



서술형 평가에서의 교사평가와 동료평가 활동 경험에 따른 고등학생들의 인식

백성혜*, 류혜정
한국교원대학교

High School Students' Perceptions on Descriptive Assessment Activity Experiences by Teacher or by Peer

Seoung Hye Paik*, Hye Jung Ryu
Korea National University of Education

ARTICLE INFO

Article history:

Received 19 August 2014
Received in revised form
31 August 2014
12 September 2014
Accepted 15 September 2014

Keywords:

peer assessment,
teacher assessment,
cognitive achievement,
affective achievement

ABSTRACT

The purpose of this study is to figure out the influence on the students' achievement of the cognitive and the affective domains by descriptive items assessment of teacher or peer and to obtain implications by analyzing the students' satisfaction and the reliability of the assessment. For this purpose, two classes of 11th grade students located in a small city were selected and took an exam related to the cognitive and the affective domains before and after the assessment. The assessment activities were carried out during 10 lessons and the teacher gave feedback to the students of the teacher assessment class in the lessons. In the peer assessment class, a small-size student group discussion and feedback were given to the students after the exam. The results show that higher level achievement group students represented relatively positive satisfaction on teacher assessment, and lower level achievement group students represented positive satisfaction on peer assessment. In spite of the same marker list, higher level achievement group students represent relatively high reliability than lower level achievement group students. The lower level achievement group students in the peer assessment class got statistically meaningful improvement of achievement than the students of teacher assessment class. The peer assessment activity was positively influenced on the affective domain of the lower level achievement group students, especially significant meaning of statistics was found on the students' perception about science.

1. 연구의 필요성 및 목적

현대 사회는 기존의 지식을 기억하고 재생하는 능력보다 학습자의 다양성을 존중하고 인성 및 창의성을 최대한 신장시키는 교육 활동을 강조하고 있으며, 이에 따라 지식의 획일성과 경직성을 극복할 수 있는 새로운 형태의 교육 평가 체제의 필요성이 나타나고 있다. 그 중 하나의 대안으로 제안되고 있는 서술형 평가는 단순한 지식의 암기에서 벗어나, 구체적인 부가 가치 창출능력, 수행 능력, 실천적 지능이 중요해진 현대 사회에서 주목받고 있는 평가 방법이라 할 수 있다(Yang, 2006).

서술형 평가는 문제 해결력 등 고등 사고 기능을 직접적으로 평가함으로써 궁극적으로 창의적인 능력을 신장시킬 수 있으며, 여러 측면에서 지식이나 능력을 지속적으로 평가함으로써 교수-학습의 결과뿐만 아니라 그 과정에 대한 자세한 정보의 수집, 학생 개개인의 특성과 수준을 지속적이면서도 종합적으로 평가할 수 있다는 특성을 가진다. 따라서 서술형 평가를 통해 학생이 인지적으로 '잘 아는 것' 뿐만 아니라, '아는 것을 실제로 적용할 수 있는지'에 대한 평가가 가능하다.

그러므로 서술형 평가는 현대사회의 변화에 따른 기존의 암기식, 주입식 위주의 학교수업 방법을 개선하는 한편, 다양한 문제 상황에 능동적으로 대응하고 새로운 방식으로 문제를 해결하는 능력을 갖춘

인재를 육성하기 위한 목적에서 도입되어 그 중요성이 점차 커지고 있으며, 단계적으로 확대 실시되고 있는 추세이다(Kim, 2010). 이에 발맞추어 많은 선행연구에서 탐구과정을 측정하는 서술형 평가 도구를 개발하거나(Ha, 2000; Hwang, 2010; Kim & Kim, 2012), 서술형 평가 자료를 개발하였다(Hong, M. & Chung, H., 2006). 그러나 서술형 평가의 경우 채점의 공정성을 유지하기 위하여 많은 시간이 걸리고, 평가자의 주관성이 개입되는 등의 문제가 발생할 수 있다(Kim & Yoo, 2006).

모든 평가에서 가장 중요한 부분은 공정한 평가와 평가 결과에 대한 피드백이다. 이런 점을 고려할 때 학습자의 입장에서 수동적으로 주어진 평가 결과를 받아들이기만 하는 것이 아니라 적극적으로 평가에 참여하여 주도적인 학습 자세를 갖출 필요가 있다. Ballantyne 등(2002), Blom과 Poole(2004), Falchikov(1986) 등은 동료평가가 학습의 주인의식을 제공하여 자아 존중감과 자기효능감, 학습 동기에 영향을 미친다고 하였다. Defina(1992)와 Green(1994), Hahn과 Oh(2008) 등은 동료평가의 본질과 역할을 확대시킬 수 있으며, 평가 과정을 통해 평가를 검토하는 기회를 제공해 주며 서로 피드백을 주는 과정을 통해 학습이 강화된다고 하였다. 또한 Cho(2004)는 동료 간의 설명을 구체화하고, 자신의 생각을 수정하고, 시작할 때의 수준과 끝날 때의 수준을 서로 비교하여 검토하면서 성장하게 해준다고 하였다. 그리고 사회

* 교신저자 : 백성혜 (shpaik@knue.ac.kr)

** 본 논문은 류혜정의 2014년도 석사 학위논문의 데이터를 활용하여 재구성하였음.

http://dx.doi.org/10.14697/jkase.2014.34.6.0593

적 구성주의 시각에서 보았을 때 동료평가의 활용은 학생과 교사, 학생과 학생간의 적극적인 상호작용 효과와 수업에 대한 집중을 가져올 수 있다고 하였다. 한편 동료 평가 방식은 교사 평가 방식에 비해 학생들 간의 활발한 언어적 상호 작용과 반성적인 사고를 촉진함으로써 메타인지 향상에 기여할 것이라고 주장하는 연구도 있다(Blom & Poole, 2004; Falchikov, 1986; Fry, 1990; Song, 2006). 특히 Wen과 Tsai(2006), Topping 등(2000)은 동료평가 과정을 통해 동료들이 어떻게 학습하는지 관찰할 수 있기 때문에 학습에 영향을 미친다고 하였다. Kang(2009)의 연구에서도 예비과학교사들이 동료 평가 결과를 객관적으로 받아들이고 수용하는 태도를 보였다.

그러나, 동료 평가는 학생들 사이의 상호 작용의 질에 달려 있기 때문에, 교사평가와 상관이 없다는 연구 결과도 있다(Chea, 2006). 특히 학생들은 동료 평가에서 관대하게 채점을 하는 경향이 있는 것으로 나타났다(Farr & Tone, 1994). 또한 오히려 정의적 측면에서 부정적인 영향을 미친다는 연구(Orsmond, *et al.*, 1996)도 있다. 그리고 Kim(2005)의 연구에 의하면, 평가자의 역할보다는 피평가자의 역할을 수행한 것이 높은 학업성취도와 상관이 있는 것으로 나타났다. 또한 이 연구에서는 잘 설계되고 고안된 학습 활동을 가진 평가자와 피평가자의 역할이 동료평가에 내재되어 있어야 교육적 효과가 나타난다고 주장하였다.

이러한 선행연구를 토대로 본 연구에서는 서술형 평가 문항을 교사가 평가하는 경우와 동료가 평가하는 경우에 대한 학생들의 만족도와 신뢰도를 비교하여 보고, 평가 활동이 학생들의 학업성취도 및 정의적 영역에 미치는 영향을 알아보려 하였다.

범교과적 수준에서 서술형 평가와 관련된 다양한 정의들을 살펴보면, ‘학생이 이미 알고 있는 지식을 활용하여 주어진 문제 상황을 논리적으로 분석, 설명, 해석하거나 창의적으로 해결하는 능력을 측정하는 평가’(Kim, *et al.*, 2010), ‘학생이 답을 직접 구성해야 하는 지필검사의 한 유형’(Park, 2008), ‘학생이 답이라고 생각하는 지식이나 의견 등을 직접 서술하도록 하는 평가 방식’(Back, 2000), ‘주어진 질문에 대해 학생들이 높은 반응 자유도를 보이면서 여러 개의 문장으로 응답하는 평가 형태’(Kyungkido Education Office, 2012), ‘창의성과 같은 고등사고기능을 묻거나 문제해결과정, 과학탐구 과정의 이해도를 측정할 수 있도록 구성된 평가’(Kang, *et al.*, 2010), ‘학생이 문제에 제시된 현상의 원인과 결과 관계, 과학적 개념간의 논리적인 순서, 연관성과 위계 등을 알고 있는지와 같은 학생의 사고 과정이 드러나는 평가’(Lee, *et al.*, 2011), 지식에 대한 적용 능력, 비판 능력, 감상 능력, 창의력, 종합적 판단능력, 문장 표현능력 등과 같은 고등정신능력을 측정하기 위한 평가’(Hahn, *et al.*, 1997) 등으로 제시되어 있다. 따라서 수행평가는 학생들의 문제해결능력과 창의력 등 고등사고 능력 향상에 중점을 두고 있으며, 학생이 답이라고 생각하는 지식, 해결 과정, 의견 등을 직접 ‘서술’하도록 하는 평가 방식을 의미한다.(Seoul Metropolitan Office of Education, 2010).

II. 연구 방법

1. 연구 대상 및 절차

연구자가 수업을 담당하고 있는 중소도시의 한 고등학교 2학년 2개

학급을 대상으로 하여 교사평가와 동료평가의 결과를 비교하였다. 특히 동료평가를 실시한 학급의 경우에는 학생들 사이의 상호작용 과정에서 나타나는 다양한 행동의 원인을 파악하기 위하여 연구자가 답임을 맡고 있는 학급으로 지정하였다. 또한, 연구 대상이 되는 두 학급 모두 연구자가 직접 수업을 진행하여 수업 변인에 대한 차이를 최소화하고자 노력하였다.

서술형 문항은 2009 개정교육과정 화학 I 교과서의 ‘개성 있는 원소’ 단원에서 추출하였으며 ‘현대의 원자 모형’, ‘주기율표’, ‘원소의 주기적 성질’ 등 3개의 주제에 해당하는 총 10차시의 내용으로 구성하였다.

연구대상이 되는 2개 학급을 대상으로 사전에 학업성취도와 정의적 영역 사전검사를 실시하였다. 이후 수업 과정은 동일하게 진행하였고 학급에 따라 교사평가 혹은 동료평가가 10차시 동안 이루어졌다. 교사평가 집단에 속한 28명의 학생들은 교사가 직접 평가하고 학생들의 풀이에 코멘트를 달아주었으며 동료평가 집단은 동료가 평가하고 풀이에 코멘트를 달아주는 것으로 진행하였다.

동료평가를 실시한 학급의 학생들은 27명으로, 사전 학업성취도 검사 점수를 기준으로 1등 ~ 7등까지는 최상위 그룹, 8등 ~ 14등까지는 중상위 그룹, 15등 ~ 21등까지는 중하위 그룹, 21등 ~ 27등까지는 최하위 그룹으로 분류하였다. 그리고 각 그룹에서 한 명씩 추출하여 3-4명이 한 집단을 구성하여 총 7개의 이질적인 소집단으로 구성하였다. 이는 협동적 상호작용에 대해 연구했던 많은 선행연구에서 소집단의 최적 규모와 수준에 대한 결과(Alexopoulou & Driver, 1996)에 따른 것이다. 교사평가 집단은 사전 학업성취도 검사의 석차 순위에 따라 1등~14등까지는 상위집단, 15등~28등까지는 하위집단으로 분류하였으며, 이들은 단지 결과의 특징을 비교하기 위한 구분으로 집단 활동을 실시하지는 않았다.

모든 평가 활동이 끝난 후에는 사후 학업성취도 검사와 정의적 영역을 실시하여 수업 처치의 효과를 분석하였다. 학업성취도의 사후검사는 2학년 1학기 2차 지필평가 중 서술형 평가로 실시하였고 같은 학년을 지도하는 교사들이 함께 협의하여 출제하였다. 정의적 영역 검사는 사전 검사와 동일한 검사지를 다시 한 번 투입하여 효과를 비교하였다. 사전 검사는 학기가 시작되는 3월 하순에, 사후 검사는 학기 말인 7월 초순에 실시되어 약 3개월 정도의 시간 간격이 있으므로 사전 검사 효과가 크지 않다고 판단된다.

2. 검사 도구

학업성취도 검사는 학생들의 전체적인 학업성취능력을 알아보기 위해서 같은 학년을 지도하는 3명의 과학 교사들이 함께 협의하여 출제하였다. 출제에 참여한 과학 교사들은 교육 경력이 5년 이상이며, 그 중 한 명은 관련 전공으로 석사 학위를 가지고 있는 평가 전문성을 갖춘 교사들이다. 정의적 영역 검사는 ‘국가수준의 과학과 관련된 정의적 영역 평가 체제(Kim, *et al.*, 1998)’를 사용하였다. 이 평가 체제는 검사지 A형 23문항, B형 25문항으로 구성되어 있으며 본 연구에서는 이들 두 검사지를 통합한 총 48문항의 설문지를 사용하였다. 각 검사 문항은 리커트 5점 척도로서 5점은 ‘매우 그렇다’ 이고, 1점이 ‘전혀 아니다’로 구성되어 있으므로, 부정적인 표현으로 질문한 응답의 경우에는 1점은 5점으로, 5점은 1점 등으로 재코딩하였다. 이 연구에서

Table 1. Reliability of affective domain questionnaire

평가 범주	평가 요소	문항 번호	Cronbach- α
1. 과학에 관한 인식	1.1 과학에 대한 인식	1, 4*, 8	0.628
	1.2 과학교육에 대한 인식	5, 9*, 12	
	1.3 과학과 관련된 직업에 대한 인식	2, 6, 10*	
	1.4 과학-기술-사회의 상호 관련성에 대한 인식	3*, 7*, 11	
2. 과학에 대한 흥미	2.1 과학에 대한 흥미	13, 18, 24	0.857
	2.2 과학 학습에 대한 흥미	14, 20, 25	
	2.3 과학과 관련된 활동에 대한 흥미	15, 21*, 26	
	2.4 과학과 관련된 직업에 대한 흥미	16, 22, 27	
	2.5 과학 불안	17*, 19*, 23*	
3. 과학적 태도	3.1 호기심	31, 37, 44	0.826
	3.2 개방성	28, 38, 45*	
	3.3 비판성	33, 39, 46	
	3.4 협동성	29, 32, 40	
	3.5 자진성	30*, 34, 41	
	3.6 끈기성	35, 42*, 47	
	3.7 창의성	36, 43, 48	
계		48 문항	

*부정형 문항

사용하기 위하여 55명 학생들을 대상으로 신뢰도를 측정된 결과를 Table 1에 제시하였다.

서술형 평가는 교과서 내용 중에서 원자의 구조와 주기적 성질 단원의 내용을 중심으로 개발하였다. 이 단원을 선택한 이유는 다른 단원에 비해 자료 해석뿐만 아니라 학생들의 사고 과정을 나타낼 수 있는 내용이 풍부하다고 판단되었기 때문이다. 평가지의 각 소재 및 문항은 서술형 평가 문항 작성 절차에 맞추어 작성하였으며, 모범 답안 및 평가 기준표를 첨부하여 평가 및 피드백의 자료로 활용하였다. 또한 서술형 문제 해결 과정은 선발이나 배치가 아닌 학습자의 반성적인 사고를 향상시키는 데 목적이 있었으므로 분석적인 평가법을 이용하여 평가하였다.

3. 평가 활동

평가 활동은 한 달 동안 주3회씩 총 10회 동안 실시하였다. 하나의 주제마다 수업을 진행한 후 개인별 문제해결 시간을 가졌다. 서술형 문제는 개별적으로 학생들에게 배부한 후 문제 해결을 하고 나서 자기 평가서를 작성하도록 하였다. 문제해결 직후 자기 평가서를 작성하게 한 것은 문제의 의도를 이해하고 있고 문제 해결 결과에 대해 확신하고 있는가를 알아보기 위한 것이었다. 또한, 자신의 문제 해결과정에 대한 평가를 적게 함으로써 학생들 스스로 자신의 문제 해결과정을 돌아보게 하는데 목적이 있었다.

개별적인 문제해결 과정 후 학생들의 답안지를 건어 잘못된 부분에 대해 코멘트를 달고 첨삭하는 방식으로 교사평가가 이루어졌다. 그 결과는 다음 수업시간에 평가 기준표와 함께 돌려주고, 학생들이 스스로 문제 해결 중에 실수와 틀린 이유에 대해 반성할 수 있도록 하였다. 또한, 각 문항의 출제 의도와 핵심 사항, 문제를 풀 때 중요한 개념들에 대해 전체 학생들을 대상으로 설명하였으며, 학생들의 실수에 대해서도 지도하였다.

Table 2. The comparison of activities with feedback of evaluation results

	교사평가 활동	동료평가 활동
공통점	<ul style="list-style-type: none"> • 한 달 동안 주 3회씩 10회 실시 • 하나의 주제 별로 30분 강의 • 20분 동안 평가활동 실시 • 개인별 문제해결시간 제공 및 자기 평가서 작성 	<ul style="list-style-type: none"> • 평가방법, 유의사항 안내 및 평가 기준표 제시
차이점	<ul style="list-style-type: none"> • 제출한 학생 답안지에 교사의 평가 피드백을 적고 평가 기준표 제공 • 학생 스스로 문제 해결 과정의 실수와 이유를 반성할 시간 제공 • 출제의도, 핵심사항, 중요개념, 보편적 학생 실수 등에 대한 설명을 전체 학생 대상으로 제시 	<ul style="list-style-type: none"> • 조별 동료 평가 실시 • 동료들의 지적 및 문제 해결을 위한 조언 등 상호작용 • 활발한 상호작용을 위한 교사의 안내 • 최종 답안 수정·제출

동료평가를 실시한 학급에서는 학생들과 교사 사이의 상호작용을 위해 학생들이 질문하고 교사가 대답하는 방식의 활동을 추가하였다. 교사의 설명과 답변이 있을 후 자신의 언어로 답안을 재수정하여 제출하도록 하여 학생들의 반성적 사고를 촉진하였다.

동료평가의 경우에도 하나의 주제마다 수업을 진행하고 나서 개인별 문제해결 시간을 가진 후 자신의 문제 해결 과정에 대한 1차 자기 평가서를 제출하게 하였다. 그리고 다음 차시에는 동료평가의 방향을 설정해 주기 위해 교사가 평가 방법에 대해 안내하고 유의할 점에 대해 설명을 하였다. 동료평가는 문제별로 학업성취도를 기준으로 같은 조로 편성된 학생 3-4명과 함께 평가해야 할 학생의 답안지를 놓고 평가 기준표 및 교재를 참고로 평가하고 첨삭 내용을 기재하는 방식으로 진행되었다. 이때 Cho(2004)의 연구에서 밝힌 바와 같이 평가 항목에 핵심적인 내용이 포함되면서 짧은 시간 안에 처리할 수 있도록 단순하고 간결하게 구성하였다. 하나의 문제에 대해 학생들의 의견을 밝히고 이견이 없거나 의견이 모여져 의견 조정이 이루어질 때까지 소집단별 상호작용을 할 수 있도록 하였다. 동료평가에서는 잘못된 부분을 지적해주거나 문제해결을 위한 조언을 제공할 수 있도록 하였다. 또한 학생은 최종적으로 다른 학생들의 조언을 참고하여 다시 답을 수정하여 제출하게 하였다. 교사는 토의 과정에 직접 개입하지 않고 방향을 알려주고 상호작용이 잘 이루어지지 않는 경우 활발한 상호작용이 일어날 수 있도록 안내해주는 역할을 하였다. 또한 학생들의 토의 과정에서 교사의 출제 및 채점 의도가 제대로 전달되어 평가가 올바르게 이루어지는 지에 대한 점검을 하였다.

비교 결과, 채점 기준이 명확한 경우에는 동료 평가 결과와 교사 평가 결과에 차이가 없었지만 진술의 해석이 필요한 애매한 상황에서는 선행연구와 달리 학생들의 동료 평가 결과가 교사 평가 결과보다 엄격하였다. 특히 학생들은 교사와 달리 해석이 다양할 수 있는 상황에서 점수를 부여하지 않으려는 경향이 있었다.

또한 교사평가 학급과 동료평가 학급간의 통제를 위해 전체 50분 수업 중 먼저 일반적인 수업 진도에 맞춰 강의식 수업을 한 후 20분을 남겨 놓고 평가 활동을 하였다. 그러나 주제에 따라 동료평가 학생들은 쉬는 시간까지 관련 문제에 대해 활발히 토의하는 모습을 보였으며 이러한 과정까지 정확하게 통제하지는 않았다. 대부분의 경우 학생 스스로가 시간제한을 염두에 두었으므로 두 집단의 활동 시간에 있어 큰 차이는 나지 않았다. 교사평가 활동과 동료평가 활동을 비교하여 Table 2에 제시하였다.

Table 3. Comparative analysis of students' satisfaction responses with feedback of evaluation results

평가유형	교사평가				동료평가			
	상위 집단	하위 집단	상위 집단	하위 집단	상위 집단	하위 집단	상위 집단	하위 집단
전혀 그렇지 않다	0 (0.0)	부정 (7.1)	부정 (7.7)	부정 (0.0)	부정 (0.0)	부정 (0.0)	부정 (0.0)	부정 (0.0)
별로 그렇지 않다	0 (0.0)	0	3 (21.4)	1 (15.4)	1 (7.7)	1 (7.1)	1 (7.1)	1 (7.1)
보통이다	2 (14.3)	중립 (14.3)	4 (28.6)	중립 (28.6)	3 (23.1)	중립 (23.1)	7 (50.0)	중립 (50.0)
조금 그렇다	9 (64.3)	긍정 (35.7)	5 (35.7)	긍정 (53.8)	7 (53.8)	4 (28.6)	4 (28.6)	4 (28.6)
매우 그렇다	3 (21.4)	12 (85.7)	2 (14.3)	7 (50.0)	1 (7.7)	8 (61.5)	2 (14.3)	6 (42.9)
계	14(100.0)		14(100.0)		13(100.0)		14(100.0)	

Table 4. Reasons for help of solution for descriptive evaluation

이유	교사평가	동료평가	계
서술형 평가 문항에 자신감이 생김	7(31.8)	3(15.7)	10(24.4)
다양한 풀이 방법의 습득	2(9.1)	4(21.1)	6(14.6)
학습내용에 대해 정확히 알게 됨	8(36.4)	7(36.8)	15(36.6)
잘못 알던 내용을 바로잡을 수 있었음	3(13.6)	1((5.3)	4(9.8)
무응답	2(9.1)	4(21.1)	6(14.6)
계	22(100.0)	19(100.0)	41(100.0)

III. 연구 결과

1. 서술형 평가에 대한 만족도 분석

Table 3에 제시한 바와 같이 교사 평가 상위 집단 학생들은 85.7%가 평가 결과의 피드백에 긍정적인 만족도를 보인 반면 동료 평가 상위 집단 학생들은 60.8%가 긍정적 만족도를 보여 상위 집단 학생들은 상대적으로 교사 평가를 선호하고 있다는 것을 알 수 있다. 한편 하위 집단 학생들의 반응을 분석해 보면 교사 평가 학생들의 부정적 답변은 21.4%, 동료 평가 학생들의 부정적 답변은 7.1%였다. 따라서 하위 집단 학생들은 교사 평가보다 동료 평가에 대해 더 긍정적인 인식을 하는 것을 알 수 있다.

긍정적인 응답의 이유에 대해 알아본 결과를 Table 4에 정리하여 제시하였다. Table 4에서 알 수 있듯이 가장 높은 비율의 학생들이 학습 내용에 대해 정확히 알게 되었다고 대답하였다. 이는 문제를 평가 하는 과정에서 왜 이런 풀이가 나왔는지 고민하고 반성적인 사고의 기회를 가졌기 때문이라고 생각된다.

또한 교사평가에서는 서술형 평가 문항에 자신감을 가져 도움이 되었다는 학생들의 반응 비율이 높은 반면 동료평가에서는 다양한 풀이 방법을 습득하여 자신의 답과는 다른 관점으로 학습 내용을 이해하여 생각의 폭을 더 넓혀주는 효과를 가졌다고 생각하는 학생의 비율이 높았다. 이는 선행연구에서 Defina(1992)와 Green(1994), Hahn과 Oh(2008)등이 동료평가의 장점으로 제안한 것과 일치하는 결과이다.

Table 5. Comparative analysis of students' reliability response with feedback of evaluation results

평가유형	교사평가				동료평가			
	상위 집단	하위 집단	상위 집단	하위 집단	상위 집단	하위 집단	상위 집단	하위 집단
전혀 그렇지 않다	0 (0.0)	부정 (0.0)	부정 (0.0)	부정 (0.0)	부정 (0.0)	부정 (0.0)	부정 (0.0)	부정 (0.0)
별로 그렇지 않다	0 (0.0)	0	0 (0.0)	0 (0.0)	1 (7.7)	0 (0.0)	0 (0.0)	0 (0.0)
보통이다	1 (7.1)	중립 (7.1)	8 (57.1)	중립 (57.1)	4 (30.8)	중립 (30.8)	5 (35.7)	중립 (35.7)
조금 그렇다	6 (42.9)	13 (92.9)	3 (21.4)	6 (42.9)	7 (53.8)	8 (61.5)	9 (64.3)	9 (64.3)
매우 그렇다	7 (50.0)	3 (21.4)	3 (21.4)	6 (42.9)	1 (7.7)	0 (0.0)	0 (0.0)	0 (0.0)
계	14(100.0)		14(100.0)		13(100.0)		14(100.0)	

2. 평가의 피드백에 대한 신뢰도 분석

서술형 평가 문항을 실시한 총 55명 중 36명의 학생이 ‘서술형 평가에서 교사(동료)가 제공한 피드백을 신뢰한다.’ 고 긍정적인 응답을 하였다. Table 5에서 알 수 있듯이 교사 평가의 경우 상위 집단 학생들은 92.9%가 평가 결과의 피드백에 긍정적 신뢰를 보였던 반면, 동료 평가에서는 상위 집단 학생들의 61.5%만이 신뢰한다고 응답하였다. 따라서 상위 집단의 경우 동일한 채점 기준을 제공한 경우에도 교사의 평가를 신뢰하는 반면, 상대적으로 동료 평가를 불신하고 있음을 확인할 수 있었다. 따라서 Cho(2004)의 제안처럼 학생들이 서로의 조언을 제공하는 데에 있어서 사회적 기술에 대한 훈련이 필요하다. 반면 하위 집단의 경우에는 교사 평가에 대한 신뢰도에 긍정적인 반응을 한 비율이 42.9%이었으나, 동료 평가에 대한 신뢰도에 긍정적 반응을 한 비율은 64.3%로 반대 현상이 나타났다.

3. 서술형 평가활동이 학습자의 학업성취도에 미치는 영향

평가 전에 먼저 두 집단이 동질 집단인지를 알아보기 위하여 교사평가 학급과 동료평가 학급에 1차 지필평가를 실시한 뒤 서술형 평가 문항의 점수를 상위 집단과 하위 집단별로 나누어 맨-휘트니 U 검정(Mann-Whitney Test)를 실시하였다(Table 6). 휘트니 U 검정은 샘플 수가 5~21 정도일 때 사용하기 적절한 분석방법이다(http://vassarstats.net/utest.html). 검증 결과 유의확률 p값이 학업성취도 상위 집단에서는 0.395, 학업성취도 하위 집단은 0.241로 유의수준 0.05에서 두 집단은 동질 집단임을 알 수 있다.

그 후 2차 지필평가를 실시한 뒤 맨-휘트니 U 검정(Mann-Whitney Test)를 실시한 결과, 학업성취도 상위 집단에서는 검증 결과 유의확률 p값이 0.190으로 학급 간 유의미한 차이가 나타나지 않았으나, 학업성취도 하위 집단은 0.048로 유의수준 0.05에서 평가 방식에 따라 학업성취도가 변화하였음을 확인할 수 있었다.

하위 집단 학생들끼리 비교해 보았을 경우 학업성취도의 향상은 동료 평가 학급에서 유의미하게 증가하였다. 이는 동료 평가에 의한 피드백으로 하위 집단 학생들이 학습에 긍정적인 자극을 받았음을 의미하며, Watson과 Marshall(1995), Webb(1982) 등의 연구 결과와도

Table 6. Homogeneity analysis of pre- and post-scholastic achievement test by levels according to class assessment types

학업 성취도	검사	학급	인원 수	평균	표준 편차	맨-휘트니 U	유의 확률
상위 집단	사전	교사 평가	14	42.74	5.907	73.500	0.395
		동료 평가	13	45.04	5.496		
	사후	교사 평가	14	39.82	7.154		
		동료 평가	13	36.89	5.409		
하위 집단	사전	교사 평가	14	16.44	11.754	72.500	0.241
		동료 평가	14	21.89	9.963		
	사후	교사 평가	14	13.21	8.987		
		동료 평가	14	20.93	9.687		

*p<0.05

Table 7. Homogeneity analysis of pre- and post-affective domain test by scholastic achievement levels according to class assessment types

학업 성취도	검사	학급	인원 수	평균	표준 편차	맨-휘트니 U	유의 확률
상위 집단	사전	교사 평가	14	161.86	19.183	60.500	0.138
		동료 평가	13	167.46	22.393		
	사후	교사 평가	14	166.00	13.068		
		동료 평가	13	169.79	18.031		
하위 집단	사전	교사 평가	14	150.07	9.335	97.500	0.982
		동료 평가	14	151.00	13.989		
	사후	교사 평가	14	146.07	11.532		
		동료 평가	14	156.50	13.716		

*p<.05

일치하는 것이다.

이 연구의 의미는 하위 집단 학생들의 학업성취도 향상에 있다. 동료 평가 활동 자체가 모든 학생들의 활발한 상호작용을 유발하는 효과가 있다고 한다면, 상위 집단 학생들의 경우에도 교사 평가 학급에 비해 학업성취도의 유의미한 향상이 나타나야 한다. 하지만, 이 연구에서는 하위집단 학생들에게만 유의미한 성취도 향상이 나타났다. 그 이유는 Table 5의 결과에서 논의한 바와 같이 상위 집단 학생들의 경우 동료 평가 결과를 신뢰하지 않았기 때문이다. 평가의 기준이 명확하게 제시되어 있고, 오히려 학생 평가가 더 엄격하였음에도 불구하고 동료에 의해 이루어진 평가 결과를 신뢰하지 않는 태도로 인해 동료 평가의 긍정적인 효과를 상위 집단 학생들은 나타내지 못하였다. 이는 성취도가 높은 학생들의 동료에 대한 신뢰의 문제가 해결되지 않는다면 동료 평가가 모든 학생들에게 긍정적일 수 없음을 의미하는 것이다.

4. 평가활동이 학습자의 정의적 영역에 미치는 영향

학생들의 학업성취도에 따라 평가 활동이 정의적 영역에 미치는 영향을 알아보기 위하여 맨-휘트니 검정을 실시하였다(Table 7). 그 결과, 동료 평가 학급의 하위 집단만이 평가 활동에 의해 정의적 영역이 유의미하게 향상되었음이 밝혀졌다. 즉, 학업성취도 상위 집단의 학생들은 평가 방법에 의해 정의적 영역이 그다지 영향 받지 않지만, 하위 집단 학생들은 동료 평가가 정의적 영역에서 긍정적인 영향을 미친 것이다. 이는 Table 5에 제시한 성취도 하위 집단 학생들의 학업 성취도 향상과 관련지어, 동료 평가는 성취도가 낮은 학생들의 인지적

Table 8. Homogeneity analysis of pre- and post-test for low scholastic achievement level students' recognition, interests, and attitude of science

항목	학급	인원수	평균	표준 편차	맨-휘트니U	유의확률
과학에 관한 인식	교사 평가	14	38.50	3.937	44.000	.013*
	동료 평가	14	43.07	5.091		
과학에 대한 흥미	교사 평가	14	42.57	3.031	72.500	.239
	동료 평가	14	44.21	5.727		
과학적 태도	교사 평가	14	65.00	7.835	72.000	.229
	동료 평가	14	69.21	7.495		

*p<.05

영역과 정의적 영역에서 매우 긍정적인 효과를 나타냄을 확인할 수 있다. 비록 성취도에 따른 구분을 한 연구들은 아니지만, 이러한 연구 결과는 Ballantyne 등(2002), Blom과 Poole(2004), Falchikov(1986), Fry(1990), Song(2006) 등의 연구에서 주장한 것과 일치한다고 볼 수 있다.

정의적 영역 중에서 과학에 관한 인식, 과학에 대한 흥미, 과학적 태도별로 나누어 분석한 결과, 과학에 관한 인식에서의 학업성취도 하위 집단의 유의확률이 0.013으로 유의수준 0.05에서 유의미한 향상이 나타났다(Table 8).

학업성취도 하위 집단 학생들은 동료 평가 활동을 통해 다른 학생들과의 문제 해결 과정과 자신의 과정을 비교해 볼 기회를 갖고, 자신이 잘 모르고 있는 부분에 대한 피드백을 받으면서 과학에 관한 인식의 변화가 가장 뚜렷하게 나타났다. 이는 Cho(2004), Wen과 Tsai(2006), Topping 등(2000)의 연구 결과와도 일치하는 것이다.

IV. 결론 및 제언

본 연구에서는 서술형 문제에 대한 교사 평가와 학생 평가 과정의 차이가 학습자의 학업성취도나 정의적 영역에 어떤 영향을 미치는지 알아보고자 하였다. 연구 결과, 학업성취도 상위 집단 학생들은 상대적으로 교사평가에 긍정적인 만족도를 나타낸 비율이 높았으며, 하위 집단 학생들은 동료평가에 긍정적인 만족도를 나타낸 비율이 높았다. 가장 높은 비율의 학생들이 학습 내용에 대해 정확히 알게 된 점을 긍정적인 만족의 이유로 답하였다. 특히 동료평가 집단에서는 다양한 풀이 방법을 습득하여 자신의 답과는 다른 관점으로 학습 내용을 이해하여 생각의 폭을 더 넓혀주는 효과를 가졌다고 생각하는 학생의 비율이 높았다. 이를 통해 교사가 일률적으로 제공하는 모범 답안 보다 다른 친구들의 답을 보면서 다양한 동료들의 생각을 알아보는 기회를 가지는 것은 자신의 답에 대해 반성적 사고를 경험하고 학습 내용을 보다 정확히 이해할 있는 기회를 제공해 줌을 알 수 있었다.

특히 학업성취도가 하위인 학생들만 비교한 결과 교사평가를 실시한 학급보다 동료평가를 실시한 학급에서 성취도의 향상이 통계적으로 유의미하게 나타났다. 이를 통해 자신보다 성취수준이 높은 동료와의 상호작용이 학업성취도가 낮은 학생들에게 체계적인 학습 내용 이해나 잘못된 개념 수정의 기회를 효과적으로 제공하였음을 알 수 있다. 정의적 영역의 평가 결과에서도 학업성취도 상위 집단의 학생들은 평가 방법에 큰 영향을 받지 않았지만, 하위 집단 학생들은 동료 평가가 정의적 영역에서 긍정적인 영향을 미쳤으며, 특별히 과학에 대한 인식 영역에서 변화가 통계적으로 유의미하게 나타났다. 따라서 인지적 영

역 뿐 아니라 정의적 영역에서도 하위 집단 학생들은 동료평가의 효과가 매우 뚜렷하게 나타남을 알 수 있다.

한편, 학업성취도 상위 집단 학생들은 교사평가와 학생평가에서 동일한 채점 기준을 제공하였음에도 불구하고 교사평가에 대한 신뢰도가 상대적으로 높은 반면, 하위 집단 학생들은 교사평가에 대한 신뢰도가 상대적으로 낮았다. 학업성취도가 높은 학생들은 동료들의 평가를 신뢰할 수 없어서 서술형 평가가 본인에게 도움이 되지 않았다고 응답한 경우가 있었다. 이를 통해 학업성취도가 상위인 학생들은 교사평가를 신뢰하고 자신보다 학업성취도가 낮은 학생들이 제공하는 평가에 대해 부정적인 편견을 가지고 있음을 알 수 있다. 반면 학업성취도 하위 집단의 학생들은 동일한 평가 기준으로 평가한 경우에도 동료평가를 신뢰하는 비율이 높고 교사평가를 신뢰하는 비율이 낮았다. 비록 교사평가에 대해 부정적인 응답을 한 경우는 없었지만, 중립적인 응답 비율이 높았으며, 과학적 지식이 부족해서 서술형 평가가 본인에게 도움이 되지 않았다는 답변이 있었다. 따라서 학업성취도가 낮은 학생들은 서술형 평가 자체에 대해 부정적인 생각을 가지고 있음을 알 수 있다.

평가 활동에서 나타나는 학생들의 반응을 관찰한 결과, 학생들은 평가를 담당한 주체에 따라 과제에 대한 긴장감 혹은 본인의 잘못된 풀이를 고치고자 하는 욕구의 경향이 다르게 나타나는 것을 관찰할 수 있었다. 동료평가를 실시한 학급의 조별 토론 과정에서 학업성취도가 높은 학생들은 적극적으로 다른 학생의 답안에 점수를 부여하고 그 과정에서 스스로 성취감을 느끼며 활동을 하는 모습이 관찰되었다. 한편 학업성취도가 낮은 학생들은 평가가 끝난 후에 틀린 문제에 대해 고민하고 조의 다른 동료들과 상호작용하면서 자신의 답의 오류를 파악하고 수정하고자 하는 적극적인 태도를 보였다. 그에 반해 교사가 평가한 학급에서는 학업성취도가 하위 집단의 학생들은 상대적으로 평가 결과에 큰 관심을 보이지 않았다.

흥미로운 현상은 동료평가를 실시한 학급에서는 평가 기준표에 없는 애매한 응답에 대해 교사평가 학급보다 더 낮은 점수를 부여하였다는 것이다. 이는 연구자가 직접 교사평가를 수행하였기 때문에 비교가 가능하였다. 이러한 경향이 나타나는 이유는 평가 활동을 주도한 학업성취도가 우수한 학생들이 정확하지 않은 답에 대한 평가의 기준을 엄격하게 적용하였기 때문이다. 교사의 경우 응답한 학생들의 상황을 참조하여 응답을 해석하는 경향이 있지만, 학생들은 이러한 경향을 보이지 않았다. 또한 활발한 토의를 통해 학생들의 합의하에 부분점수를 새롭게 설정하여 평가하는 모습도 관찰되었다.

이러한 자기주도적인 평가 활동은 동료평가의 장점으로 특히 학업성취도가 낮은 학생들의 과학에 대한 인식 변화 뿐 아니라 과학 개념에 대한 이해를 높이고 평가에 대한 신뢰도 및 만족도를 높이는데 중요하게 작용하므로, 서술형 평가의 경우 동료평가를 적극적으로 활용할 필요가 있다고 본다. 학업성취도가 높은 학생들의 경우에도 평가 활동에 주도적으로 참여하고 성취도가 낮은 학생들과 협의하여 평가 기준을 새롭게 설정하고 채점 과정에 대해 토론하면서 동료에 대한 신뢰를 구축하고, 교사도 평가의 주체로서의 역할보다는 평가활동의 지지자로서 역할을 수행함으로써 많은 학생들에게 보다 공정하게 받아들여지는 평가활동을 제공하는 역할을 수행할 필요가 있다.

많은 교사들이 서술형 평가문항의 개발과 채점의 공정성을 확보하는데 어려움을 겪고 있다. 이러한 문제 때문에 서술형 평가를 기피하거

나 단답형에 가까운 서술형 평가문항을 개발하는 경향도 있다. 그러나 서술형 평가의 목적을 고려할 때 채점의 용이성을 고려한 단답형 평가를 지양하고, 동료평가라는 방안을 통해 교사들이 서술형 평가에서 겪는 어려움을 일부 해결할 수 있을 것이다. 동료평가는 학생들이 조별로 협의하여 판단하고 채점한 결과에 대해 신뢰하고 만족하는 경향이 높기 때문에 평가 결과에 대한 불만이 훨씬 적을 수 있으며, 학생들의 다양한 생각에 대한 평가도 용이해질 수 있기 때문이다.

연구의 시작 단계에서는 동료평가에 대한 학생들의 부담감이 큰 것으로 관찰되었으나, 평균적으로 1주일에 1~2회 정도의 실시하는 과정에서 학생들이 점차 평가활동에 익숙해지면서 부담감이 줄어들고 토론이 활성화되어가는 과정을 관찰할 수 있었다. 따라서 동료평가로 인해 발생하는 시간적 문제나 평가의 질적 문제는 크게 발생하지 않을 것으로 판단되지만, 이에 대한 구체적인 연구는 추후에 더 이루어질 필요가 있다고 본다. 특히 이 연구에서 수행하지 못했던 동료평가 과정에서 이루어지는 학생들의 의사소통에 대한 정밀한 분석을 통해 평가에 대한 학생들의 사고과정 및 개념 습득 과정에 대한 연구를 진행한다면 교육적으로 도움이 되는 더 많은 시사점을 찾을 수 있으리라 생각된다.

국문요약

이 연구에서는 서술형 문항에 대한 교사평가와 동료평가가 고등학교 학생들의 학업성취도와 정의적 영역에 미치는 효과를 알아보고, 평가에 대한 학생들의 만족도와 신뢰도를 분석하여 시사점을 얻고자 하였다. 이를 위하여 중소도시에 있는 고등학교 2학년 2개반을 선정하여 학업성취도, 정의적 영역의 검사를 사전과 사후에 실시하였다. 서술형 문항의 평가는 10차시 수업 동안 진행되었으며, 교사평가를 실시한 반에서는 학생들이 서술형 문항을 푼 후에 교사가 이에 대한 피드백을 수업 중에 실시하였다. 동료평가를 실시한 반에서는 학생들이 서술형 문항을 푼 후에 소집단 별로 동료평가를 실시하고 피드백을 제공하도록 하였다. 연구 결과, 학업성취도 상위 집단 학생들은 상대적으로 교사평가에 긍정적인 만족도를 나타내었고, 하위 집단 학생들은 동료평가에 긍정적인 만족도를 나타내었다. 한편, 학업성취도 상위 집단 학생들은 교사평가와 학생평가에서 동일한 채점 기준을 제공하였음에도 불구하고 교사평가에 대한 신뢰도가 상대적으로 높은 반면, 하위 집단 학생들은 교사평가에 대한 신뢰도가 상대적으로 낮았다. 학업성취도가 하위인 학생들만 비교한 결과 교사평가를 실시한 학급보다 동료평가를 실시한 학급에서 성취도의 향상이 통계적으로 유의미하게 나타났다. 정의적 영역의 평가 결과에서도 하위 집단 학생들은 동료 평가가 정의적 영역에서 긍정적인 영향을 미쳤으며, 특별히 과학에 대한 인식 영역에서 변화가 통계적으로 유의미하게 나타났다.

주제어 : 동료평가, 교사평가, 학업성취도, 정의적 영역

References

- Alexopoulou, E., & Drive, R. (1996). Small-group discussion in physics: Peer interaction modes in peers and fours. *Journal of Research in Science Teaching*, 33(10), 1099-1114.
- Ballantyne, R., Hughies, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action

- research process. *Assessment and Evaluation in Higher Education*, 27(5), 427-441.
- Back, S. K. (2000). *The principle of descriptive evaluation*. Seoul: Education Science Company.
- Blom, D., & Poole, K. (2004). Peer assessment of tertiary music performance. *British Journal of Music Education*, 21(2), 111-125.
- Chea, S. J. (2006). A study on concordance of student assessment in problem based learning of medical schools. *The Journal of Yeolin Education*, 14(1), 331-349.
- Cho, H. M. (2004). A theoretical exploration of peer evaluation in the perspective of social constructivism. *The Journal of Education*, 23, 229-254.
- Defina, A. A. (1992). *PortFolio assessment. Getting started*. New York, NY: Scholastic Professional Books.
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education*, 11, 146-166.
- Farr, R., & Tone, B. (1994). *Portfolio and performance assessment*. NY: Arlex publishing.
- Fry, S. (1990). Implementation and evaluation of peer marking in higher education. *Assessment and Evaluation in Higher Education*, 15, 177-189.
- Green, M. (1994). Children as evaluation understanding from the inside. In B. Cambourne & J. Turbil (Eds.), *Responsive evaluation* (pp. 83-103). Portsmouth, NH: Heinemann.
- Ha, B., Lee, Y., Ha, O., & Sin, M. (2000). Development of practical instrument for the measurement of science process skills of elementary school students. *Journal of Research in Science & Mathematic education*, 26, 33-72.
- Hahn, A. J., Kang, H., Kwan, C., Kim, H., & Woo, J. (1997). *New elementary science teaching method*. Seoul: Education science company.
- Hahn, J. S., & Oh, J. S. (2008). Effects of peer rating on the web-based online discussion. *Journal of Educational Technology*, 24(1), 87-107.
- Hong, M., & Chung, H. (2006). The development of assessment items with assay & description type test science in biology for high school students. *Journal of Science Education Kyungpook National University*, 30, 65-100.
- <http://vassarstats.net/utest.html> The logic and computational details of the Mann-Whitney test. chapter 11a. *Concepts and Applications of Inferential Statistics*. provided by the Department of Mathematics & Statistics at the University of Saskatchewan: ©Richard Lowry 2001-2014.
- Hwang, J., Lee, H., & Kwack, D. (2010). An assesment of high school students' performance on science process skills in biology, *Research for Subject Matter Education*, 14(1), 67-84.
- Kang, K. H. (2009). The analysis of results and recognition on pre-service science teachers' peer evaluation in presentation activity. *Korean Journal of Teacher Education*, 25(3), 261-275.
- Kang, H., Gong, Y., Kwon, H., Kim, J., Bae, J., Song, M., Sin, Y., Yang, I., Yun, H., Lee, D., Lee, M., Im, C., Im, H., Jang, S., Jun, Y., & Chae, D. (2010). *Elementary science education theory*. Seoul: Education Science Company.
- Kim, M. J. (2005). Peer assessment as a learning method: The effects of assessor and assessee's roles on metacognition, performance, and motivation. *Journal of Educational Technology*, 21(4), 1-28.
- Kim, H. N., Jung, W. H., & Jung, J. W. (1998). National science evaluation system development of affective characteristics. Chungbuk: Science Education Institute of Korea National University of Education.
- Kim, J. H. (2010). *Descriptive evaluation ROAD VIEW*, Seoul: ParkLeeJung Publisher.
- Kim, K., Seo, Y., & Lee, Y. (2013). An Application Method Study on Written Performance Assessments in Middle School Informatics according to the Revised Curriculum in 2009. *Korean Journal of Teacher Education*, 29(4), 187-203.
- Kim, S., & Yoo, Y. (2006). Developing descriptive essay problems in mathematics for high school students. *Journal of Science Education Kyungpook National University*, 30, 1-22.
- Kim, Y., & Kim, Y. (2012). The Development of a free-response test for the assessment of science process skill. *Biology education*, 40(1), 167-177.
- Kyungkido Education Office. (2012). *Guidebook of descriptive & Essay evaluation*. North government office building, Support section of teaching and learning.
- Lee, J., You, Y., Sin, E., Jung, J., Im, K., Gu, E., Kim, H., Ham, D., Kwak, J., Park, S., Park, S., Kim, S., & Kim, J. (2011). *Descriptive evaluation ROAD VIEW-Science*, Seoul: ParkLeeJung Publisher.
- Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in peer and self-assessment. *Assesment and Evaluation in Higher Education*, 21(3), 239-249.
- Park, D. S. (2008). *Evaluation of education*. Seoul: Education Science Company.
- Seoul Metropolitan Office of Education. (2010). *Reinforcing method of descriptive and assay assessment for students' improvement of scholastic ability in Seoul*. Report paper 2007.
- Song, H. M. (2006). *The analysis of the effect of peer evaluation and the verbal interaction in small groups in solving descriptive questions of metabolism* (Doctoral dissertation). Korea National University of Education, Chungbuk.
- Topping, K., Smith, F. F., Swanson, I., & Elliot, A. (1998). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education*, 25(2), 149-169.
- Yang, G. S. (2006). *Descriptive and essay assessment source book*. Seoul: Seoul Metropolitan Office of Education.
- Watson, S. B., & Marshall, J. E. (1995). Effects of cooperative incentives and heterogeneous arrangement on achievement and interaction of cooperative learning groups in a college life science course. *Journal of Research in Science Teaching*, 32(3), 291-299.
- Webb, N. M. (1982). Group Composition, Group Interaction, and Achievement in small groups. *Journal of Educational Psychology*, 74(4), 475-484.
- Wen, M. L., & Tsai, A. (2006). University students' perceptions of and attitudes toward online peer assessment. *Higher Education*, 51, 27-44.