸쳐 이루어졌다. 교사는 학생들에게 [Table 1]과 같이 네 가지 개방적 탐구과제를 제시하였고, 학생들은 3명이 한 조가 되어 자신의 능력과 관심에 맞는 한가지 과제를 선택하였다. 학생들은 선택한 과제를 해결하기 위해 스스로 방법을 계획하고, 수행하고, 자료를 해석하여 결론을 도출하는 탐구의 전 과정을 진행하였으며, 이를 보고서로 정리하여 제출하였다. 학생들이 작성한 탐구 보고서는 학생들이 수행한 탐구과정의 기록이 되며, 따라서 이것이 자신의 탐구과정기능을 평가하는 자기평가의 대상이 되었다.

자기평가는 약 3차시에 걸친 개방적 탐구 활동이 끝난 뒤, 점검표를 이용하여 1차시 동안 이루어졌다. 자기평가에서는 평가의 기준을 교사와 학생이 공유하는 것이 중요하기 때문에, 자기평가의 과정은 학생들에게 평가의 기준을 명확히 제시하는 단계, 교사가 평가 기준을 어떻게 적용하는지 보여주는 단계, 학생들이 스스로 자신의 결과물에 평가 기준을 적용하도록 돕는 단계로 이루어진다(Adey, et al., 1996). 학생들이 자기평가라는 새로운 평가 방법에 익숙하지 않다는 점과 교사와 학생이 평가기준을 공유하는 것이 중요하다는 것을 고려하여, 본 연구를 수행하기 전에 총 4차시에 걸쳐 '우유펙 밀어내기'라는 주제로 개방적 탐구와 자기평가를 진행하였다.

점검표는 총 19개의 항목으로 이루어졌고, 탐구과정기능에 초점을 맞추었다(부록 1 참조). 각 항목에 대한 자기평가는 O, X로 이루어졌고, 학생들의 자기평가가 올바르게 이루어졌는지 점검하기 위해 학생들에게

각 항목별로 자신의 보고서에서 해당 내용을 찾아 기록하게 하였다. 점검표의 항목들은 어떤 과제를 선택했는가와 관계없이 변인 사이의 관계를 알아보는 정량적 탐구에서 반드시 고려해야 할 탐구과정기능을 나타내는 것이었고, 본 연구에서는 학생의 자기평가를 교사의 평가와 비교하는 것이 주요 목적이었으므로 네 가지 과제에 동일한 점검표를 적용하였다.

연구에서 얻은 결과를 분석하기 위한 방법은 다음과 같다. '자기평가총점(ST: Self-assessment Total)'은 학생이 자신의 보고서를 스스로 평가한 점수이며, 점검표의 각 항목에 대해 O를 1점, X를 0점으로 하여 총합을 구한 것이다. 학생별 '교사평가총점(TT: Teacher assessment Total)'은 두 명의 교사가 학생들이 자기평가에 사용한 점검표로 각각 채점을 한 뒤 항목별로 평군을 구하여 합한 총점이다. 두 명의 교사가 평가한 점수의 상관계수는 0.9로 나타났다. 탐구과정기능에 대한 학생의 수행 '성취도'는 교사의 평가를 기준으로 하였고, 따라서 '교사평가총점'이 '성취도'를 나타낸다고 보았다. '성취도'는 0에서 19점 사이의 값을 가진다. 항목별 '교사평가총점'은 각 항목별로 학생들의 교사평가점수의 평군을 구한 값으로 0에서 1사이의 값을 가진다.

'자기평가능력(SAC: Self-Assessment Capability)'은 탐구과정기능에 대한 학생의 자기평가가 교사의 평가와 얼마나 일치하는지 나타내는 값으로, 학생별 '자기평가능력'은 다음과 같은 방법으로 구한다. 여기서 S_i 는 학생, T_i 는 점검표의 항목을 나타낸다.

Table 1. Investigation tasks given to the students

Task	Content
1	Design the method to find out the rate of energy-loss when a free-falling ball is bounded. Compare the energy loss rate in various cases(the kinds of balls, number of bounding, etc).
2	Find out the energy efficiency of a rubber band by comparing elastic energy and transferred mechanical energy.
3	To keep hot water warm in a container, what would be the influencing factors? Find out them, and investigate the relationship in detail.
4	To keep some pieces of ice frozen in a container, what would be the influencing factors? Find out them, and investigate the relationship in detail.

$$(\text{자기평가능력})_i = 1 - \frac{\sum[(\text{자기평가점수})_i - (\text{교사평가점수})_i]}{\text{점점표 항목의 개수}}$$

학생별 '자기평가능력'은 0에서 1사이의 값을 가지며, 값이 1이면 교사의 평가와 학생 자신의 평가가 완전히 일치함을 뜻하고 값이 0이면 교사의 평가와 학생 자신의 평가가 완전히 불일치함을 뜻한다.

항목별 '자기평가능력'은 각 항목별로 학생의 자기평가가 교사의 평가와 얼마나 일치하는지 나타내는 것으로 다음과 같은 방법으로 구한다.

$$(\text{자기평가능력})_j = 1 - \frac{\sum[(\text{자기평가점수})_j - (\text{교사평가점수})_j]}{\text{전체 학생수}}$$

항목별 '자기평가능력'은 0에서 1사이의 값을 가지며, 값이 1인 항목은 교사와 학생의 평가가 완전히 일치하는 항목이며, 값이 0인 항목은 교사와 학생의 평가가 완전히 불일치하는 항목이다.

'자기평가경향(SAT: Self-Assessment Tendency)'은 학생들이 교사와 비교하여 자신을 얼마나 높게 혹은 낮게 평가하는지 나타내는 것으로 학생별 '자기평가경향'은 다음과 같이 구한다.

$$(\text{자기평가경향})_i = 1 - \frac{\sum[(\text{자기평가점수})_i - (\text{교사평가점수})_i]}{j}$$

학생별 '자기평가경향'은 -19에서 +19사이의 값을 가지며 '자기평가경향'이 (+)값인 학생은 교사보다 자신의 탐구능력을 높게 평가하는 학생이고, (-)값이면 교사보다 자신을 낮게 평가하는 학생이다.

또한 항목별 '자기평가경향'은 각 항목별로 학생들이 교사보다 높거나 낮게 평가하는 정도를 나타내는 값으로 다음과 같은 방법으로 구한다.

$$(\text{자기평가경향})_j = 1 - \frac{\sum[(\text{자기평가점수})_j - (\text{교사평가점수})_j]}{\text{전체 학생수}}$$

이 값은 -1에서 1 사이의 값을 가지며, (+)인 항목은 학생들이 교사보다 자신을 높게 평가하는 항목이고, (-)인 항목은 학생들이 교사보다 자신을 낮게 평가하

는 항목이다.

III. 연구 결과

1. '자기평가능력(SAC)' 분석

전체 학생의 '자기평가능력'은 평균 0.73으로 이는 학생의 자기평가와 교사평가의 일치도가 73%임을 의미한다. 성별과 성취도별로 구한 '자기평가능력' 값은 [Table 2]와 같다. 여학생이 남학생보다 '자기평가능력'이 조금 높았으나 유의한 차이를 나타내지 않았고($p > .01$), '성취도'에 따라 구분한 상·중·하 집단의 '자기평가능력'은 서로 유의한 차이가 있는 것으로 나타났다($p < .01$).

Table 2. Self-Assessment Capability(SAC) by the gender and the achievement level

	High	Middle	Low	Total
Boys	0.75	0.75	0.65	0.71
Girls	0.82	0.75	0.59	0.75
Total	0.80	0.75	0.63	0.73

학생의 '자기평가능력'과 '성취도'는 [Fig. 1]과 같은 분포를 나타내어($r = 0.5$, $p < .01$), 상관이 있는 것으로 나타났다. 분포의 모양으로 보아 상관이 아주 높지는 않으나, 대체로 '성취도'가 높은 학생일수록 '자기평가능력'도 높음을 알 수 있었다. 또한 이러한 상관관계는 여학생이 두드러졌다($r = 0.7$, $p < .01$).

[Table 3]과 같이 점점표의 각 영역과 항목별로 교사평가점수(TT)와 자기평가능력(SAC)을 구하였다. 예를 들어 1번 문항의 자기평가능력이 0.75인 것은 1번 문항에 대한 학생의 자기평가와 교사평가의 일치도가 75%임을 뜻한다. 다섯 가지 영역 중에서 '자기평가능력'이 가장 높았던 영역은 '결과정리 및 해석' 영역으로 12번 항목을 제외한 모든 항목의 값이 높았다. '자기평가능력'이 가장 낮았던 영역은 결론도출 영역이었다. 결과정리 및 해석 영역은 실제로 수행은 잘 하지 못했지만 스스로 못했다고 평가하여 '자기평

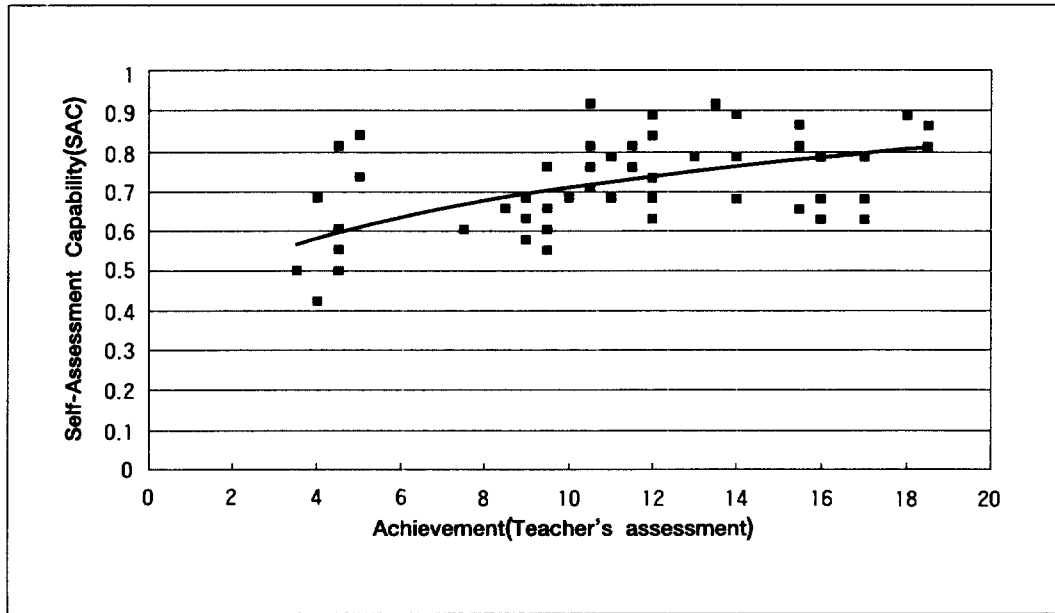


Fig. 1. The relationship between 'Achievement' and 'Self-Assessment Capability(SAC)'

가능력이 높았고, 결론 도출 영역은 가장 수행하지 못한 영역이었는데 교사에 비해 스스로를 높게 평가하여 자기평가를 가장 못하였다.

구체적으로 자기평가를 잘 한 항목 중에 '측정값을 표로 나타내었는가?(10번)', '측정을 하기 위한 도구를 올바르게 선택하였는가?(7번)'는 탐구수행과 자기평가를 모두 잘 한 항목들이고, '그래프를 그릴 때 평균값을 이용하여 그렸는가?(14번)', '그래프를 올바르게 그렸는가?(15번)'는 탐구수행은 잘 못했지만 자기평가를 잘 한 항목들이다. 비록 수행은 잘 못하였지만 이를 알고 바르게 평가하였다. 반면 자기평가를 잘 수행하지 못한 대표적인 항목들은 '영향을 주는 것 이외의 것을 일정하게 유지하였는가?(6번)', '탐구 결과에 비추어 맞는 결론을 내렸는가?(17번)', '탐구 방법 및 결과를 자신의 생각에 비추어 검토하였는가?(19번)'으로 변인통계 관련 항목과 결론 도출에 관련된 항목이었다.

2. '자기평가경향(SAT)' 분석

학생들이 교사에 비해 자신을 높게 혹은 낮게 평가하는 정도를 나타내는 '자기평가경향'은 최소값이 -6, 최대값이 +11이었고, [Table 4]에서 보듯이 전체 학생의 자기평가 경향은 약 1.0으로 평균적으로 교사보다 자신의 능력을 한 문항 정도 높게 평가하고 있었다. 교사평가총점을 기준으로 학생들을 상·중·하성취도 집단으로 나누어 분석한 결과, 탐구 수행을 잘하는 학생일수록 교사보다 자신의 탐구능력을 낮게 평가하고, 탐구수행을 잘 못하는 학생일수록 자신의 탐구능력을 교사보다 높게 평가한다는 것을 알 수 있었다. 이러한 상·중·하 집단에 따른 '자기평가 경향'의 차이는 통계적으로 유의한 것으로 나타났다 ($p < .01$). 남녀학생의 '자기평가경향'을 비교하였을 때 남학생 1.1, 여학생 1.0으로 모두 교사보다 자신의 능력을 높게 평가하고 있었고, 성별에 따른 '자기평가경향'은 유의한 차이가 없었다 ($p > .05$).

Table 3. Teacher assessment Total(TT), Self-Assessment Capability(SAC), Self-Assessment Tendency(SAT) by achievement levels for each checklist item

Area	Item	TT	SAC	SAT			
				High	Middle	Low	Total
Deciding Task	1	.74	.75	.00	.00	.64	.21
	2	.60	.68	-.24	.02	-.07	-.10
	Mean	.67	.72	-	-	-	-
Planning	3	.90	.84	.00	.00	-.38	-.13
	4	1.00	.81	-.10	-.19	-.29	-.19
	5	.76	.57	-.38	-.12	.17	-.11
	Mean	.89	.74	-	-	-	-
Performing Plan	6	.63	.44	-.55	-.19	.05	-.23
	7	.87	.89	.00	.00	.14	.05
	8	.66	.58	-.33	-.07	.52	.04
	9	.52	.87	.00	.24	.14	.13
	Mean	.67	.70	-	-	-	-
Managing and Interpreting Results	10	.90	.97	.00	-.05	.05	.00
	11	.46	.92	.00	.05	.19	.08
	12	.67	.55	-.40	.05	.05	-.10
	13	.47	.88	.14	-.05	.17	.09
	14	.37	.89	.00	.14	.19	.11
	15	.44	.92	.00	.00	.14	.05
Mean	.51	.85	-	-	-	-	
Drawing Conclusions	16	.28	.80	.02	-.02	.12	.04
	17	.43	.48	.62	.24	.48	.44
	18	.59	.59	.12	.33	.40	.29
	19	.20	.44	.17	.43	.52	.37
Mean	.40	.50	-	-	-	-	
Total Mean		.61	.73	-	-	-	-

Table 4. Students' Self-Assessment Tendency

	SAT
High-level	-.93
Middle-level	.81
Low-level	3.24
All	1.04

'자기평가경향'을 좀더 자세히 알아보기 위해 항목 별로 '자기평가경향'을 분석하였다. [Table 3]은 성취도별로 학생들의 항목별 자기평가경향을 보여주는데,

예를 들어 8번 문항에서 전체 학생의 자기평가경향이 0.04인 것은 평균적으로 볼 때 전체 학생의 4%가 교사보다 자신을 높게 평가했다는 뜻이다. 또한 평균적으로 상위권 학생의 33%가 교사보다 자신을 낮게 평가했고, 평균적으로 하위권 학생의 52%가 교사보다 자신을 높게 평가했다는 뜻이다.

전체 학생의 자기평가경향을 살펴보면, 교사에 비해 자기를 낮게 평가한 항목은 주로 탐구계획 영역의 항목들(3, 4, 5번)과 변인통제 항목(6번)이었고, 교사에 비해 자기를 높게 평가한 항목은 주로 결론도출 영역

의 항목(17, 19번)이었다. 학생들이 교사보다 자신을 높게 또는 낮게 평가한 이유를 알기 위해, 학생들이 자기평가를 하면서 해당 내용으로 기록한 것을 분석하였다. 변인통제와 관련된 6번 항목의 경우 교사보다 낮게 평가한 학생이 25명이었는데 이 학생들 중에는 변인통제를 하고도 '3가지 물건으로만 실험을 하였다'와 같이 독립변인을 달리해 가면서 여러종류의 실험을 해야한다고 설명하여, 항목의 내용을 잘못 이해하고 있는 경우가 있었고(28%), 해당 내용을 기록하지 않은 학생들이 많았다(44%). 해당 내용을 기록하지 않은 이유가 항목의 내용을 이해하지 못했기 때문이라고 해석한다면, 항목의 내용이 이해가 안 갈 때 학생들은 부정적으로 평가하는 경향이 있다고 볼 수 있다. 17번과 19번 항목은 교사보다 자신을 높게 평가한 학생이 각각 34명과 32명이었다. 17번 문항의 경우 이들 중 94%의 학생이 결과에 맞게 결론을 내렸다고 설명하여 학생들의 기대 수준이 교사보다 낮음을 알 수 있었고, 19번 문항의 경우에도 94%의 학생이 결과 및 결론은 자신의 생각에 비추어 검토하였다고 응답하여 결론 도출 단계에서 논의에 대한 기대 수준도 교사보다 낮음을 알 수 있었다.

성취도에 따라 자기평가경향이 다른 이유를 알아보기 위해 상·중·하 집단별로 각 항목별 '자기평가경향'을 비교하였다. 상위권 학생들이 교사에 비해 자신을 두드러지게 낮게 평가한 5번, 6번, 8번, 12번 항목에서 교사는 'O'라고 평가하였지만 학생은 'X'라고 평가한 사례를 뽑아 학생의 보고서와 자기평가 내용을 분석하였다. 5번 항목의 경우 과제 2번을 선택한 학생들은 고무줄의 탄성에너지가 위치에너지로 전환되는 것을 측정하기 위해 종이 조각을 튀겨 높이를 측정하는 방법을 사용했는데, 교사는 이 방법이 적절하다고 판단한 반면 학생들은 '적절하지 못한 것 같다. 고무줄을 놓았을 때 종이 가 튀어 올라간 높이를 측정할 때 순식간이기 때문에 정확하게 측정하지 못했다'고 평가하여 적절한 기대 수준이 교사에 비해 높음을 알 수 있었다. 6번 항목의 경우에도 '단지 탄성에너지만 측정하였을 뿐 다른 저항들은 생각하지 않았다'고 설명하여 교사보다 높은 기대 수준을 가지고 있는 학생도 있었고, 설명을 빈칸으로 둔 채 단지

'X'라고만 평가한 학생들도 있었다. 8번의 경우 5번 항목과 마찬가지로 '종이가 튀어 올라간 높이를 정확하게 잴 수 없었다'고 설명하여, 1cm 단위까지 읽은 경우 최대한 읽었다고 판단한 교사의 기준보다 학생들의 기대 수준이 더 높음을 알 수 있었다. 12번의 경우 교사는 교과서에 제시된 것과 같이 잘 정리된 표가 아니어도 단위와 변인이 표시되어 있으면 적절하다고 판단한 반면, 학생은 단위와 변인을 '정확히 기록하지 않았다'고 설명하였고, 설명을 빈칸으로 둔 채 단지 'X'라고만 평가한 학생들도 있었다. 전체적으로 볼 때 상위권 학생들은 일부 항목에 대해 교사보다 기대 수준이 더 높았다. 그리고 항목의 내용을 이해하지 못해서 자기평가 설명을 빈칸으로 두었다고 해석하면, 전체학생의 경향과 마찬가지로 상위권 학생들도 항목의 내용이 이해가 안 갈 때 부정적으로 평가하는 경향이 있음을 알 수 있다.

하위권 학생들이 교사에 비해 자신을 두드러지게 높게 평가한 1번, 8번, 17번, 19번 항목에서 교사가 'X'라고 했지만 학생이 'O'라고 평가한 사례를 분석하였다. 1번 항목의 경우 얼음을 오래 보관하는데 영향을 주는 요인을 찾는 탐구를 하면서 '영향을 주는 것: 차가운 물, 뜨거운 물, 햇빛, 그늘, 영향을 받는 것: 얼음'과 같이 자기평가 내용을 설명하여 '영향을 주는 것'이나 '영향을 받는 것'이라는 변인 개념이 명확하지 않음을 보였다. 8번 항목의 경우 많은 학생들이 소수 첫째자리까지 읽을 수 있는 메스실린더의 눈금을 일의 자리까지 읽고 '최대한 읽을 수 있는 데까지 읽었다'고 응답하였고, 17번과 18번 항목의 경우 교사는 결과에 바탕을 두고 알 수 있는 것만을 결론에 서술하는 것이 적절하다고 생각하는 반면, 학생들은 실험 결과로 알 수 없는 내용을 결과에 서술하고도 적절하다고 평가하였다. 전체적으로 볼 때 하위권 학생들은 일부 항목에 대해 교사보다 기대 수준이 더 낮았다.

IV. 결론 및 논의

중학교 3학년 학생들이 탐구과정기능에 대해 얼마나 신뢰롭게 자기평가를 수행할 수 있는지 알아보기 위해 '자기평가능력'과 '자기평가경향'을 분석한 결과

는 다음과 같이 정리할 수 있다.

학생들의 자기평가가 교사의 평가와 일치하는 정도는 평균 73%로 남녀간 '자기평가능력'은 차이가 없었다. '자기평가능력'은 '성취도'와 정적상관이 있었고, 이 관계는 여학생의 경우 두드러졌다. 점검표의 각 영역 중에서 '자기평가능력'이 높은 영역은 결과정리 및 해석 영역이었고, 가장 낮은 영역은 결론도출 영역이었다. 전체적으로 성취도가 높을수록 자기평가를 잘하였지만, 어떤 항목은 성취도가 낮음에도 불구하고 매우 높은 자기평가를 할 수 있었다.

학생들은 평균적으로 교사보다 자신의 탐구과정기능을 조금 높게 평가하고 있었다. 남녀차이는 없었으나 '성취도'에 따른 차이는 있었다. 상위권 학생들은 교사보다 자신을 낮게 평가하였고, 하위권으로 갈수록 높게 평가하는 경향이 있었다. 교사에 비해 자기를 낮게 평가한 항목은 주로 탐구계획 영역의 항목들과 변인통제 항목이었고, 교사에 비해 자기를 높게 평가한 항목은 주로 결론도출 영역의 항목이었다. 상위권 학생들은 교사에 비해 기대수준이 높기 때문에 자신을 낮게 평가하였고, 하위권 학생들은 교사에 비해 기대수준이 낮기 때문에 자신을 높게 평가하였다.

결론적으로 학생들은 자신의 탐구과정기능을 전반적으로 신뢰롭게 평가하였고 남녀차이는 없었다. 탐구과정기능의 항목별로는 대체로 결과정리 및 해석영역 항목에 대해 자기평가를 신뢰롭게 수행하였고, 변인통제와 결론도출영역의 자기평가를 신뢰롭게 수행하지 못했다. 상위권 학생들은 항목에 대한 기대수준이 교사에 비해 높아서 전반적으로 교사보다 자신을 낮게 평가했고, 이와 반대로 하위권 학생들은 항목에 대한 기대수준이 교사에 비해 낮아서 자신을 높게 평가했다.

학생이 자기평가를 잘 한다는 것은 교사의 평가 기준을 잘 이해하고 교사와 같은 기대수준을 가지고 있는 것이다. 학생들이 자신이 수행한 탐구에 대해 교사의 평가 기준을 잘 이해하고 신뢰롭게 적용하여 자기평가를 잘 수행하는 것은 평가가 형성적 목적으로 기능하기 위해 매우 중요하다. 본 연구를 수행하기 전에 이미 4차시에 걸쳐 개방적 탐구와 자기평가를 진행했음에도 불구하고 일부 항목에 대한 학생들

의 기대수준이 교사와 달랐다는 점은 평가 항목의 기준을 공유하는 것이 한 두 번의 연습으로 가능하지 않음을 시사한다. 자기평가능력이 높았던 결과정리 및 해석 항목은 전통적인 학교 실험에서 주로 강조하는 항목들이다. 학생들은 교사로부터 주로 결과정리 및 해석에 대해 평가를 받아왔을 것이고, 따라서 이 영역의 평가 기준을 교사와 공유할 수 있는 기회가 많았을 것이다. 하위권 학생들의 경우에는 교사에 비해 자신을 두드러지게 높게 평가하였고 대개 과제 설정과 결론 도출 영역의 항목들이 이에 해당했다. 이 영역의 항목들이 평소 학생들이 학교 실험에서 경험하지 못한 내용임을 고려한다면 탐구 수행을 통해 각 탐구과정기능이 무엇을 의미하는지 이해하도록 지도하는 것이 중요함을 시사한다.

본 연구에서는 탐구과정기능에 대한 학생들의 '자기평가능력'과 '자기평가경향'을 평가 항목과 학생의 성취도를 바탕으로 해석하였으나 앞으로는 그 외의 다른 요인들이 미치는 영향도 연구해야 할 것이다. 또한 평가기준에 대한 이해를 바탕으로 한 자기평가가 학생의 탐구과정기능 향상에 기여하는 형성평가의 역할을 얼마나 잘 할 수 있는지 계속 연구해야 할 것이다.

적 요

중학교 3학년 학생들이 과학적 탐구과정기능에 대한 자기평가를 신뢰롭게 할 수 있는지 알아보기 위해 학생의 평가가 교사의 평가와 얼마나 일치하는지, 교사에 비해 자신을 얼마나 높게 혹은 낮게 평가하는지 알아보았다. 학생들은 탐구과정 기능에 초점이 맞추어진 점검표를 이용하여 개방적 탐구 후에 작성한 보고서를 자기평가 하였다. 학생의 자기평가와 교사의 평가가 일치하는 정도는 평균 73%이었다. '자기평가능력'의 남녀간 차이는 없었으나, '성취도'를 기준으로 한 상·중·하 집단의 차이가 있었다. '자기평가능력'은 '성취도와 상관'이 있는 것으로 나타났고, 여학생의 경우 상관 관계가 더 높았다. 탐구과정기능의 영역 중에서 '자기평가능력'이 높은 영역은 결과정리 및 해석 영역이었고, 가장 낮은 영역은 결론도출영역

이었다. '자기평가경향'을 분석한 결과 전체 학생들은 평균적으로 교사보다 자신의 능력을 조금 높게 평가하였다. 탐구계획 영역의 항목들과 변인통제 항목은 교사에 비해 자신을 낮게 평가하였고, 결론도출 영역의 항목들은 교사보다 자신을 높게 평가하였다. 상위권 학생들은 교사보다 평가항목에 대한 기대 수준이 높아 자신을 낮게 평가하였고, 특히 하위권 학생들은 평가항목에 대한 기대수준이 교사보다 낮아 자신을 교사보다 높게 평가하는 경향이 두드러졌다.

참 고 문 헌

- 김호진, 광대오, 성민웅(2000). 중등학교 과학교사들의 학습평가에 관한 실태조사, 한국과학교육학회지, 20(1), 101-111.
- 김미경, 오희균, 박종원(1996). 물리 탐구 실험의 평가를 위한 도구의 개발과 분석, 한국과학교육학회지, 16(1), 51-60.
- 우종욱, 이항로, 김승훈(1997). 과학 실험 평가 도구 개발을 통한 탐구 능력 평가의 타당화에 관한 연구, 한국과학교육학회지, 17(1), 65-73.
- 우종욱, 김범기, 한안진, 허명(1998). 국가 수준의 과학탐구능력 평가체제 개발, 한국과학교육학회지, 18(4), 617-626.
- Adey, P., Fairbrother R., William D.(1996). *Learning Styles and Strategies*, OFSTED Reviews of Research, School of Education, King's College London.
- Black, P.(1998). *Testing: Friend or Foe? The Theory and Practice of Assessment and Testing*, The Falmer press.
- Balck, P. & Wiliam, D.(1998). *Inside the black box*, King's college London.
- Fairbrother, R., Black, P., Gill, P.(1995). *Teachers Assessing Pupils: Studies from Science Classrooms*, Hatfield.
- Pottenger, F. M.(1990). *The Testing-Grading Model of Teaching Contrasted with the FAST Model*, Curriculum Research and Development Group, University of Hawaii.
- Stein, M., Miller, D.(1998), Density Explorations, *The Science Teacher*, 65(2), 45-47.
- Tamir, P.(1999). Self-assessment: the use of self-report knowledge and opportunity to learn inventories, *International Journal of Science Education*, 21(4), 401-411.
- White, R. T. & Gunstone, R. F.(1989). Metalearning and conceptual change, *International Journal of Science Education*, 11(5), 577-586.
- Zoller, U.(1999). Scaling-Up of Higher-Order Cognitive Skills-Oriented College Chemistry Teaching: An Action-Oriented Research, *Journal of Research in Science Teaching*, 36(5), 583-596.
- Zoller, U., Tsaparlis G., Fatsow, M., & Lubezky, A.(1997). Student self-assessment of higher order cognitive skills in college science teaching, *Journal of College Science Teaching*, 27, 99-101.

부록 1. 탐구과정기능 자기평가를 위한 점검표

	내 용
탐구 문제를 정할 때	1. '영향을 주는 것' 과 '영향을 받는 것' 을 정하였는가? 2. 두 가지 모두 숫자로 나타낼 수 있는 것인가?
탐구 방법을 계획할 때	3. '영향을 주는 것' 의 값을 3가지 이상으로 다르게 하였는가? 4. '영향을 주는 것' 과 '영향을 받는 것' 의 값이 너무 크거나 작지 않게 하였가? 5. '영향을 받는 것' 을 측정하기 위한 방법이 적절한가?
탐구를 수행할 때	6. 공정한 실험을 하기 위해, '영향을 주는 것' 으로 선택한 것 이외에, 또 영향을 줄 수 있는 '다른 것' 들을 찾아서 일정하게 유지하였는가? 7. 측정을 하기 위한 도구를 올바르게 선택하였는가? 8. 측정 도구의 눈금을 읽을 때, 최대한 읽을 수 있는 데까지 읽었는가? 9. 같은 값을 두 번 이상 반복 측정하였는가?
탐구 결과를 정리하고 해석할 때	10. 측정값을 표로 나타내었는가? 11. 표를 만들 때, 여러 번 측정하여 얻은 값을 모두 기록하고 평균값도 기록하였는가? 12. 표에 '영향을 주는 것' 과 '영향을 받는 것' 을 표시하고, 단위를 기록하였는가? 13. 실험 결과를 적절한 종류의 그래프로 나타내었는가?(막대, 꺾은선, 선 그래프) 14. 그래프를 그릴 때, 평균값을 이용하여 그렸는가? 15. (막대 그래프) 측정값들을 크기에 맞게 막대로 그렸는가? (꺾은선 그래프) 측정값들을 올바르게 점으로 나타내고 꺾은선으로 연결했는가? (선 그래프) 점을 찍고 그래프를 그릴 때, 비례 관계인지, 반비례 관계인지를 생각하여 점들에 최대한 가까운 직선 또는 곡선으로 그렸는가? 점들의 분포가 비례 관계에 가깝다면 직선으로, 그렇지 않다면 점들의 모양에 가까운 곡선으로 그렸는가? 16. 그래프의 모양을 살펴보아 결과를 해석해서 관계를 설명하였는가?
결론을 내릴 때	17. 탐구 결과에 비추어 맞는 결론을 내렸는가? 18. 결론에 '영향을 주는 것' 과 '영향을 받는 것' 이 모두 나타나있는가? 19. 탐구 방법 및 탐구 결과를 자신의 생각에 비추어 검토하였는가? (자신의 생각과 차이가 났다면 그 이유를 생각해보거나, 차이가 나지 않았더라도 실험에 대해서 생각해보아야 할 점을 지적하였는가?)