

고등학교 논의기반 탐구 과학수업에서 학생 평가활동이 반성적 사고에 미치는 영향

이선우 · 남정희*
부산대학교

Impact of Student Assessment Activities on Reflective Thinking in High School Argument-Based Inquiry

Seonwoo Lee, Jeonghee Nam
Pusan National University

ARTICLE INFO

Article history:

Received 24 March 2016
Received in revised form
20 April 2016
Accepted 26 April 2016

Key words:

peer assessment, self assessment,
argument-based inquiry
reflective thinking metacognition

ABSTRACT

This study focused on the use of student assessment activities to investigate the impact on reflective thinking in Argument-based Inquiry. The participants of the study were 166 10th grade students (six classes). Over one semester, students participated in five ABI programs that we developed. The experimental group (84 students) was taught Argument-Based Inquiry with students' self and peer assessment activities. The comparative group (82 students) was taught without the activities. We analyzed students' reflective writing to investigate how the student assessment activities influenced the students' reflective thinking. We also used the interviews and surveys to examine the validity of student assessment activities.

According to analysis of the reflective writing, the experimental group had a significantly higher mean score than the comparative group in the 3rd and 5th writing. The ratio of students who showed a metacognitive level of reflection with regard to analysis of inquiry process, understanding of learning, and change of thinking increased in both groups, but the experimental group's ratio was higher than the comparative group's. The result of analysis of the reflective practice showed that the ratio of the experimental group's students who reached the metacognitive level of reflection in their writing increased, while the comparative group's decreased. Therefore, we conclude that student assessment activities can create a learning environment that facilitates student participation, increases the students' engagement in the learning process, and can be used as a tool to scaffold learning.

1. 서론

2015년 9월에 발표된 개정 과학 교육과정에서는 다양한 탐구 중심의 학습을 강조하면서 기본 개념에 대한 통합적인 이해를 가지고 탐구 경험을 통한 과학적 사고력, 과학적 탐구 능력, 과학적 문제 해결력, 과학적 의사소통 능력, 과학적 참여와 평생 학습 능력 등의 과학과 핵심역량을 함양할 수 있는 교육을 추구한다(Ministry of education, 2015).

과학 탐구과정에서 중요시 되고 있는 과학적 의사소통 능력은 과학적 문제 해결 과정과 결과를 공동체 내에서 공유하고 발전시키기 위해 자신의 생각을 주장하고 타인의 생각을 이해하여 조정하는 능력으로, 이는 탐구 중심의 학습이 이루어지는 데 있어 학생들이 갖추어야 할 핵심 역량이다(Ministry of education, 2015). 또한 과학적 의사소통의 방법으로써 논의 활동은 과학 개념에 대한 학생들의 이해력을 향상시키고(Sampson *et al.*, 2013; Tavares, Jiménez-Aleixandre, & Mortimer, 2010), 비판적인 사고력을 향상시킬 수 있을 뿐만 아니라(Nam *et al.*, 2011; Erduran & Jiménez-Aleixandre, 2007; Osborne, Eduran, &

Simmon, 2004; Phillips & Norris, 1999; Yore, Bisanz, & Hand, 2003), 과학적 사고력을 향상시킬 수 있다(Ministry of education, 2015).

과학 학습에서 협력적 활동의 기회를 학생들에게 제공하여 학생들의 의사소통 능력과 논의 능력을 향상시킬 수 있는 교수방법으로 제안되어온 논의기반 탐구 과학수업(Argument-Based Inquiry, ABI)은 교사가 제시하는 문제 상황에 대한 의문을 가지고 이를 해결하기 위해 실험을 설계하고 수행하여 주장과 증거를 형성하는 전 과정에 학생들 간의 논의에 의한 협상과 합의과정으로 이루어지는 교수방법이다(Keys *et al.*, 1999; Jang, Nam, & Choi, 2012). 특히 협상을 통한 주장과 증거를 형성하는 과정에서 모둠원들 사이 또는 모둠 간의 협력적 관계가 형성되었을 때 활발한 논의가 이루어질 수 있으므로(Keys *et al.*, 1999; Jang, Nam, & Choi, 2012), 논의기반 탐구 과학수업에서는 활동에 참여하는 학생들의 적극적인 수업 태도가 뒷받침되어야 한다(Duschl, Schweingruber, & Shouse, 2007).

학습에서 학생들의 참여도를 높이는 방법으로 평가 활동에 학생들을 참여시켜 평가를 학습의 일부분으로 통합시키는 방법을 제안한다(Lee, Bak, & Nam, 2015; Purchase, 2000; Rust, Price, & O'Donovan,

* 교신저자 : 남정희 (jhn@pusan.ac.kr)

** 본 논문은 이선우의 2016년도 박사 학위논문의 데이터를 활용하여 재구성하였음.

*** 이 논문은 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음.

<http://dx.doi.org/10.14697/jkase.2016.36.2.0347>

2003; Smyth, 2004). 구성주의적 관점에서 학생들은 자신의 학습에 주체적이고 능동적으로 참여해야 하며 학업 성취에 대해서도 스스로 책임감을 가져야 한다. 또한 학생들은 학습에서 자신이 목표를 설정하고 학습을 점검하며 평가를 수행하는 모든 과정에 능동적으로 참여해야 한다(Fiske & Taylor, 1991). 즉, 교사의 고유한 권한이었던 평가 활동에 학습자를 참여시키는 것은 학생을 자신의 학습 과정에 능동적으로 참여시키고, 이로써 학습에 대한 책임감을 갖도록 하는 방법이 될 수 있다. 또한 학습과정에서 이루어지는 학생들의 평가 전략 자체가 학습 전략이 될 수 있으므로 학생들에게 평가하는 방법을 가르쳐야 한다(Fairbrother *et al.* 1995)고 제안하고 있다.

학생들을 평가활동에 참여시켜 평가를 학습의 과정으로 통합시키는 방법으로, 동료가 수행한 과제에 등급을 정하거나 피드백을 제공하는 것을 의미하는 동료평가활동(Boud & Falchikov, 2007)과 학습자 스스로 자신의 학습 성과에 대해 인식하고 반성하여 수정하는 자기평가활동이 있다(Andrade & Du, 2007). 논의 활동에서 동료평가는 2명 이상의 학생 사이에서 일어나는 협력적 활동이므로 학습 내용에 대해 서로 피드백을 제공하는 과정을 통해 학습을 향상시킬 수 있으며(Kollar & Fischer, 2012; Spiller, 2009), 동료들에게 즉각적인 피드백을 제공하면서 논의의 질을 판단하는 능력을 향상시킬 수 있다(Larson, Britt & Kurby, 2009). 또한 자기평가활동을 통해 학생들은 자신이 알고 있는 것에 대해 점검하고 평가하는 메타인지능력을 향상시킬 수 있다(Biggs & Moore, 1993; Boud & Falchikov, 2007). 이처럼 논의 활동에서도 피드백 및 평가의 기회를 학생들에게 제공하는 학생 평가활동은 학생의 참여 문화를 촉진시킬 수 있다(Kollar & Fischer, 2012).

학생들의 학습 능력을 향상시키기 위해서는 자신의 학습에 대한 반성 과정이 필요하다(Dewey, 1933). 논의기반 탐구 과학수업의 마지막 단계에 학생들이 작성하는 반성 글쓰기는 학습자가 자신의 학습 과정에 대해 인식하고 모니터링하는 메타인지적 활동 과정이다(Bangert-Drown, Hurley, & Wilkinson, 2004, Nam *et al.*, 2008; Sung & Nam, 2013). 학생들은 반성 글쓰기를 통해 자신이 수행한 활동 전반에 대해 되돌아보는 기회를 가질 수 있으며, 이러한 반성은 평가 활동을 통해 피드백을 제공함으로써 효과적으로 유도할 수 있다(Cheng & Warren, 1999, Mayer, 2004; Moreno & Mayer, 2005; Pope, 2001; Strange & Mumford, 2005). 학습자들은 피드백을 제공받지 않은 상태에서 자신의 수행의 성공과 실패여부에 대하여 반성할 때 학습수행에 대하여 잘못된 전략을 적용할 수 있기 때문에 학습자에게 아무런 피드백을 제공하지 않는 경우보다 다른 사람으로부터 피드백을 통한 안내를 제공받은 교수전략이 더 효과적일 수 있다(Mayer, 2004). 따라서 동료평가 및 자기평가활동을 통한 피드백 제공은 학생들의 반성적 사고에 영향을 미칠 수 있다.

이와 같이 동료평가 및 자기평가와 같은 학생 평가활동이 학습에 긍정적인 영향을 미친다는 연구결과들이 있지만, 학생들은 동료가 제공하는 피드백의 질을 의심하거나(Li, Steckelberg & Srinivasan, 2008), 평가자의 능력에 대해 부정적인 인식을 가지고 있는 경우도 있으며(Brindley & Scofield, 1998), 자기평가에서는 자신의 능력을 높이 평가하는 경향이 나타나기도 한다(Hwang *et al.*, 2001; Zoller, Fastow, & Lubezky, 1997; Zoller, 1999). 동료평가활동이 학생들의 논의 능력을 향상시킬 수 있다(Lee, Bak, & Nam, 2015)는 연구결과

도 있지만, 평가활동에서 학생들이 평가자와 피평가자의 역할을 수행하는 과정이 어떻게 학생의 학습 향상에 기여하는지 분명하지 않다. 그리고 학생 평가활동이 학습에 대하여 어떤 가치가 있는지에 대한 증거가 부족하므로(Li, Liu, & Steckelberg, 2010), 이에 대한 심층적인 연구가 필요하다. 또한 학생 평가활동에 의해 제공되는 피드백의 효과는 피드백의 내용, 형태 그리고 기능에 따라 달라질 수 있으며(Narciss, 2004), 이러한 피드백의 내용은 학생들의 학습과 활동에 결정적인 영향을 미칠 수 있다(Cho & MacArthur, 2010). 학생 평가활동에서 사용된 평가 기준은 수업에서 평가자와 피평가자의 수준 높은 활동을 이끌 수 있기 때문에 올바른 평가 기준이 제공되어야 하며(Prins, Sluijsmans, & Kirschner, 2006), 학생 평가활동에서 적용된 평가 기준과 피드백 내용에 대해서도 타당성과 신뢰성을 분석하여 피드백의 질을 검토할 필요가 있다(Strijbos, Van Goozen, & Prins, 2012; Gielen & De Wever, 2013).

따라서 이 연구에서는 논의기반 탐구 과학수업에 동료평가활동과 자기평가활동을 포함하는 학생 평가활동을 적용하였을 때 이러한 학생 평가활동이 학생들의 반성적 사고에 어떻게 영향을 미치는지 알아보는 것을 목적으로 하였다. 이를 위해 논의기반 탐구 과학수업에서 작성한 반성 글쓰기를 분석하여 활동이 진행됨에 따라 학생들의 반성적 사고가 어떻게 변화하는지 분석하였다. 그리고 학생 설문조사와 인터뷰를 실시하여 평가 결과의 타당성과 활용정도를 분석하였으며, 이와 같은 평가활동에 대한 학생들의 인식에 대한 탐색을 통해 학생 평가활동의 영향을 알아보고자 하였다.

II. 연구 방법

1. 연구 참여자

이 연구는 광역시에 위치한 인문계 고등학교 1학년 6개 학급의 학생 166명을 대상으로 하였다. 연구 대상 학교는 과학중점학교로 지정되어 2학년과 3학년은 인문과정반과 과학중점반으로 분리되어 교육과정이 운영되고 있다. 1학년은 일반과정과 과학중점과정으로 학급은 분리되어 있으나 동일한 교육과정으로 운영되고 있다. 이 연구에서는 1학년 학생 중 과학중점반 6학급(166명)을 연구 대상자로 선정하였으며, 3개 학급 84명은 실험집단으로, 다른 3개 학급 82명은 비교집단으로 선정하였다. 두 집단 모두 남학생으로 구성되어 있다. 2014년 7월부터 한 학기동안 5개 주제의 논의기반 탐구 과학수업을 실시하였다. 실험집단에는 동료평가 및 자기평가를 포함하는 학생 평가활동을 적용한 논의기반 탐구 과학수업을, 비교집단에서는 학생 평가활동을 적용하지 않은 논의기반 탐구 과학수업을 실시하였다.

연구에서 논의 기반 탐구 과학수업을 적용한 교사는 인문계 고등학교에 재직 중인 교직 경력 9년의 교사로서 사범대학에서 화학교육을 전공하였으며, 과학교육 전공 박사과정에 재학 중이었다. 또한 이 연구에 참여하기 전 3년의 논의 기반 과학 탐구수업 적용 경험을 가지고 있다.

2. 학생 평가활동을 적용한 논의기반 탐구 과학수업(ABI) 프로그램

논의기반 탐구 과학수업(Argument-Based Inquiry, ABI)은 학생들의 과학적 탐구와 개념이해를 향상시키고 메타인지를 촉진하기 위해

논의와 글쓰기를 활용한 학습전략으로, Keys *et al.*(1999)이 개발한 수업전략인 탐구적 과학 글쓰기(Science Writing Heuristic, SWH) 수업을 우리나라의 교육실정에 맞게 재구성한 프로그램이다(Nam *et al.*, 2008). 이 프로그램은 의문만들기, 실험설계 및 수행, 관찰, 주장과 증거, 읽기, 반성하기의 6단계로 구성되어 있다(Keys *et al.*, 1999; Nam *et al.*, 2008).

논의기반 탐구 과학수업(Argument-Based Inquiry, ABI) 1개 주제에 대한 활동의 모든 단계를 수행하는데 평균 2차시(50분)가 소요된다. 따라서 이 연구에서 5개의 주제의 논의기반 탐구 과학수업을 수행하는데 총 10차시가 소요되었으며, 수업은 일주일에 1차시씩 진행하였다. 개발된 프로그램 중 일부는 실험 수행 가능 여부에 따라 A유형과 B유형으로 나누어 개발하였다. A유형은 논의기반 탐구 과학수업의 모든 단계를 수행하는 프로그램이고, B유형은 실험 설계 단계를 생략하고 제시된 재료로 분자의 구조를 제작하는 과정을 실험 수행단계로 대체한 프로그램이다.

논의기반 탐구 과학수업에 적용할 활동 주제는 고등학교 1학년 학생들이 과학 교과에서 학습하는 내용 중에서 5개의 주제를 선정하였다. 교육과정 상의 진도 계획을 고려하여 서로 관련성이 있는 주제들을 선택하여 프로그램을 개발하였다. 프로그램 개발을 위해 2개의 주제는 분자의 구조와 극성과 관련한 주제를 선정하였으며, 3개의 주제는 화학반응속도와 관련한 주제를 선정하였다(Table 1). 개발된 프로그램은 과학 교육 전문가 1명, 과학 교육 박사과정 3명, 과학 교육 석사과정 3명에게 타당도를 검증받았다.

2014년 7월부터 한 학기 동안 총 5개 주제에 대해 실험집단에는

Table 1. Topics of Argument-based inquiry activities

Topic	Type	Period
분자의 극성	B	2시간
분자의 구조	B	2시간
화학반응속도와 온도와의 관계	A	2시간
화학반응속도와 농도와의 관계	A	2시간
화학반응속도와 표면적과의 관계	A	2시간



Figure 1. Argument-based inquiry applying students' assessment

모둠의 주장과 증거에 대한 동료평가 및 자기평가활동을 포함하는 학생 평가활동을 적용한 논의기반 탐구 과학수업을 처치하였으며, 비교집단에는 학생 평가활동 없이 기존의 논의기반 탐구 과학수업을 처치하였다. 수업은 세 명 혹은 네 명으로 이루어진 모둠을 구성하여 이루어졌으며 학생 평가활동을 적용한 논의기반 탐구 과학수업 수행 과정은 다음과 같이 이루어졌다(Figure 1).

학생 평가활동을 적용한 논의기반 탐구 과학수업은 기존의 논의기반 탐구 과학수업에서 주장과 증거 단계에 동료평가와 자기평가를 포함하는 학생 평가활동을 추가하였다. 동료평가는 다른 모둠의 학생들이 작성한 주장과 증거를 평가하는 것을 말하며, 자기평가는 자기 모둠에서 작성한 주장과 증거를 다른 모둠에서 작성한 주장과 증거와 비교하여 평가하는 것을 말한다. 따라서 실험집단 학생들은 주장과 증거 단계에서 작성한 각 모둠의 주장과 증거가 칠판에 게시되면 학급 토론 전에 다른 모둠의 주장과 증거에 대해 평가하는 동료평가와 자기 모둠의 주장과 증거에 대해서 평가하는 자기평가 과정을 수행하였다.

학생 평가활동은 ‘등급 정하기’, ‘평가 근거 작성하기’, ‘피드백 제공하기’의 세 가지 과정으로 구성하였다(Figure 2).

첫 번째 ‘등급 정하기’ 과정은 논의기반 탐구 과학수업의 주장과 증거 단계에서 모둠의 주장과 증거가 완성된 이후에 실시된다. 주장과 증거 단계에서 완성된 모둠의 주장과 증거 자료는 A3 용지에 작성하여 칠판에 게시한다. 이 자료는 전체 논의에서 발표 자료로 이용되기도 하며 학급의 주장과 증거를 결정할 때 의견을 공유하는 데에 사용된다. 학생들은 게시된 다른 모둠의 주장과 증거를 읽고 이 중에서 가장 우수한 모둠과 가장 부족하다고 생각되는 모둠을 한 모둠씩 선정하는데 이 과정이 ‘등급 정하기’ 과정이다. 제한된 시간 안에 모든 팀을 평가하기에 어려움이 있으며, 만약 한 팀에 대해서 좋은 점과 아쉬운 점을 평가한다면 칠판에 게시된 각 모둠의 모든 자료를 읽지 않을 수 있으므로 칠판에 게시된 자료 중에서 가장 우수한 모둠 한 팀과 가장 부족한 모둠 한 팀을 선정하도록 하였다.

두 번째 단계인 ‘평가 근거 작성하기’ 과정에서는 학생들이 활동 보고서에 자신이 가장 우수한 모둠과 가장 부족한 모둠을 선정한 근거를 작성하는 과정으로, 주장 및 근거자료의 적절성과 타당성에 대한 평가 근거를 작성하였다. 또한 가장 우수하다고 선택한 모둠과 가장 부족하다고 선택한 모둠과 비교했을 때 자신의 모둠은 이 두 모둠과 어떻게 비교되는지 자기 모둠의 주장과 증거자료를 평가하였다.

마지막으로 ‘피드백 제공하기’ 과정은 평가 자료를 공유하기 위한 과정으로 즉각 피드백과 지연 피드백의 두 단계로 이루어졌다. 즉각 피드백은 학급 논의과정에서 이루어지며, 실험결과를 바탕으로 작성한 모둠의 주장과 증거에 대해서 한 팀이 발표를 하면 실험 결과에서 오차나 의문사항에 대해 질의·응답하는 과정을 가졌다. 지연 피드백은 학생들이 활동 보고서에 작성한 평가 결과를 교사가 모두 수집하고 정리하여 다음 활동을 시작하기 전 학생들에게 배부하였다. 교사는 학생들이 다른 모둠의 학생들이 평가한 자료와 자기 모둠의 학생들이 평가한 자료를 받고 이를 피드백 자료로 활용하여 이전 활동에서 지적받은 문제점들을 개선할 수 있도록 지도하였다. 지연 피드백 과정은 학급 토론에서 제한된 시간으로 인해 모든 학생들이 발표 및 질의·응답에 참여할 수 없다는 단점을 보완하는 방법이 될 수 있다.

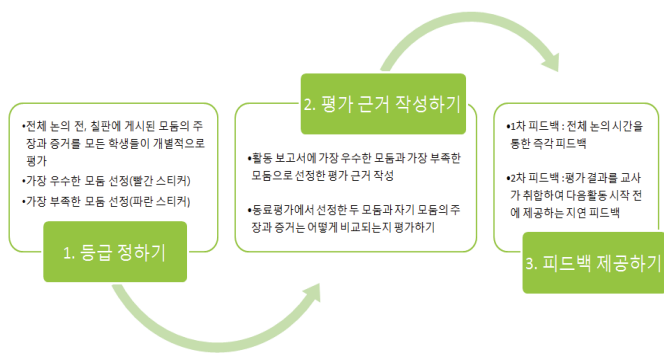


Figure 2. Stage of students' assessment

이상과 같이 실험집단에서는 동료평가 및 자기평가를 포함하는 평가활동을 실시하였으며, 비교집단에서는 이와 같은 평가 단계만 생략하고 모든 과정은 실험집단과 동일하게 실시하였다.

3. 자료 수집

학생 평가활동을 적용한 논의기반 탐구 과학수업에서 평가활동이 학생들의 반성적 사고에 미치는 효과를 알아보기 위해 논의기반 탐구 과학수업의 마지막 단계에서 작성한 반성 글쓰기 자료를 수집하였다. 또한 평가활동에 대한 학생들의 심층적인 인식을 알아보기 위해 논의기반 탐구 과학수업 후에 실험집단 학생들을 대상으로 설문조사와 인터뷰를 실시하였다.

반성 글쓰기는 논의기반 탐구 과학수업의 반성 단계에서 작성한 글쓰기이다. 반성 글쓰기를 통해 학생들은 수업 전 과정에 대해 점검하고, 학습한 내용과 활동 전 후의 생각변화를 점검하고 생각을 변화시킨 요인에 대해 되돌아본다. 활동이 진행됨에 따라 학생들의 반성적 사고에 어떠한 변화가 나타나는지 알아보기 위해 총 다섯 개 주제의 논의기반 탐구 과학수업의 활동 보고서 중 활동 초기, 활동 중기, 활동 후기라고 간주되는 1차, 3차, 5차 활동 보고서의 반성 글쓰기 자료를 수집하여 분석하였다.

설문조사는 학생 평가활동을 적용한 논의기반 탐구 과학수업을 실시한 실험집단 학생 84명을 대상으로 마지막 프로그램을 끝낸 이후에 실시하였다. 설문지 문항은 Likert 5점 척도 문항, 자유서술형 문항으로 구성되었으며, 설문조사를 통해 학생 평가활동의 타당성 및 평가활동의 결과가 피드백으로써 다음 활동에 어떻게 영향을 미쳤는지 알아보았다. 학생 평가활동에 대한 인식조사로 평가활동이 자신의 학습에 미치는 긍정적인 영향, 평가 결과의 타당성 여부, 수궁할 수 없는 평가 결과 등에 대해 질문하였고, 평가결과의 활용 정도를 알아보기 위해 평가 결과는 이후 활동에서 어떻게 활용되었는지 알아보았다.

설문조사로 나타나지 않은 학생들의 심층적인 생각을 알아보기 위해 인터뷰를 실시하여 분석하였다. 인터뷰에 참여한 학생은 실험집단 84명의 학생 중 학업성취도를 기준으로 상위권, 중위권, 하위권에 해당하는 학생을 각 학급당 2-3명씩 선발하여 총 20명의 학생을 대상으로 실시하였다. 인터뷰 문항은 설문지 문항과 동일하게 구성하였으며, 논의기반 탐구 과학수업을 진행한 교사와 학생이 개인별로 만나서 반구조화 된 형태로 진행되었고 시간은 개인당 약 10~20분 정도 소요되었다. 학생 설문지와 인터뷰 문항은 연구자가 초안을 작성하고 과학 교육 전문가 1인과 협의를 통해 수정하여 개발되었다.

4. 자료 분석

학생 평가활동이 학생들의 반성적 사고에 미치는 영향을 알아보기 위해 논의기반 탐구 과학수업의 마지막 단계에서 작성한 반성 글쓰기를 분석하였다. 평가들은 반성적 사고를 수준별 유형으로 분류한 Hatton & Smith(1995)의 분석틀을 기초로 하여 Sung *et al.*(2012)이 개발한 반성 글쓰기 평가틀을 수정하여 사용하였다. 이 연구에서는 Sung *et al.*(2012)이 개발한 평가틀의 하위요소인 탐구과정 분석, 학습의 이해, 생각의 변화, 의문 제기 4개 요소에 반성적 실천 요소를 추가하여 사용하였으며, 각 요소별 반성 수준은 Sung *et al.*(2012)의 평가틀을 기초로 하여 4단계의 반성 수준으로 분류하여 분석하였다. 반성 글쓰기에서 증거 없이 단순히 자신의 생각을 서술한 경우를 1수준, 증거를 제시하여 자신의 생각을 서술한 경우를 2수준, 증거와 함께 자신의 생각을 제시하며, 과학적 수준의 문제 해결 방안이나 과학적 개념을 함께 제시하는 경우를 3수준으로 분류하였다.

탐구과정 분석은 수업에서 학생들이 수행했던 탐구과정을 설명하고 탐구과정에서 발생했던 오류와 원인을 점검하는 요소이다. 따라서 탐구과정을 설명하고 과정상의 오류와 원인을 분석하여 해결방안을 제시하는 경우를 3수준으로 정하였다. 탐구과정을 설명하고 과정상의 오류와 원인은 제시하였지만 해결방안은 제시하지 않은 경우는 2수준, 탐구과정을 나열하거나 과정상의 오류만 제시하는 경우는 1수준, 탐구과정에 대한 설명이 나타나지 않으면 0수준으로 분류하였다.

학습의 이해는 수업을 통해 학습한 내용을 점검하는 요소로 학습한 내용을 탐구과정과 과학적 원리의 자료를 근거로 설명하면 3수준, 탐구과정 또는 과학적 원리의 자료 중 한 가지만 근거로 설명하는 경우는 2수준, 학습 내용을 단순히 설명하는 경우는 1수준, 학습 내용을 점검하지 않으면 0수준으로 분류하였다.

생각의 변화는 수업을 통해 자신의 생각이 어떻게 변화하였는지를 점검하는 요소로 처음에 가졌던 생각은 무엇이고 어떻게 변화하였는지를 점검하며 생각이 변화된 근거를 제시하면서 정교화된 나중의 생각을 기술하는 경우 3수준으로 분류하였다. 또한 처음의 생각과 나중에 생각이 어떻게 변화하였는지 근거를 제시하고 설명하는 경우에는 2수준, 근거를 제시하지 않고 처음의 생각과 나중의 생각만 제시하는 경우는 1수준, 생각의 변화가 나타나지 않는 경우는 0수준으로 분류하였다.

의문 제기는 실험 결과나 학습 내용과 관련하여 의문을 제기하면서 과학적 수준의 해결방안을 제시하면 3수준, 개인적 수준의 해결방안을 제시하면 2수준, 단순히 의문만 제기하는 경우는 1수준, 의문을 제기하지 않으면 0수준으로 분류하였다.

반성적 실천은 학습자가 이전 활동에서 반성한 내용을 다음 활동에 실천으로 옮겨 행동의 변화가 나타났는지를 확인하는 요소이다. 따라서 이전 활동에서 자신의 행동에 대한 문제점을 언급하고 어떻게 개선하였는지를 설명하고 있는 경우 3수준으로 분류하였다. 자신의 행동에 대한 문제점을 언급하고, 다음 활동에 개선해야 할 구체적인 해결방안을 제시하면 2수준, 활동에서 단순히 자신의 행동에 대한 반성만 제시하면 1수준, 반성적 실천이 나타나지 않으면 0수준으로 분류하였다.

과학 학습에서 실험을 계획 및 수행하고 탐구과정을 점검하는 과정은 메타인지적 활동에 속한다(Baird *et al.*, 1991). 학생들은 반성에

서 학습 과정을 점검하고 평가하며 계획하는 메타인지 과정과 학습 내용을 확인하고 기존의 지식 및 경험과 연관지어 생각하는 정교화 과정이 요구되므로 메타인지 수준이 높을수록 고차원적인 반성 활동을 수행할 수 있다(Heo, 2011). 따라서 이 연구에서는 반성 글쓰기에 서 나타나는 반성적 사고 수준을 메타인지 수준으로 재분류하여 분석하였다.

반성 글쓰기에서 나타나는 메타인지 수준의 분류에서 반성적 사고 0수준과 1수준의 경우, 단순히 자신의 생각을 나열한 비반성적 과정이므로(Hatton & Smith, 1995) 자신의 인식에 대해 점검하고 조정하는 과정이 나타나지 않기 때문에 메타인지 수준의 반성이라고 볼 수 없다. 반성적 사고 2수준의 경우 자신이 왜 그렇게 생각하는지에 대한 이유와 증거를 제시하여 자신의 생각을 서술한 것이고, 반성적 사고 3수준의 경우 과학적 수준의 문제 해결 방안을 함께 제시한 것이므로 자신이 가지고 있는 인식에 대해 모니터링하고 평가하여 조정하는 메타인지적 활동이 일어났다고 볼 수 있다. 따라서 반성적 사고 2수준과 3수준을 메타인지 수준의 반성으로 분류하였다. 특히 3수준은 자신의 인식에 대해 모니터링하고 평가한 이후 결과를 바탕으로 인식을 조정할 것이라고 할 수 있으므로 2수준보다 더 높은 메타인지 수준이라고 할 수 있다. 그러므로 반성적 사고 2수준은 낮은 메타인지 수준의 반성, 반성적 사고 3수준은 높은 메타인지 수준의 반성으로 분류하였다.

논의기반 탐구 과학수업에서 학생 평가활동에 대한 탐색을 위해 논의기반 탐구 활동 후 실험집단 학생들을 대상으로 실시한 설문조사와 인터뷰를 분석하였다. 설문지는 학생 평가활동에 대한 인식을 조사하기 위해 평가활동이 학습에 도움을 주었는지에 대한 Likert 척도 문항과 구체적인 사례를 서술하는 자유서술형 문항으로 구성되었다. 또한 동료평가와 자기평가활동의 타당성과 활용 정도를 알아보기 위한 Likert 척도 문항, 평가결과의 구체적인 사례를 서술하는 자유서술형 문항으로 구성하였으며, 동료평가활동에 대해서는 추가적으로 수급이 되지 않았던 결과의 사례를 자유서술형으로 작성하도록 하여 분석하였다. Likert 5점 척도 문항은 전혀 그렇지 않다 1점, 그렇지 않다 2점, 보통이다 3점, 그렇다 4점, 매우 그렇다 5점으로 하여 평균 점수를 구하였다. 자유서술형 문항은 응답을 유형별로 범주화하여 세부요소를 추출하고, 요소별 빈도수를 측정하였다. 범주화된 요소들은 과학 교육 전문가 1명에게 타당성을 검증받고 수정하였으며, 최종 범주화된 요소를 바탕으로 설문조사지를 다시 분석하여 세부요소에 대한 빈도수를 결정하였다. 인터뷰는 전사하여 분석하였으며 응답을 유형별로 범주화하여 세부요소를 추출하였다.

모든 검사지와 평가들은 과학 교육 전문가 1명과 과학 교육 박사과정 3명, 석사과정 5명으로부터 내용타당도를 검증받았다. 평가들의 타당도와 신뢰도를 높이기 위해 분석자 6명이 무작위로 20명의 반성 글쓰기를 채점한 후 협의과정을 거쳐 분석틀을 수정하였다. 수정된 분석틀로 과학 교육 박사과정 2명과 석사과정 4명이 각각 두 반씩 채점하여 하나의 검사지마다 두 명의 분석자가 채점하였으며, 채점자 간에 점수 차이가 있는 경우 합의에 이를 때까지 지속적으로 협의하여 분석의 신뢰도를 높이고자 하였다.

이 연구에서 수집한 자료는 SPSS WIN 20.0 통계패키지 프로그램을 사용하여 정량분석을 하였고, 공변량 분석(ANCOVA), t-test로 통계 처리하였다. 반성 글쓰기는 반성적 사고 변화를 분석하기 위해 분석하였으며, 집단 간의 초기 반성적 사고 수준을 분석하기 위해

1차 반성 글쓰기를 독립 표본 t-test를 통해 두 집단의 평균 차이를 검증하였다. 중기와 후기에 해당하는 3차와 5차 반성 글쓰기는 1차 반성 글쓰기 값을 공변량(covariate)으로 하여 일원 공변량 분석(one-way ANCOVA)을 하였다.

설문조사와 인터뷰에 대한 분석의 신뢰도를 높이기 위해 응답별로 추출한 사례를 바탕으로 하여 모든 분석자들이 공동으로 1차 범주화를 실시하였고, 1차로 범주화된 내용을 바탕으로 하여 전체 설문 자료와 인터뷰 자료를 2차 범주화하였다.

III. 연구 결과

이 연구는 논의기반 탐구 과학수업에서 동료평가 및 자기평가를 포함하는 학생 평가활동이 고등학생들의 반성적 사고에 미치는 영향에 대해 알아보기 위해 학생들이 작성한 반성 글쓰기를 분석하였으며, 평가활동에 대한 학생들의 인식을 탐색하기 위해 설문지 및 인터뷰 자료를 분석하여 결과를 제시하였다.

1. 학생 평가활동이 반성적 사고 변화에 미치는 영향

논의기반 탐구 과학수업에서 활동이 진행됨에 따라 학생 평가활동을 적용한 집단과 적용하지 않은 집단의 반성적 사고 수준을 비교하기 위해 학생들이 작성한 총 다섯 개의 반성 글쓰기 중 1차, 3차, 5차 반성 글쓰기를 분석하였다. 반성적 실천은 이전 활동에서 반성한 내용을 다음 활동에서 어떻게 개선되었는지를 확인하는 요소이므로 1차 반성 글쓰기에 대한 반성적 실천은 분석에서 제외하였다. 따라서 반성적 실천 요소는 반성 글쓰기 요소별 분석에서 별도로 제시하였다.

가. 반성 글쓰기 활동별 분석 결과

학생들이 작성한 반성 글쓰기 분석 결과는 반성 글쓰기 평가들의 하위요소인 생각의 변화, 탐구과정 분석, 학습의 이해, 의문 제기의 수준을 점수화하여 총점을 제시하였다. 1차 반성 글쓰기의 총점과 각 요소별 평균 점수는 독립표본 t-test로 분석하여 검증하였으며, 3차와 5차 반성 글쓰기의 총점과 각 요소별 평균 점수는 두 집단의 차이를 알아보기 위해 첫 번째 반성 글쓰기의 점수를 공변량으로 하여 일원 공변량 분석(one-way ANCOVA)을 실시하였다.

1차 반성 글쓰기 분석 결과, 실험집단의 총점 평균은 0.9점, 비교집단의 총점 평균 1.0점으로 비교집단이 실험집단에 비해 높았으나 유의미한 차이는 나타나지 않았다($t=-.630, p=.530$). 하위요소인 탐구과정 분석($t=.350, p=.727$), 학습의 이해($t=-1.338, p=.183$), 생각의 변화($t=.354, p=.724$), 의문 제기($t=-.464, p=.643$)에서도 실험집단과 비교 집단 사이에 통계적으로 유의미한 차이가 나타나지 않았다(Table 2).

3차 반성 글쓰기 분석 결과, 실험집단의 총점 평균은 3.21점, 비교집단의 총점 평균은 2.09점으로 실험집단이 비교집단에 비해 유의미하게 높게 나타났다($F= 19.008, p=.000***$). 하위요소에 대한 분석 결과, 학습의 이해와 생각의 변화 요소의 평균은 실험집단이 비교집단보다 유의미하게 높게 나타났다($p<.05$). 탐구과정 분석과 의문 제기 요소의 평균은 실험집단이 비교집단에 비해 높았으나 유의미한 차이는 나타나지 않았다(Table 3). 따라서 3차 반성 글쓰기에서는 실험집

Table 2. Analysis of the 1st reflective writing

	Experimental groups		Comparative groups		t	p
	M	SD	M	SD		
탐구과정 분석	.05	.221	.04	.195	.350	.727
학습의 이해	.41	.589	.55	.717	-1.338	.183
생각의 변화	.43	.746	.39	.691	.354	.724
의문 제기	.01	.113	.03	.228	-.464	.643
총점	.90	.942	1.00	1.064	-.630	.530

p>.05 통계적으로 유의미한 차이 없음.

Table 3. Analysis of the 3rd reflective writing

	Experimental groups		Comparative groups		t	p
	M	SD	M	SD		
탐구과정 분석	.90	.816	.68	.849	2.781	.097
학습의 이해	1.26	.877	.81	.692	12.660	.001**
생각의 변화	.96	.901	.55	.867	8.044	.005**
의문 제기	.09	.324	.04	.268	1.657	.200
총점	3.21	1.506	2.09	1.687	19.008	.000***

p<.01, *p<.001

Table 4. Analysis of the 5th reflective writing

	Experimental groups		Comparative groups		t	p
	M	SD	M	SD		
탐구과정 분석	2.18	.734	1.87	.754	6.582	.011*
학습의 이해	2.04	.904	1.57	.789	13.605	.000***
생각의 변화	1.10	1.150	.86	1.092	1.843	.177
의문 제기	.38	.743	.18	.605	3.273	.072
총점	5.72	2.056	4.47	1.929	17.379	.000***

*p<.05, ***p<.001

Table 5. Analysis of inquiry process in reflective writing

Reflective thinking level	Metacognitive level	Experimental groups(%)			Comparative groups(%)		
		1 st	3 rd	5 th	1 st	3 rd	5 th
0	.	95	36	0	96	52	0
1	.	5	41	19	4	32	36
2	low-metacognition	0	21	43	0	12	42
3	high-metacognition	0	2	38	0	4	22

단의 반성적 사고 수준이 비교집단의 반성적 사고 수준보다 높으며, 이러한 차이는 학습의 이해와 생각 변화의 차이에서 오는 것으로 볼 수 있다.

5차 반성 글쓰기 분석 결과, 실험집단의 총점 평균은 5.72점, 비교집단의 총점 평균은 4.47점으로 실험집단이 비교집단에 비해 유의미하게 높게 나타났다(F= 17.379, p=.000***). 하위요소에 대한 분석 결과, 탐구과정 분석과 학습의 이해 요소의 평균은 실험집단이 비교집단보다 유의미하게 높게 나타났다(p<.05). 생각의 변화와 의문 제기의 평균은 실험집단이 비교집단보다 높게 나타났지만 유의미한 차이

는 나지 않았다(Table 4). 이를 종합해 보면, 5차 반성 글쓰기에서도 실험집단의 반성적 사고 수준이 비교집단의 반성적 사고 수준보다 높은 것으로 볼 수 있다. 그러나 5차 반성 글쓰기의 실험집단과 비교집단의 반성적 사고 수준의 차이는 3차 반성 글쓰기에서와는 달리 탐구과정과 학습의 이해에서의 차이에서 오는 것으로 볼 수 있다.

나. 반성 글쓰기 요소별 분석 결과

논의기반 탐구 과학수업에서 활동이 진행됨에 따라 학생 평가활동을 실시한 실험집단과 실시하지 않은 비교집단 간의 요소별 반성적 사고 수준 변화를 비교하기 위해 하위요소에 대한 수준별 학생 비율 변화를 분석하였다. 또한 반성 글쓰기에서 나타나는 반성적 사고 수준을 메타인지 수준으로 재분류하여 분석하였다. 반성 글쓰기에서 하위요소별 분석 결과는 다음과 같다.

1) 탐구과정 분석

탐구과정 분석은 수업에서 학생들이 수행했던 탐구과정을 설명하고 탐구과정에서 발생했던 오류와 원인을 점검하여 개선 방안을 제시하는 요소이다.

분석 결과, 1차 반성 글쓰기에서 반성적 사고 0수준 학생이 실험집단 95%, 비교집단 96%로 대부분을 차지했으며, 반성적 사고 1수준 학생은 실험집단 5%, 비교집단 4%로 낮은 비율로 나타났다. 두 집단 모두 메타인지 수준의 반성적 사고 2수준 및 3수준의 학생은 나타나지 않았다(Table 5).

3차 반성 글쓰기에서 반성적 사고 0수준 학생은 실험집단 36%, 비교집단 52%로 실험집단이 비교집단에 비해 더 낮은 비율을 보였으며, 반성적 사고 1수준 학생은 실험집단에서는 41%, 비교집단에서는 32%로 비교집단에서 더 낮은 비율로 나타났다. 메타인지 수준의 반성에 해당하는 반성적 사고 2수준 및 3수준 학생은 실험집단 23%, 비교집단 16%로 실험집단이 높은 비율을 차지하고 있다. 반성적 사고 2수준(낮은 메타인지 수준)의 학생은 실험집단 21%, 비교집단 12%로 실험집단이 더 높은 비율을 나타냈으나, 반성적 사고 3수준(높은 메타인지 수준)의 학생은 실험집단 2%, 비교집단은 4%로 비교집단이 실험집단에 비해 더 높은 비율을 보였다.

5차 반성 글쓰기에서 두 집단 모두 반성적 사고 0수준의 학생은 나타나지 않았다. 반성적 사고 1수준 학생은 실험집단 19%, 비교집단 36%로 실험집단에서 낮은 비율을 나타냈다. 또한 두 집단 모두 3차 반성 글쓰기에 비해 반성적 사고 1수준 학생 비율이 많이 감소된 것으로 나타났다. 메타인지 수준에 해당하는 반성적 사고 2수준 및 3수준 학생의 비율은 실험집단에서는 81%, 비교집단에서는 64%로 실험집단이 비교집단보다 높은 비율로 나타났다. 특히 반성적 사고 2수준(낮은 메타인지 수준) 학생은 실험집단 43%, 비교집단 42%로 비슷하게 나타났으나, 반성적 사고 3수준(높은 메타인지 수준) 학생은 실험집단 38%, 비교집단 22%로 비교집단에 비해 실험집단에서 높은 비율을 보였다.

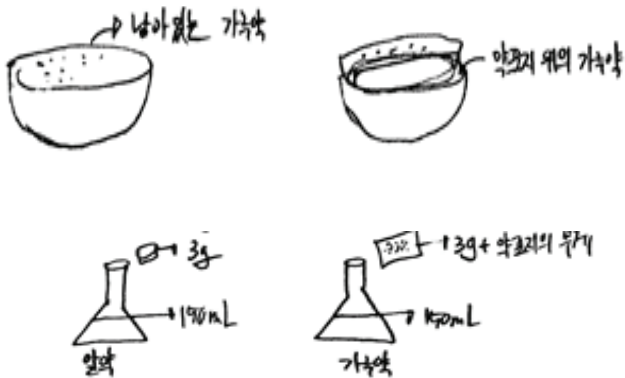
위의 결과로부터, 탐구과정 분석에서 실험집단이 비교집단보다 3차에서 유의미한 차이가 나타나지 않았으나, 5차에서는 유의미하게 높았던 이유를 추측할 수 있다. 3차 반성 글쓰기에서는 반성적 사고

1, 2수준의 학생 수가 실험집단이 비교집단보다 높은 비율을 보였고 탐구과정에 대해 언급하지 않은 0수준의 학생 수는 실험집단이 비교집단보다 낮은 비율을 나타냈으나, 가장 높은 수준인 3수준의 학생은 비교집단이 실험집단보다 높게 나타났기 때문에 두 집단 간의 유의미한 차이가 나타나지 않은 것으로 보인다. 5차 반성 글쓰기에서는 이와는 반대로, 두 집단 모두 반성적 사고 0수준의 학생은 나타나지 않았으며 반성적 사고 2수준의 학생 수는 두 집단 모두 비슷한 비율을 보였으나, 반성적 사고 1수준의 학생 수가 실험집단이 비교집단보다 낮은 대신 반성적 사고 3수준의 학생 수가 실험집단이 비교집단보다 높은 비율로 나타났기 때문에 두 집단 간의 유의미한 차이가 나타난 것으로 보인다. 즉, 활동이 진행될수록 반성적 사고 3수준의 학생 수가 실험집단이 비교집단보다 크게 증가하였기 때문에 두 집단 간의 반성 수준의 차이가 나타난 것이라고 할 수 있다.

다음은 실험집단과 비교집단의 반성 글쓰기에서 탐구과정 분석 3수준에 대한 사례이다(사례1).

<사례1>

ST1: 실험이 시작되고 우리는 발포정을 가루약과 알약 형태로 물 속에 넣어 시간을 측정해야 했다. 이 부분에서 우리는 오차가 발생하였음을 알게 되었다. 알약을 부수고 막자사발에 붙어 있는 가루가 있었던 것이다. 이 오차는 막자사발 위에 약포지를 올려 놓고 알약을 부숨으로써 해결할 수 있었다. 이 오차를 해결하고 난 뒤에도 또 다른 오차가 바로 발생하였다. 바로 가루약이 떨어질 때 걸리는 시간에 반응이 일어나므로 이 오차는 약포지 채로 넣어 흔들어 보기로 해보았고 실험은 성공적으로 끝났다. 다시 생각해 보니 가루약과 함께 들어간 약포지의 무게가 오차가 될 수 있다는 것을 알았다. 우리 모두는 알약일 때는 약포지를 깜박하고 넣지 않았다. (실험집단)



ST2: 세 번째 실험에서 발포정을 뺀 가루로 만든 후 약포지에 싸서 플라스크에 넣고 흔들어 반응시키는 방법을 사용했다. 이 방법이 지닌 큰 문제는 흔들었을 때 약포지에 있는 가루가 모두 퍼지지 않을 수 있다는 것이다. 차라리 꽃봉오리 모양으로 접어서 흔들는 게 더 잘 퍼지지 않을까 하는 생각이 든다. 또 가루라고 해서 생기는 기체의 양이 달라지지 않을 것이라고 생각했는데 오히려 크게 차이가 났다. (비교집단)

사례1에서, 학생들은 실험에서 발생한 오차를 언급하고 오차가 발생한 원인에 대하여 자신들의 생각을 제시한 후 이를 해결할 수 있는 구체적인 방안을 제시하고 있다. 실험집단의 경우 실험에서 발생한 오차에 대해 여러 가지를 제시하고 이에 따른 원인과 해결방안을 각각 제시하고 있으며, 그림을 그려서 설명을 추가하기도 하였다. 그러나 비교집단에서는 동일한 3수준이지만 실험에서 발생한 오차를 하나만 제시하고 원인과 해결방안을 제시하는 경우가 많았다. 따라서 탐구과정 분석에 대한 반성적 사고 수준은 비교집단보다 실험집단이 높은 것을 알 수 있다.

2) 학습의 이해

학습의 이해는 반성 글쓰기에서 수업을 통해 학습한 내용을 학습자가 점검하는 요소이다.

분석 결과, 1차 반성 글쓰기에서 반성적 사고 0수준의 학생은 실험집단 65%, 비교집단 56%로 높은 비율을 보였으며, 반성적 사고 1수준의 학생은 실험집단 30%, 비교집단 37%로 실험집단이 비교집단보다 더 낮은 비율로 나타났다. 메타인지 수준의 반성적 사고 2수준 및 3수준의 학생은 실험집단 5%, 비교집단 7%로 두 집단 모두 낮은 비율로 나타났으나 비교집단이 실험집단보다 더 높은 비율을 보였다.(Table 6).

3차 반성 글쓰기에서 두 집단 간의 반성적 사고 수준에 유의미한 차이가 나타났다. 반성적 사고 0수준의 학생은 실험집단 23%, 비교집단 33%로 두 집단 모두 1차 반성 글쓰기에 비해서 많이 감소하였다. 반성적 사고 1수준의 학생은 실험집단 32%, 비교집단 54%로 비교집단이 실험집단보다 더 높게 나타났다. 그러나 메타인지 수준에 해당하는 반성적 사고 2수준 및 3수준 학생의 비율은 실험집단 45%, 비교집단 13%로 실험집단이 비교집단에 비해 훨씬 높은 비율을 차지하고 있다. 반성적 사고 2수준(낮은 메타인지 수준) 학생은 실험집단 40%, 비교집단 12%로 큰 차이를 보였으며, 반성적 사고 3수준(높은 메타인지 수준) 학생은 실험집단 5%, 비교집단 1%로 비교집단에 비해 실험집단에서 높은 비율을 보였으나 두 집단 간의 큰 차이는 없다. 따라서 3차 반성 글쓰기에서 학습의 이해에 대하여 두 집단 간의 유의미한 차이는 낮은 메타인지 수준에 해당하는 반성적 사고 2수준의 학생 비율이 비교집단보다 실험집단이 높았던 것과 반성적 사고 0수준과 1수준의 학생 비율이 비교집단보다 실험집단이 낮았던 것 때문인 것으로 보인다.

5차 반성 글쓰기에서도 3차 반성 글쓰기와 동일하게 두 집단 간의 반성적 사고 수준에 유의미한 차이가 나타났다. 반성적 사고 0수준의

Table 6. Analysis of understanding of learning in reflective writing

Reflective thinking level	Metacognitive level	Experimental groups(%)			Comparative groups(%)		
		1 st	3 rd	5 th	1 st	3 rd	5 th
0	.	65	23	6	56	33	6
1	.	30	32	19	37	54	43
2	low-metacognition	5	40	38	5	12	39
3	high-metacognition	0	5	37	2	1	12

학생은 두 집단 모두 6%로 낮은 비율을 보였다. 반성적 사고 1수준의 학생은 실험집단 19%, 비교집단 43%로 실험집단이 비교집단보다 훨씬 낮은 비율로 나타났다. 두 집단 모두 3차 반성 글쓰기에 비해 반성적 사고 1수준의 학생 비율이 감소하였음을 알 수 있다. 메타인지 수준에 해당하는 반성적 사고 2수준 및 3수준 학생의 비율은 실험집단 75%, 비교집단 51%로 실험집단이 비교집단보다 더 높은 비율을 보였으며, 두 집단 모두 3차 반성 글쓰기에 비해 학생 비율이 증가하였음을 알 수 있다. 특히 반성적 사고 2수준(낮은 메타인지 수준) 학생은 실험집단 38%, 비교집단 39%로 비슷하게 나타났으나, 반성적 사고 3수준(높은 메타인지 수준) 학생은 실험집단 37%, 비교집단 12%로 비교집단에 비해 실험집단에서 높은 비율을 보였다. 따라서 5차 반성 글쓰기에서 학습의 이해에 대하여 두 집단 간의 유의미한 차이는 반성적 사고 1수준의 학생 비율이 비교집단보다 실험집단이 낮고, 반성적 사고 3수준의 학생 비율이 비교집단보다 실험집단이 높았기 때문인 것으로 보인다.

다음은 실험집단과 비교집단의 반성 글쓰기에서 학습의 이해 3수준에 대한 사례이다(사례2).

<사례2>

ST1: 이번에 알게 된 사실은 표면적이 넓을수록 반응속도가 빨라진다는 것이다. 이 번 실험을 포함해 반응속도에 관한 실험을 3개나 했다. 한 개는 온도, 또다른 것은 농도, 마지막으로 표면적에 관한 실험이었는데 온도가 높을수록, 농도가 진할수록, 표면적이 넓을수록 반응속도가 빨랐다. 이 3가지 실험들의 공통점이 있는데 그것은 바로 분자끼리 충돌횟수가 상승한다는 것이다. 여기서 충돌횟수가 증가할수록 반응속도가 증가하는 것이다. 또 활성화에너지에 대해서도 알게 되었다. 이 활성화에너지를 넘어가야지만 반응을 시작하는 것이다. 또 마지막으로 유효충돌인데 위에서 충돌횟수가 증가할수록 유효충돌이 많아져 반응을 일으키는 것이다. 즉, 유효충돌이란 반응이 일어날 수 있는 방향으로 입자들이 충돌해서 화학반응을 일으킬 수 있는 분자끼리의 충돌을 말한다. 이것으로 화학반응을 일으키기 위해서는 활성화에너지보다 에너지가 많고 유효충돌이 일어나야지만 화학반응을 일으키고 유효충돌이 빨리 많이 일어나면 반응속도가 빨라지는 것 같다. (실험집단)

ST2: 가루낸 발포정은 가루내지 않은 발포정에 비해서 입자의 크기가 작으므로 표면적이 커진다. 표면적이 커지게 되면 반응물질간의 접촉 면적이 커지면 반응물질 입자간의 충돌횟수가 증가한다. 충돌횟수가 증가하게 되면 반응이 일어날 수 있는 유효충돌이 일어나기 더 쉬워지기 때문에 반응이 더 빨리 일어나기 쉬워져서 반응속도가 더 빨라지게 된다. 따라서 표면적이 클수록 반응속도가 빨라진다. (비교집단)

사례2에서 비교집단의 경우 실험에서 가루약이 알약보다 반응속도가 빨랐던 실험결과와 충돌횟수의 과학 개념으로 표면적이 클수록 반응속도가 빨라지는 것을 설명하고 있다. 그러나 실험집단의 경우 이번 활동을 통해 알게 된 표면적과 반응속도의 관계뿐만 아니라 이전 활동에서 학습한 온도, 농도와 반응속도의 관계에 대한 실험결과도 제시하며, 활성화에너지와 유효충돌의 과학적 개념을 이용하여 세 가지의 실험 결과를 설명하고 있다. 따라서 실험집단은 비교집단

에 비해 입자의 크기가 작을수록 표면적이 커지는 이유나 이전 활동과 연관 지어 결과를 제시하는 등 학습한 내용을 더욱 다양하게 제시하였다. 따라서 학습의 이해에 대하여 실험집단 학생이 비교집단 학생보다 반성적 사고 수준이 더 많이 향상되었음을 알 수 있다.

3) 생각의 변화

생각의 변화는 수업을 통해 자신의 생각이 어떻게 변화하였는지를 점검하는 요소이다.

분석 결과, 1차 반성 글쓰기에서 반성적 사고 0수준의 학생은 실험집단 68%, 비교집단 73%로 두 집단 모두 높은 비율로 나타났으며, 반성적 사고 1수준의 학생은 실험집단과 비교집단에서 19%의 동일한 비율로 나타났다. 메타인지 수준에 해당하는 반성적 사고 2수준 및 3수준의 학생은 실험집단 13%, 비교집단 8%로 두 집단 모두 낮은 비율로 나타났다. 1차 반성 글쓰기에서 두 집단 간의 수준별 학생 수는 비슷한 양상으로 나타났기 때문에 두 집단 간의 유의미한 차이가 없었던 것으로 보인다(Table 7).

3차 반성 글쓰기에서 두 집단 간의 반성적 사고 수준에 유의미한 차이가 나타났다. 반성적 사고 0수준의 학생은 실험집단 36% 비교집단 64%로 두 집단 모두 1차 반성 글쓰기에서 나타난 학생 비율보다 감소하였으나 실험집단이 비교집단보다 더 많이 감소하였으며, 비교집단은 큰 차이가 없는 것을 볼 수 있다. 반성적 사고 1수준의 학생은 실험집단 38%, 비교집단 23%로 실험집단이 비교집단보다 더 높은 비율로 나타났으며, 두 집단 모두 1차 반성 글쓰기보다 증가한 비율을 보였다. 메타인지 수준에 해당하는 반성적 사고 2수준 및 3수준의 학생은 실험집단 26%, 비교집단에서는 13%로 실험집단이 비교집단보다 높은 비율을 나타냈다. 이 중 반성적 사고 2수준(낮은 메타인지 수준)의 학생 비율은 실험집단 20%, 비교집단 7%로 실험집단이 높게 나타났으며, 반성적 사고 3수준(높은 메타인지 수준)의 학생 비율은 실험집단과 비교집단 모두 6%의 동일한 비율을 보였다. 따라서 생각의 변화에 대하여 두 집단 간의 유의미한 차이가 나타난 이유는 반성적 사고 0수준과 1수준의 학생 비율이 실험집단이 비교집단보다 낮고, 반성적 사고 2수준의 학생 비율은 실험집단이 비교집단보다 높았기 때문인 것으로 보인다.

5차 반성 글쓰기에서 반성적 사고 0수준의 학생은 실험집단 42%, 비교집단 53%로 실험집단이 비교집단보다 낮은 비율로 나타났으나, 3차 반성 글쓰기에서보다 실험집단의 반성적 사고 0수준 학생 비율은 오히려 증가하였고 비교집단은 감소하였다. 반성적 사고 1수준의 학생은 실험집단 25%, 비교집단 19%로 실험집단이 비교집단보다 더

Table 7. Analysis of change of thinking in reflective writing

Reflective thinking level	Metacognitive level	Experimental groups(%)			Comparative groups(%)		
		1 st	3 rd	5 th	1 st	3 rd	5 th
0	.	68	36	42	73	64	53
1	.	19	38	25	19	23	19
2	low-metacognition	12	20	14	7	7	13
3	high-metacognition	1	6	19	1	6	14

높은 비율을 나타냈다. 메타인지 수준에 해당하는 반성적 사고 2수준 및 3수준 학생의 비율은 실험집단 33%, 비교집단 27%로 실험집단이 비교집단보다 높게 나타났으며, 두 집단 모두 3차 반성 글쓰기보다 증가한 것을 알 수 있다. 이 중 반성적 사고 2수준(낮은 메타인지 수준)의 학생은 실험집단 14%, 비교집단 13%로 두 집단에서 유사하게 나타났으나, 반성적 사고 3수준(높은 메타인지 수준)의 학생은 실험집단 19%, 비교집단 14%로 실험집단이 비교집단에 비해 더 높은 비율을 보였으나 차이는 크지 않았다. 따라서 5차 반성 글쓰기에서 두 집단 간에 수준별 학생 비율이 유사하게 나타나고 있으므로 두 집단 간의 유의미한 차이가 나타나지 않은 것으로 보인다.

다음은 실험집단과 비교집단의 반성 글쓰기에서 생각의 변화 3수준에 대한 사례이다(사례3).

<사례3>

ST1: 가루를 내서 뿜으면 더 빨리 반응이 일어난다는 것을 책을 읽기 전부터 알고 있었다. 하지만 세밀한 이유 같은 것을 잘 몰랐다. 하지만 이번에 책을 읽으면서 반응물을 뿜아 결국엔 표면적이 넓어져서 반응속도가 빨라지는 것이었다. 한마디로, 반응물을 가루로 만드는 것은 표면적을 넓히기 위한 것이고 표면적이 반응속도와 관련이 있는 것이다. 표면적이 넓어져, 반응물질간의 접촉 면적이 커지면 반응물질 입자 간의 충돌횟수가 증가하고 유효충돌횟수가 증가해서 반응속도가 빨라지는 것이다. 그림으로 그리면 다음과 같다. (실험집단)



ST2: 실험하기 전에는 ‘표면적이 넓으면 반응속도가 빨라진다’에 대한 근거는 잘 모르고 있었지만 책을 보면서 반응물질이 고체인 경우, 고체 물질이 표면적이 넓으면 접촉면이 증가하여 반응 물질 입자 간의 충돌횟수 증가로 반응속도는 빨라진다는 것을 알 수 있었다. (비교집단)

사례3의 경우, 학생들은 알약보다 가루약 상태일 때가 반응속도가 더 빠르다는 것은 알고 있었으나 정확한 이유는 알지 못하고 있었다고 서술하였다. 읽기 자료를 통해 가루약 상태일 때가 표면적이 넓어지고, 이로 인해 충돌횟수가 증가하기 때문이라는 근거를 제시하며, 표면적이 클수록 반응속도가 빨라진다는 정교화된 나중의 생각을 서술하고 있다. 실험집단의 학생은 비교집단의 학생보다 자신의 생각을 자세하게 설명하고 이를 뒷받침할 수 있는 그림 자료를 함께 제시하는 경우가 더 많이 나타났다.

4) 의문 제기

의문 제기는 실험 결과나 학습 내용과 관련하여 새로운 의문을 제기하였는지를 확인하는 것이다.

분석 결과, 1차 반성 글쓰기에서 반성적 사고 0수준의 학생은 실험집단 99%, 비교집단 99%로 두 집단 모두 매우 높은 비율을 보였으며, 대부분의 학생들이 반성 글쓰기에서 학습한 내용에 대하여 의문을 제기하지 않았다는 것을 알 수 있다. 반성적 사고 1수준의 학생은 실험집단 1%로 나타났으며 비교집단에서는 나타나지 않았다. 메타인지 수준에 해당하는 반성적 사고 2수준 및 3수준의 학생은 비교집단 1%로 나타났으며, 실험집단은 나타나지 않았다. 반성적 사고 2수준(낮은 메타인지 수준)의 학생은 비교집단 1%로 나타났으며, 반성적 사고 3수준(높은 메타인지 수준)의 학생은 두 집단 모두 나타나지 않았다(Table 8).

3차 반성 글쓰기의 반성적 사고 수준별 학생 비율은 1차 반성 글쓰기의 수준별 학생 비율과 유사하게 나타났다. 반성적 사고 0수준의 학생은 실험집단 93%, 비교집단 97%로 실험집단이 비교집단에 비해 낮아졌으나, 두 집단 모두 여전히 높은 비율을 유지하고 있다. 반성적 사고 1수준의 학생은 실험집단 6%, 비교집단 1%로 실험집단이 비교집단보다 높은 비율을 나타냈다. 메타인지 수준에 해당하는 반성적 사고 2수준 및 3수준의 학생은 실험집단 1%, 비교집단 1%로, 두 집단 모두 반성적 사고 2수준(낮은 메타인지 수준)의 학생 비율이 1%로 동일하게 나타났다. 반성적 사고 3수준(높은 메타인지 수준)의 학생은 두 집단 모두 나타나지 않았다.

5차 반성 글쓰기에서 반성적 사고 0수준의 학생은 실험집단 75%, 비교집단 91%로 실험집단이 비교집단에 비해 낮은 비율로 나타났으며, 두 집단 모두 3차 반성 글쓰기에서의 반성적 사고 0수준 학생 비율보다 감소하였음을 알 수 있다. 반성적 사고 1수준의 학생은 실험집단 15%, 비교집단 1%로 실험집단이 비교집단보다 높은 비율로 나타났다. 메타인지 수준에 해당하는 반성적 사고 2수준 및 3수준 학생은 실험집단이 11%, 비교집단이 7%로 실험집단이 비교집단에 비해 높은 비율을 보였다. 이 중 반성적 사고 2수준(낮은 메타인지 수준)의 학생은 실험집단 8%, 비교집단 6%로 실험집단이 비교집단보다 높게 나타났으며, 반성적 사고 3수준(높은 메타인지 수준)의 학생은 실험집단 3%, 비교집단 1%로 실험집단이 비교집단보다 높은 비율로 나타났다.

활동이 진행됨에 따라 두 집단 모두 반성적 사고 0수준의 학생 비율은 감소하고, 반성적 사고 1, 2, 3수준의 학생 수가 증가하였다. 실험집단이 비교집단에 비해 반성적 사고 0수준의 학생 수가 감소하는 비율이 더 크고, 반성적 사고 1, 2, 3수준의 학생 수가 증가하는 비율이 더 컸지만, 실험집단도 5차 반성 글쓰기에서 반성적 사고 0수준의 학생 비율이 여전히 높게 나타났다. 따라서 1차, 3차, 5차 반성 글쓰기에서 의문 제기에 대한 두 집단 간의 유의미한 차이는 나타나지 않은 것으로 보인다.

Table 8. Analysis of asking questions in reflective writing

Reflective thinking level	Metacognitive level	Experimental groups(%)			Comparative groups(%)		
		1 st	3 rd	5 th	1 st	3 rd	5 th
0	.	99	93	75	99	97	91
1	.	1	6	15	0	1	1
2	low-metacognition	0	1	8	1	1	6
3	high-metacognition	0	0	3	0	0	1

Table 9. Analysis of reflective practice in reflective writing

Reflective practice level	Metacognitive level	Experimental groups(%)		Comparative groups(%)	
		3 rd	5 th	3 rd	5 th
0	.	8	6	22	36
1	.	44	31	49	49
2	low-metacognition	36	40	21	12
3	high-metacognition	12	22	8	3

5) 반성적 실천

반성적 실천은 학습자가 이전 활동에서 반성한 내용을 다음 활동에 실천으로 옮겨 행동의 변화가 나타났는지 알아보는 것이다.

3차 반성 글쓰기 분석 결과, 반성적 실천 0수준의 학생은 실험집단 8%, 비교집단에서는 22%로 실험집단이 비교집단보다 낮게 나타났다. 또한 반성적 실천 1수준의 학생은 실험집단 44%, 비교집단 49%로 실험집단이 비교집단보다 더 낮은 비율로 나타났지만 큰 차이는 보이지 않았다. 메타인지 수준에 해당하는 반성적 실천 2수준 및 3수준의 학생은 실험집단 48%, 비교집단 29%로 실험집단이 비교집단보다 더 높은 비율을 보였다. 이 중 반성적 실천 2수준(낮은 메타인지 수준)의 학생은 실험집단 36%, 비교집단 21%로 실험집단이 비교집단보다 높게 나타났으며, 반성적 실천 3수준(높은 메타인지 수준)의 학생은 실험집단 12%, 비교집단 8%로 실험집단이 비교집단보다 높은 비율을 보였다(Table 9).

5차 반성 글쓰기에서 반성적 실천 0수준의 학생은 실험집단 6%, 비교집단 36%로 실험집단에서는 거의 나타나지 않는 반면, 비교집단은 3차 반성 글쓰기에서보다 오히려 증가한 것을 보이고 있다. 반성적 실천 1수준의 학생은 실험집단 31%, 비교집단 49%로 실험집단보다 비교집단에서 더 높은 비율로 나타났다. 메타인지 수준에 해당하는 반성적 실천 2수준 및 3수준의 학생은 실험집단 62%, 비교집단 15%로 실험집단이 비교집단보다 높게 나타나 큰 차이를 보였다. 이 중, 반성적 실천 2수준(낮은 메타인지 수준)의 학생은 실험집단 40%, 비교집단 12%로 실험집단이 비교집단보다 더 높은 비율을 보였으며, 반성적 실천 3수준(높은 메타인지 수준)의 학생은 실험집단 22%, 비교집단 3%로 실험집단이 비교집단에 비해 더 높은 비율을 보였다.

다음은 실험집단과 비교집단의 반성 글쓰기에서 반성적 실천 3수준에 대한 사례이다(사례4).

<사례4>

ST1: 지난번 실험 때 실험과정을 어눌하게 세우고서 마음만 앞섰다가 막상 실험할 때 난감했던 경험으로 이번에는 제대로 실험계획을 세우고서 실험에 임했다. 실험주장과 근거를 작성하여 칠판에 제시할 때 글씨가 너무 작았던 탓에 멀리서 알아볼 수 없다는 것과 발표 자료에 오차를 추가적으로 서술할 수 있었을 텐데도 빼먹었다는 점이 마음에 걸린다. 지난 평가때 발표 자료에 오타를 내어 대충 마커로 덮었던 기억을 고려했어 신중하게 글씨와 내용에만 집중했는데 이때문인 것 같다. 그리고 그래프로 나타낼 수 있는 부분을 주제와 직접적이면

서 알기 쉽도록 반응속도를 구해서 그래프를 작성했던 점이 괜찮은 것 같다. 다른 모둠의 발표를 통해 너무 많은 정보를 넣지는 않되, 오차같은 좀 더 이해와 주의에 대한 내용을 추가하는 것이 우리 모둠에 좀 더 필요하다는 것을 알게 되었다. (실험집단)

ST2: 이때까지 실험을 할 때는 열심히 하고 적을 때는 너무 부실하게 적어서 발표도 못해보고 그랬었다. 그래서 이번에는 우리 모둠도 발표를 해보려고 하여 우리 모둠원 모두 실험을 어떻게 할지부터 정한 뒤 서로 역할분담을 하여 나는 발표정과 물을 넣고 생성물이 빠져나가는 것을 막는 역할을 하였고 OO이는 시간 체크와 기록, OO이는 메스실린더에 눈금이 20ml 씩 증가할 때마다 체크하였다.

(비교집단)

사례4에서, 실험집단은 이전활동에서 제시한 주장과 증거의 문제점과 어떻게 개선하였는지, 수업 태도도 어떻게 개선되었는지 구체적으로 서술하고 있다. 반면 비교집단은 참여도가 높아졌다와 같이 수업 전반에 대한 언급이 많았으며, 증거의 개수를 보강하고 다양한 방법으로 제시하거나 정확하게 제시하게 되었다는 모둠의 주장과 증거에 대한 반성적 실천은 잘 나타나지 않았다.

실험집단은 활동이 진행됨에 따라 반성적 실천 2, 3수준의 학생 수가 증가하였으나, 비교집단은 오히려 감소하였다. 이를 통해 실험집단 학생들이 비교집단 학생들보다 반성한 것을 실천으로 옮기는 반성적 실천능력이 더 높다는 것을 알 수 있다.

2. 학생 평가활동에 대한 인식 탐색

논의기반 탐구 과학수업에서 평가활동에 대한 심층적인 인식을 조사하여 학생 평가활동의 타당성을 확인하고 평가활동의 결과가 피드백으로써 다음 활동에 영향을 미쳤는지에 대한 인식 탐색을 위해 논의기반 탐구 과학수업 후 실험집단 학생들을 대상으로 설문조사와 인터뷰를 실시하였다. 분석 결과는 평가활동에 대한 인식, 동료평가 활동에 대한 인식, 자기평가활동에 대한 인식의 3가지 영역으로 나누어 제시하였다.

가. 평가활동에 대한 학생들의 인식

논의기반 탐구 과학수업에서 학생 평가활동을 실시하는 목적에 대해 학생들이 어떤 인식을 가지고 있는지를 알아보기로 인터뷰를 실시하였다. 인터뷰 분석 결과, 학생들은 평가활동의 목적으로 피드백을 통한 메타인지 향상, 반성과 반성적 실천의 향상을 제시하였다. 학생들은 자신이 알고 있는 것과 모르는 것이 무엇인지 확인하고, 잘 했던 점과 부족했던 점을 파악하는 메타인지 과정의 향상을 평가의 목적으로 인식하고 있었다. 또한 평가활동을 통해 수업에 대한 반성을 할 수 있었다고 응답하였다. 이를 통해 자신이 부족한 점을 파악하고, 다음 활동에서는 이를 개선하여 더 잘 수행하기 위한 반성적 실천의 측면을 평가활동의 목적이라고도 인식하고 있음을 알 수 있다.

다음은 평가활동의 목적에 대한 학생들의 인터뷰 응답 사례이다(사례5).

<사례5>

ST1: 우리 모둠이 다른 모둠보다 부족했던 걸 알 수 있고 다른 모둠보다 잘했던 것도 알 수 있으니깐 다음에 할 때는 더 잘 적을 수 있으니깐 (메타인지 향상)

ST2: 자기평가를 하면서 자기 반성을 많이 했다고 생각해요 실험을 하면서 실험을 하고 실험 결과가 나오고 그걸 보고 자신이 부족했던 점을 다시 한 번 생각할 수 있고 다음에 그걸 고쳐나갈 수 있고 그래서 자기평가를 하는 것 같고 동료평가는 자신이 깨닫지 못하는 부족한 점을 남이 평가해 주니깐 남이 보는 시선으로 개선할 수 있게 하는 것 같아요 (반성과 반성적 실천 향상)

평가활동의 활용 정도에 대한 설문조사 결과, 평가활동이 자신의 학습에 도움을 주는가? 라는 질문에 매우 그렇다 32명, 그렇다 39명, 보통이다 12명, 그렇지 않다는 1명 나타났으며, 전혀 그렇지 않다고 응답한 학생은 나타나지 않아 4.21점의 평균점수를 보였다. 이를 통해 학생들은 평가활동이 학습에 도움을 준다고 인식하고 있음을 알 수 있다.

학생들에게 평가활동이 도움이 된 구체적인 사례를 서술형으로 작성하도록 한 질문에 대해 학생들이 작성한 응답을 인지적, 행동적, 정의적 영역의 세 요소로 범주화 하였다. 인지적 영역에는 메타인지 향상, 논의 능력 향상, 학업 능력 향상, 글쓰기 능력 향상의 하위요소를 추출하였으며, 행동적 영역에는 반성적 실천 유도, 정의적 영역에는 흥미도 향상과 협동심 향상의 하위요소를 추출하였다(Figure 3).

평가활동이 인지적 영역 향상에 도움을 주었다고 서술한 응답은 전체 응답의 63.0%를 차지하였으며, 이 중 메타인지 향상에 도움을 주었다고 응답한 학생은 36명(30.3%)으로 가장 높은 빈도를 나타냈다. 다음으로 학업 능력이 향상 17명(14.3%), 논의 능력이 향상 16명(13.4%), 글쓰기 능력 향상(6명, 5.0%) 순으로 나타났다. 평가활동이 행동적 영역 향상에 도움을 주었다고 서술한 응답은 전체의 31.1%를 차지하였으며, 이들은 모두 반성적 실천을 유도하였다고 서술한 학생들(37명, 31.1%)이다. 평가활동이 정의적 영역 향상에 도움을 주었다고 서술한 응답은 전체응답의 5.9%를 차지하였으며, 이 중 과학에 대한 흥미를 향상시켰다고 응답한 학생은 4명(3.4%)이었으며, 협동심을 길러주었다고 응답한 학생은 3명(2.5%)이었다.

나. 동료평가활동에 대한 학생들의 인식

논의기반 탐구 과학수업에서 학생들이 수행한 동료평가는 타당하

였는지 알아보고, 동료평가 결과를 다음 활동에서 활용하였는지 알아보기 위해 동료평가 결과의 활용 정도와 사례를 설문조사와 인터뷰를 통해 분석하였다.

동료평가활동의 타당성을 조사하기 위해 자기 모둠의 주장과 증거를 다른 모둠의 학생들이 평가한 동료평가 결과는 타당하다고 생각 하는지 질문하였으며, 이에 대한 응답 결과는 매우 그렇다 19명, 그렇다 43명, 보통이다 17명, 그렇지 않다 4명, 전혀 그렇지 않다는 1명 응답하여 평균점수 3.89점을 보였다. 이로부터 학생들은 동료평가의 결과에 대해 대체적으로 타당하다고 인식하였다고 볼 수 있다. 학생들은 단순히 ‘성의가 없다’라거나 ‘이해하기 어렵다’ 등 미흡하다고 지적한 부분에 대한 평가 근거가 정확하게 서술되어 있지 않는 동료평가 결과가 수궁이 안 된다고 응답하였다. 또한 다른 친구들이 미흡하다고 평가했기 때문에 자신도 미흡하다고 평가했다는 경우처럼 자기 주관이 명확하게 드러나 있지 않는 근거 없는 평가 결과에 수궁이 되지 않는다고 응답하였다. 이를 통해 평가 결과에 대한 타당한 근거가 제시되지 않을 경우 학생들은 평가 결과를 수궁할 수 없다고 생각한다는 것을 알 수 있다. 따라서 동료평가를 통해 올바른 피드백을 제공하기 위해서는 반드시 타당한 평가 근거를 함께 제시해야 한다는 것을 알 수 있다.

학생들이 이전 활동에 대한 동료평가 결과를 다음 활동에서 자신의 활동에 반영하고 활용하였는지를 알아보기 위해 학생들에게 동료평가 결과의 활용 정도에 대해 질문하였다. 응답의 결과는 매우 그렇다 23명, 그렇다 45명, 보통이다 15명, 그렇지 않다는 1명 나타났으며, 전혀 그렇지 않다고 응답한 학생은 나타나지 않아 평균점수 4.07점을 보였다. 이로부터 학생들은 동료평가 결과를 다음 활동에 반영하고 있음을 알 수 있다.

동료평가 결과를 다음 활동에 반영하여 활용하였던 구체적인 사례를 제시하라는 질문에 대해 학생들이 서술한 응답 결과로 증거의 보강에 활용하였다고 서술한 학생은 45명(65.2%)으로 가장 높은 비율을 보였으며, 증거의 보강을 하는 방법 중 다중표상의 사용빈도가 증가하였다고 서술한 학생은 30명(43.5%), 증거의 개수가 증가하였다고 서술한 학생은 15명(21.7%)이었다. 증거제시 방법 개선에 활용되었다고 서술한 학생은 13명(18.8%), 증거의 정확성 점검 6명(8.7%), 수업 태도 개선 5명(7.2%), 주장과 증거의 적절성을 점검 4명(5.8%)의 순서로 나타났다(Figure 4).

학생들은 동료평가 결과를 다음 활동을 위한 피드백으로 지속적으로 제공받아 평가 결과를 활용하여 주장과 증거에 대한 설득력을 높

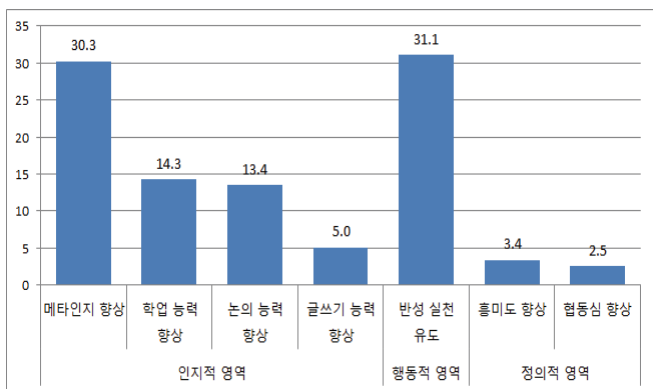


Figure 3. Perceptions of Students' assessment

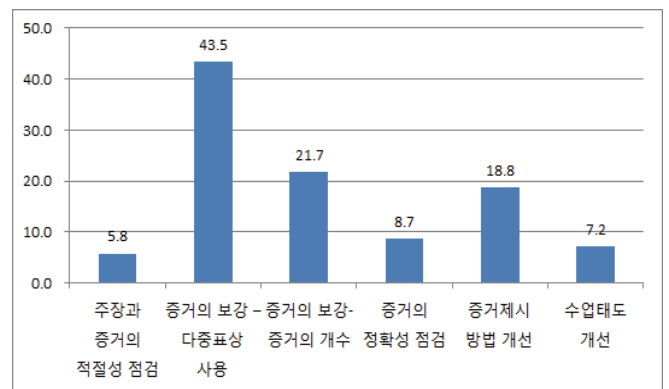


Figure 4. Outcome regarding application of peer assessment

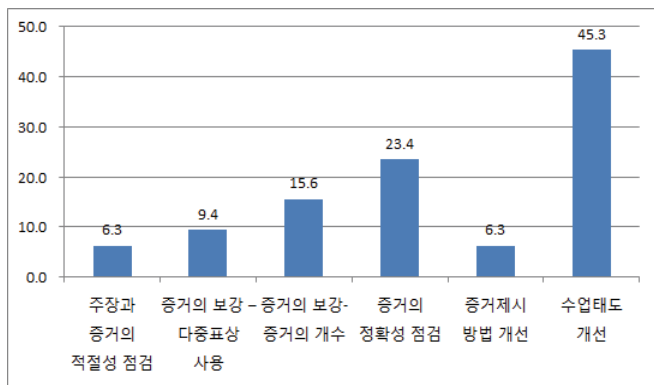


Figure 5. Outcome regarding application of self assessment

이기 위해 더 많은 증거를 제시하게 되었고, 그림이나 그래프와 같은 다중표상을 활용하게 되어 증거제시 방법이 다양해지게 되었음을 알 수 있다. 이에 따라 주장을 뒷받침하는 증거를 많이 활용하게 되면서 논의에서 증거와 보강 및 보강의 요소를 많이 활용하게 되었음을 알 수 있다. 따라서 학생들은 동료평가를 통해 지속적으로 피드백을 제공받음으로써 학습에서 자신의 수행에 대해 반성하고 이를 다음 활동에서 개선함으로써 반성적 실천이 유도되었음을 알 수 있다.

다. 자기평가활동에 대한 학생들의 인식

논의기반 탐구 과학수업에서 학생들이 자기 모듈의 주장과 증거를 평가하는 자기평가 결과가 타당하였는지 알아보고, 평가 결과를 어떻게 활용하였는지 알아보기 위해 동료평가와 동일한 방법으로 설문조사와 인터뷰를 실시하였다.

자기평가활동의 타당도에 대한 학생들의 생각을 조사하기 위해 자기 모듈에서 작성한 주장과 증거에 대해 평가한 결과는 타당하다고 생각하는지 질문하였다. 이에 대하여 매우 그렇다 21명, 그렇다 40명, 보통이다 22명, 그렇지 않다는 1명 나타났으며, 전혀 그렇지 않다고 응답한 학생은 나타나지 않아 평균점수 3.96점을 보였다. 이를 통해 학생들은 자기평가 결과는 대체적으로 타당하다고 인식하고 있음을 알 수 있다.

자기평가 결과가 학생들의 반성적 사고를 유도하여 반성적 실천을 이끌었는지에 대해 알아보기 위해 학생들에게 자기평가 결과의 활용 정도에 대해 질문하였다. 이에 대해 매우 그렇다 28명, 그렇다 40명, 보통이다 14명, 그렇지 않다는 2명 나타났으며, 전혀 그렇지 않다고 응답한 학생은 나타나지 않아 평균점수 4.12점을 보였다. 이러한 결과로부터 학생들은 자기평가 결과를 다음 활동에 반영하고 있음을 알 수 있다.

자기평가 결과를 다음 활동에 반영하여 활용하였던 구체적인 사례를 제시하라는 질문에 수업 태도 개선에 활용되었다고 응답한 학생은 29명(45.3%)으로 가장 높은 빈도를 보였다. 증거를 보강하는데 활용하였다고 서술한 학생은 16명(25.0%)이었으며, 이 중 다중표상의 사용 빈도가 증가하였다고 서술한 학생은 6명(9.4%), 증거의 개수가 증가하였다고 서술한 학생은 10명(15.6%)이었다. 증거의 정확성을 점검하는데 활용하였다고 서술한 학생은 15명(23.4%)이었으며, 주장과 증거의 적절성을 점검하는데 활용되었다고 서술한 학생은 4명(6.3%), 증거제시 방법을 점검하는데 활용되었다고 서술한 학생은

4명(6.3%) 나타났다(Figure 5).

자기평가 결과를 주장과 증거의 적절성과 정확성을 점검하고 증거를 보강하며 증거제시 방법을 개선하는 데 활용하였다는 응답은 동료평가 결과를 활용한 사례와 유사하였지만, 자기평가 결과를 수업 태도 개선에 활용하였다고 응답한 빈도가 가장 높았던 점에서 동료평가 결과를 활용한 사례와 차이가 있다.

학생들은 자기평가활동을 통해 자신에 대한 반성을 할 수 있었으며, 이를 통해 모둠원들 간의 협동 학습능력을 향상시킬 수 있었으며 수업참여도를 높이는 등의 수업 태도를 개선할 수 있었다고 응답하였다. 학생들은 자기평가 결과를 증거의 정확성을 점검하고 증거를 보강하며 주장과 증거의 적절성을 점검하는데 활용하였으며, 자신의 학습 과정을 점검하고 개선하는데 활용하였다. 특히 자기평가를 수업 태도 개선에 가장 많이 활용한 것으로 보아 평가활동이 학생들의 반성과 반성적 실천을 유도하였다고 볼 수 있다.

IV. 결론 및 제언

이 연구는 학생 평가활동을 적용한 논의기반 탐구 과학수업이 고등 학생들의 반성적 사고에 미치는 영향에 대한 탐색을 목적으로 한다. 이를 위해 학생 평가활동의 적용 유무에 따라 반성 글쓰기에서 나타나는 반성적 사고를 분석하였으며, 인터뷰와 설문조사를 통해 평가활동에 대한 학생들의 인식을 탐색하여 학생 평가활동의 타당성과 활용 정도를 알아보았다.

학생 평가활동을 적용한 논의기반 탐구 과학수업의 마지막 단계에서 작성한 학생들의 반성 글쓰기를 탐구과정 분석, 학습의 이해, 생각의 변화, 의문 제기 요소별로 수준을 점수화하여 총점을 분석한 결과, 1차 활동에서는 두 집단 간의 유의미한 차이가 없었으나, 3차와 5차 반성 글쓰기에서 실험집단이 비교집단에 비해 유의미하게 높은 점수를 나타내어 실험집단이 비교집단보다 반성적 사고 수준이 향상되었음을 보였다.

논의기반 탐구 과학수업이 진행될수록 탐구과정 분석과 학습의 이해, 생각의 변화 요소에서 실험집단과 비교집단 모두 메타인지 수준의 반성을 나타내는 학생 비율이 증가하였으나, 실험집단이 비교집단보다 메타인지 수준의 반성을 보이는 학생 비율이 더 높았다. 특히 5차 활동에서 탐구과정 분석과 학습의 이해 및 생각의 변화에서 높은 메타인지 수준에 해당하는 실험집단 학생의 비율이 비교집단에 비해 더 높았다. 이를 통해 실험집단이 비교집단보다 메타인지를 사용한 반성적 사고능력이 향상되었음을 알 수 있다.

학생 평가활동을 적용한 논의기반 탐구 수업에서 학생들은 동료평가활동을 통해 평가자로부터 피드백으로써 제공받은 결과를 증거를 보강하고 증거제시 방법을 개선하는데 활용하였으며, 증거의 적절성과 정확성을 점검하는 등 자신의 학습을 점검하고 개선하는데 활용하였다. 또한 학생들은 자기평가 결과를 이전 활동과 비교하여 자신의 수업 태도가 얼마나 개선되었는지 점검하는데 가장 많이 활용한 것으로 보아 학생 평가활동이 반성과 반성적 실천을 유도하였다고 볼 수 있다.

반성 글쓰기에서 학생들은 자신의 탐구과정의 오류와 문제점을 분석하고 원인을 찾아 이를 해결하는 방안을 제시하는 탐구과정 분석에서 메타인지를 사용하였다. 자신이 알고 있는 것과 모르는 것이

무엇인지, 자신의 생각이 어떻게 변했는지를 점검하고, 이에 따라 인지 과정을 조절 및 통제하는 과정에서도 메타인지 능력이 요구되며, 이러한 메타인지적 과정은 학습에서 중요한 부분을 차지한다고 볼 수 있다. 학생들은 반성을 통해 객관적으로 자신의 사고를 점검하며, 학습을 모니터링하며 활동에 대해 평가하였다. 이 때 학생들이 행하는 높은 수준의 반성에는 메타인지 기술이 요구되므로(Hatton & Smith, 1995), 학생들의 반성적 사고를 유도하고 명시적인 자신만의 과정을 형성하기 위해서는 메타인지 기술을 향상시켜야하며 이러한 메타인지 기술의 향상을 통해 반성적 사고능력을 촉진시킬 수 있다고 본다.

피드백을 제공받지 않은 상태에서 자신의 학습에 대해 반성할 때, 학습자들은 학습수행에 대하여 잘못된 기준으로 판단할 수 있으므로, 학습자에게 아무런 피드백을 제공하지 않는 경우보다 다른 사람으로부터 올바른 전략과 객관적인 관점과 관련한 피드백을 제공받았을 때 더 효과적으로 반성이 일어날 수 있다(Mayer, 2004). 따라서 학생들은 동료평가를 통해 지속적으로 피드백을 제공받음으로써 학습에서 자신의 수행에 대한 성공과 실패를 판단하는 효과적인 반성을 유도할 수 있었다고 본다. 또한 스스로에 대한 의문을 제기하여 특정 상황에 대한 자신의 행동과 성과를 검토하면서 자신이 무엇을 왜 하고 있는지에 대한 질문에 답하는 과정 자체가 반성이 될 수 있기 때문에 자기평가활동은 학생들의 메타인지와 반성을 촉진시키는 학습전략이 될 수 있다.

결론적으로, 학생 평가활동은 학생들의 메타인지를 촉진시켜 높은 수준의 반성을 이끌었으며, 이를 실천으로 옮기는 반성적 실천능력을 향상시켜 학습과 행동의 개선을 이끌었다고 생각된다. 학생들은 학생 평가활동의 목적을 자신이 미처 알지 못했던 부분에 대해 동료의 피드백을 통하여 알게 되고, 반성과 메타인지 과정을 통해 자신이 무엇을 알고 있는지 부족한 점은 무엇인지를 점검하며 반성을 실천으로 이끌어 부족한 점을 개선하기 위한 것이라고 인식하였다. 이를 통해 학생들이 학습의 결과로써 평가가 실시되는 것이 아니라, 학습 과정을 점검하고 향상시키기 위하여 피드백으로 제공되는 형성적 평가의 개념으로 인식하고 있음을 알 수 있다.

학생을 평가의 주체자로서 학습에 능동적으로 참여시키는 동료평가 및 자기평가활동은 구성주의적 학습관에서 볼 때 평가활동 자체가 학습의 일부분으로 통합되어 학생들의 사고력을 발달시키고 학습의 질을 향상시키며 과제에 대한 책임감을 증가시키는 등 학생들의 학습에 긍정적인 영향을 주었음을 알 수 있다. 또한 참여문화를 촉진시키는 학습 환경을 조성하여 성공적인 수행을 이끄는 방법이 될 수 있음을 보였다. 이를 통해 학생 평가활동은 학습의 과정으로 통합되어 학생들을 능동적으로 학습에 참여시킬 뿐만 아니라 학습에 대한 책임감을 향상시켜 학습을 스캐폴딩하는 도구로써 활용이 가능하다고 생각된다.

국문 요약

이 연구는 학생 평가활동을 적용한 논의기반 탐구 과학수업이 고등 학생들의 반성적 사고에 어떠한 영향을 미치는지 알아보고자 하였다. 이를 위해 광역시의 인문계 고등학교 1학년 6개 학급의 학생 166명을 대상으로 한 학기에 걸쳐 5개의 논의기반 탐구 과학수업을 실시하였

다. 실험집단(84명)은 논의기반 탐구 과학수업의 주장과 증거 단계에 동료평가활동과 자기평가활동을 적용하여 논의의 질을 평가하는 경험을 제공하였으며 비교집단(82명)은 이 단계를 생략하였다.

학생 평가활동이 학생들의 반성적 사고에 미치는 효과를 알아보기 위해 학생들이 작성한 반성 글쓰기를 분석하였으며, 인터뷰와 설문조사를 통해 평가활동에 대한 학생들의 인식을 탐색하여 학생 평가활동의 타당성과 활용 정도를 알아보았다. 반성 글쓰기 분석 결과, 1차 반성 글쓰기에서 두 집단 간의 유의미한 차이가 없었으나 3차와 5차 반성 글쓰기에서 실험집단의 점수가 비교집단보다 유의미하게 높았다. 반성 글쓰기의 하위 요소에서 탐구과정 분석, 학습의 이해, 생각의 변화 요소는 실험집단과 비교집단 모두 메타인지 수준의 반성을 나타내는 학생 비율이 증가하였으나, 실험집단이 비교집단보다 메타인지 수준의 반성을 보이는 학생 비율이 더 높았다. 반성적 실천 요소의 분석 결과, 실험집단은 메타인지 수준의 반성적 실천 학생 비율이 증가하였으나 비교집단은 오히려 감소하였다. 학생 설문조사와 인터뷰를 분석한 결과, 학생들은 평가활동이 자신의 학습에 도움이 된다고 인식하고 있었다. 또한 평가활동을 통해 자신이 무엇을 얼마나 알고 있는지를 깨닫게 되고 문제점을 파악하여 개선할 수 있는 방법을 탐색하는 메타인지 능력이 향상되었다고 응답하였다. 이러한 메타인지적 반성을 통해 부족한 부분을 개선하려 노력하였으므로 평가활동이 반성적 실천을 이끌었다고 인식하였다.

따라서 학생 평가활동은 학습의 과정으로 통합되어 학생들을 능동적으로 학습에 참여시킬 뿐만 아니라 학습에 대한 책임감을 향상시켜 학습을 스캐폴딩하는 도구로써 활용이 가능하다고 생각된다.

주제어 : 동료평가활동, 자기평가활동, 논의기반 탐구 과학수업, 반성적 사고, 메타인지,

References

- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-Assessment. *Assessment and Evaluation in Higher Education*, 32(2), 159-181.
- Bangert-Drowns, R., Hurlley, M. M., & Wilkinson, B. (2004). The effect of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, 74, 29-58.
- Biggs, J. B., & Moore, P. J., (1993). *The Process of Learning* (3rd ed.). New York: Prentice Hall.
- Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education*. London: Kogan Page.
- Brindley, C. & Scoffield, S. (1998). Peer assessment in undergraduate programmes. *Teaching in Higher Education*, 3(1), 79-80.
- Cheng, W., & Warren, M. (1999). Peer and teacher assessment of the oral and written tasks of a group project. *Assessment & Evaluation in Higher Education*, 24(3), 301-314.
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328-338.
- Dewey, J. (1933). *How we think: A Restatement of the relation of reflective thinking to the Educative Process*. Boston; D.C. Heath.
- Duschl, R., Schweingruber, H., & Shouse, A. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.
- Erduran, S., & Jimenez-Alexandre, M. (2007). Argumentation in science education: an overview. *Argumentation in Science Education*, 35, 3-27.
- Fairbrother, B., Black, P., & Gill, P. (Eds.). (1995). *Teachers Assessing Pupils*. London: The Association of Science Education.

- Fiske, S., & Taylor, S. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Gama, C. (2004). Metacognition in interactive learning environments: The reflection assistant model. In *Intelligent Tutoring Systems* (pp. 668-677). Springer Berlin Heidelberg.
- Gielen, M., & De Wever, B. (2013). Structuring the peer assessment process: Impact on feedback quality. In N. Rummel, M. Kapur, N. Mitchell, & S. Puntambekar (Eds.), *To see the world and a grain of sand: Learning across levels of space, time, and scale: CSCL 2013 conference proceedings*, 2, 255-256.
- Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation, *Teaching and Teacher Education*, 11(1), 33-49.
- Heo, E. *The Impact of General Chemistry Laboratory Using Reading Frame-Based Science Writing Heuristic Approach on College Students' Reflective Thinking*. Ph.D. Dissertation, Pusan National University, Busan, Korea, 2011.
- Hwang, S., Kim, H., Yoo, J., & Pak, S. (2001). Ninth Graders' Self-Assessment of Scientific Process Skills in Open Investigation. *Journal of the Korean Association for Science Education*, 21(3), 506-515.
- Jang, K., Nam, J., & Choi, A. (2012). The Effects of Argument-Based Inquiry Using the Science Writing Heuristic (SWH) Approach on Argument Structure in Students' Writing. *Journal of the Korean Association for Science Education*, 32(7), 1099-1108.
- Keys, C., Hand, B., Prian, V., & Collins, S. (1999). Using the Science Writing Heuristic as a Tool for Learning from Laboratory Investigations in Secondary Science. *Journal of Research in Science Teaching*, 36(10), 1065-1084.
- Kollar, I. & Fischer F. (2012). Peer assessment as collaborative learning: a cognitive perspective. *Learning and Instruction*, 20(4), 344-348.
- Larson, A., Britt, M., & Kurby, C. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education*, 77(4), 339-366.
- Lee, S., Bak, D., & Nam, J. (2015). Impact of Peer Assessment Activities on High School Student's Argumentation in Argument-Based Inquiry. *Journal of the Korean Association for Science Education*, 35(3), 353-361.
- Li, L., Liu, X., & Steckelberg, A. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41(3), 525-536.
- Li, L., Steckelberg, A. L., & Srinivasan, S. (2008). Utilizing peer interactions to promote learning through a computer-assisted peer assessment system. *Canadian Journal of Learning and Technology*, 34(2), 133-148.
- Mayer, R. E. (2004). Should there be a three-strike rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59, 14-19.
- Ministry Of Education. (2015). 2015 Revised national curriculum. MOE, Notice No. 2015-74.
- Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, 97, 117-128.
- Nam, J., Koh, M., Bak, D., Lim, J., Lee, D. & Choi, A. (2011). The Effects of Argumentation-based General Chemistry Laboratory on Preservice Science Teachers' Understanding of Chemistry Concepts and Writing. *Journal of the Korean Association for Science Education*, 31(8), 1077-1091.
- Nam, J., Kwak, K., Jang, K., & Hand, B. (2008). The implementation of argumentation using Science Writing Heuristic (SWH) in Middle School Science. *Journal of the Korean Association for Science Education*, 28(8), 922-936.
- Narciss, S. (2004). The impact of informative tutoring feedback and self-efficacy on motivation and achievement in concept learning. *Experimental Psychology*, 51(3), 214-228.
- O'Malley, J. M. & Chamot, A. U. (1990). *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.
- Osborne, J., Erduran, S., & Simmon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*. 41(10), 994-1020.
- Phillips, L., & Norris, S. (1999). Interpreting popular reports of science: What happens when the readers' world meets the world on paper?. *International Journal of Science Education*, 21(3), 317-327.
- Pierce, A. E. (2003). Irruptions of voice: A critique of "Rhizovocality: by a white feminist researcher. *International Journal of Qualitative Studies in Education*, 16(5), 721-723.
- Pope, N. (2001). An examination of the use of peer rating for formative assessment in the context of the theory of consumption values. *Assessment & Evaluation in Higher Education*, 26(3), 235-246.
- Prins, F., Sluijsmans, D., & Kirschner, P. (2006). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education*, 11, 289-303.
- Purchase, H. (2000). Learning about interface design through peer assessment. *Assessment and Evaluation in Higher Education*, 27(4), 341-352.
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education*, 28(2), 147-164.
- Sampson, V., Enderle, P., Grooms, J., & Witte, S. (2013). Writing to Learn by Learning to Write During the School Science Laboratory: Helping Middle and High School Students Develop Argumentative Writing Skills as They Learn Core Ideas. *Science Education*, 97(5), 643-670.
- Smyth, K. (2004). The benefits of students learning about critical evaluation rather than being summatively judged. *Assessment and Evaluation in Higher Education*, 29(3), 369-378.
- Spiller, D. (2009). *Assessment matters: Self assessment and peer assessment*. New Zealand: University of Waikato.
- Strange, J. M., & Mumford, M. D. (2005). The origins of vision: Effects of reflection, models, and analysis. *The Leadership Quarterly*, 16(1), 121-148.
- Strijbos, J. W., Van Goozen, B., & Prins, F. (2012). Developing a coding scheme for analysing peer feedback messages.
- Sung, H., Hwang, S., & Nam, J. (2012). Examining the Relation Between Students' Reflective Thinking and the Reading Framework in the Science Writing Heuristic (SWH) Approach. *Journal of the Korean Association for Science Education*, 32(1), 146-159.
- Sung, H. & Nam, J. (2013). The Impact of Reading Framework as a Reading Strategy on Writing for Reflection of Middle School Students. *Journal of the Korean Association for Science Education*, 33(2), 249-265.
- Tavares, M., Jiménez-Aleixandre, M., & Mortimer, E. (2010). Articulation of conceptual knowledge and argumentation practices by high school students in evolution problems. *Science & Education*, 19(6-8), 573-598.
- Yore, L., Bisanz, G., & Hand, B. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, 25(6), 689-725.
- Zoller, U. (1999). Teaching tomorrow's college science courses are we getting it right?. *Journal of College Science Teaching*, 29(6), 409-414.
- Zoller, U., Fastow, M. & Lubezky, A. (1997). Student self- assessment in chemistry examinations requiring higher and lower order cognitive skills. *Journal of Chemical Education*, 76(1), 112-113.