

## 고등학교 논의기반 탐구 과학수업에서 학생 평가활동이 주장과 증거 형성에 미치는 영향

이선우·남정희\*

부산대학교 화학교육과

(접수 2017. 12. 21; 게재확정 2018. 3. 13)

### Impact of Student Assessment Activities on Claim and Evidence Formation in High School Argument-Based Inquiry

Seonwoo Lee and Jeonghee Nam\*

Department of Chemistry Education, Pusan National University, Busan 46241, Korea. \*E-mail: jhnam@pusan.ac.kr

(Received December 21, 2017; Accepted March 13, 2018)

**요약.** 이 연구는 논의기반 탐구 과학수업에서 학생 평가활동이 고등학생들의 주장과 증거 형성에 어떠한 영향을 미치는지 알아보려고 하였다. 이를 위해 인문계 고등학교 1학년 166명을 대상으로 논의기반 탐구 과학수업을 실시하였으며, 실험집단 학생(84명)들은 모두의 주장과 증거에 대한 평가 기회를 제공받았고 비교집단 학생(82명)들은 이 단계를 생략하였다. 학생 평가활동이 학생들의 주장과 증거 형성에 미치는 영향 알아보기 위해 학생들이 작성한 주장과 증거 그리고 학생 평가활동의 결과를 분석하였으며, 인터뷰와 설문조사를 통해 학생 평가활동의 타당성과 활용 정도를 알아보았다. 연구 결과, 논의활동에서 증거를 형성하는 능력이 비교집단보다 실험집단이 통계적으로 유의미하게 높게 나타났으며, 실험집단 학생들이 비교집단 학생들보다 더 정확하고 충분한 증거를 제시하는 것으로 나타났다. 또한 실험집단 학생들은 평가활동을 통해 주장에 대한 증거가 적절한지, 증거가 충분한지, 제시된 증거 자료는 정확한지 대해 평가하며 지속적으로 평가결과를 피드백으로 활용하였음을 알 수 있었다. 학생들은 평가 결과가 대체적으로 타당하다고 인식하고 있으며, 평가 결과를 다음 활동에 반영하여 다양한 증거를 제시하는 능력이 향상되었다고 인식하고 있음을 알 수 있었다. 이를 통해 논의기반 탐구 과학수업에서 학생 평가활동이 학생들의 주장과 증거를 형성할 때 주장의 타당성과 정당성을 높이는 증거 형성에 효과적이라는 것을 알 수 있다.

**주제어:** 논의기반 탐구 과학수업, 논의, 동료평가, 자기평가

**ABSTRACT.** The purpose of this study was to investigate the impact of student assessment activities on claim and evidence formation in argument-based inquiry (ABI) for high school students. The participants of the study were 166 grade 10 students from six different classes in the same high school. 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. Over one semester students participated in five ABI programs that we developed. According to the analysis of the claim and evidence from groups, the experimental group had a significantly higher mean score than the comparative group. The result of analysis of students' assessment in the experimental group, the frequency about accurate and sufficient evidence revealed to be high and students assessed whether peers' claims fit with the evidence and whether peers' explanations of the evidence's validity was sufficient. Students' answers in the survey and interviews showed that the students thought they could improve the accuracy of their ideas, appropriateness of their evidence, and the method of presenting evidence based on the assessment results.

**Key words:** Argument-based inquiry, Argumentation, Peer assessment, Self assessment

### 서론

2015 개정 과학 교육과정에서는 기본 개념에 대한 통합적인 이해를 바탕으로 과학적 사고력, 탐구 능력, 문제 해결력, 의사소통 능력 등의 과학과 핵심역량을 함양할 수 있는 교육을 추구한다. 이를 위해 교육과정에 과학교과 역량을 도입하고 내용 체계를 재구조화하여 학생들의 학

습량을 감축하였으며, 교과내용의 기준 시수를 감소하고 교사의 여유 시간을 확보하여 학생들이 깊이 있게 탐구할 수 있도록 제도를 개선하였다.<sup>1</sup> 이처럼 학생 참여 중심의 교육환경을 구축하기 위해서는 교사가 과학개념을 일방적으로 전달하는 수업보다는 동료들과 논의할 수 있는 기회를 제공하여 학생들이 스스로 지식을 구성하고 능동적으로 사고하는 학습 환경을 제공해야 한다.<sup>2-4</sup>

과학 교육 현장에서는 학생들에게 학생 참여 중심인 협력적 활동을 제공하여 탐구 능력과 논의 능력, 의사소통 능력을 향상시킬 수 있는 기회를 제공해야 하는데, 이러한 교수방법 중 하나가 논의기반 탐구 과학수업(Argument-Based Inquiry, ABI)이다.<sup>4-6</sup> 탐구 과정 속에 논의와 글쓰기를 접목시켜 메타인지를 촉진시키고 과학적 사고력을 향상시키는 논의기반 탐구 과학수업은 과학적인 증거를 수집하고, 활동의 전체 과정에서 학생들 간에 협상과 합의 과정을 중요시하므로 의사소통 능력을 향상시킬 수 있는 프로그램이다.<sup>6-10</sup> 논의기반 탐구 과학수업은 탐구과정에 논의와 읽기, 글쓰기를 통합함으로써 논의 능력과 글쓰기 능력, 탐구 능력을 향상시킬 수 있으며 과학 개념에 대한 이해력을 높이고 메타인지를 향상시키는 데 도움을 줄 수 있다.<sup>8-11</sup> 또한 학생들 사이의 의사소통 과정을 통해 합리적인 의사결정 능력도 향상시킬 수 있다.<sup>12,13</sup>

학생들은 학습에서 자신이 목표를 설정하고 학습을 점검하며 평가를 수행하는 모든 과정에 능동적으로 참여해야 한다.<sup>14</sup> 학생 참여 중심 학습으로의 교수·학습방법 개선이 지속적으로 강조됨에 따라 학생들의 수행에 대한 평가도 결과 중심의 평가가 아니라 과정 중심의 평가로 변화를 요구하고 있다.<sup>1</sup> 2015 개정 교육과정에서도 학습 과정을 중시하는 평가를 강화하여 학습결과 뿐만 아니라 학습 과정에서 학생들에게 양적, 질적 피드백을 제공하여 성취기준에 도달하도록 도와주고 학생들 스스로 자신의 학습을 성찰하고 개선할 수 있도록 안내하는 평가가 이루어지기를 요구하고 있다. 구성주의 학습관의 영향으로 학생중심의 학습 및 자기주도적인 학습에 대해 강조하면서, 평가에서 수동적인 역할을 해온 학습자를 평가의 주체로서 능동적인 학습자의 역할을 강조하게 되었다.<sup>15</sup> 이러한 학습과 평가에서 학습자의 역할에 대한 변화는 과학 교육에서의 학습과 평가에도 영향을 미친다. 학습자들을 평가 활동에 참여하게 하여 평가를 학습 과정에 포함시키면 수업에서 학생들의 참여도를 더욱 높일 수 있다.<sup>15-17</sup>

학생들을 평가활동에 참여시켜 평가를 학습의 과정으로 통합시키는 방법으로 동료가 수행한 과제에 등급을 정하거나 피드백을 제공하는 것을 의미하는 동료평가활동<sup>18</sup>과 학습자 스스로 자신의 학습 성과에 대해 인식하고 반성하여 수정하는 자기평가활동이 있다.<sup>19</sup> 동료평가를 통해 동료의 과제를 학습자가 평가해 보고 피드백을 제공함으로써 자신의 과제 결과물에 대해서 검토해 볼 뿐만 아니라 동료 평가자가 제공하는 정보를 토대로 자신의 주장을 더 견고하게 할 수 있다.<sup>18,20</sup> 동료평가활동에서 평가자는 비판적인 사고능력을 향상시킬 수 있고 피평가자는 적절한 피드백을 제공 받을 수 있으므로 평가자와 피평가자 모두의 학습을 향상시킬 수 있는 방법이다.<sup>21-25</sup> 자기평가

는 학생들 자신의 수준과 특성을 인정하는 평가로, 교사가 제시한 내용에 따라 자신이 수행한 학습을 자신이 평가하는 과정에서 반성적 사고를 통해 자신의 장점과 단점을 파악할 수 있는 기회를 제공해주는 평가이다.<sup>19,20</sup> 학습을 시작할 때부터 학생들에게 자기평가의 중요성을 인지시키고 자신들의 학습과정과 학습결과에 있어서 객관적인 평가를 할 수 있는 기회를 제공하여 준다면, 자기평가는 효율적인 학습 방법이 될 수 있을 것이다.<sup>18,26</sup>

논의활동에서 특히 중요한 것은 과학적 주장과 그에 대한 증거를 평가하는 것이다.<sup>27,28</sup> 과학 연구자들이 과학적 지식에 대한 주장과 증거를 구성하는 과정에서 동료와 지속적인 협의와 논의를 거쳐 지식을 수정하고 정교화하는 것처럼,<sup>29</sup> 학습자들의 과학 지식 또한 학습자 간의 논의과정을 통한 사회적 상호작용으로 구성된다고 볼 수 있다. 모둠과 학급의 논의를 통한 논의과정은 본질적으로 동료들의 진술을 고려하여 개인과 동료가 만들어 가는 대화의 과정이다.<sup>30,31</sup> 학습자들은 과학 질문에 대한 답을 이끌어내는 과정을 통해 모둠에 참여할 수 있고 과학 지식의 구성 과정을 경험할 수 있다. 이 과정에서 학생들은 자신의 활동에 대해 반성할 기회를 가질 수 있고, 이러한 과정은 학습자에게 메타인지를 촉진시켜 과학 지식이 내면화되도록 도와준다.<sup>13,24-25</sup>

따라서 이 연구에서는 학생들의 능동적인 참여를 유도하고 높은 수준의 주장과 증거를 형성하는 논의 능력의 향상을 위해 논의기반 탐구 과학수업에서 동료평가와 자기평가를 이용한 학생 평가활동을 논의 활동에 적용하여 학생 평가활동이 고등학생의 주장과 증거 형성에 어떻게 영향을 미치는지 알아보려고 하였다.

## 연구 방법

### 연구 대상 및 참여자

이 연구는 인문계 고등학교 1학년 학생 중 과학중점반 6개 학급의 학생 166명을 연구 대상으로 선정하였으며, 6개 학급 중 무작위로 3개 학급 84명은 실험집단으로, 3개 학급 82명은 비교집단으로 구성하였다. 수업은 학급당 8개의 모둠으로 구성되어 모둠활동으로 이루어졌으며 모둠 구성원은 1학기 과학성적을 고려하여 한 모둠 당 3-4명의 이질집단으로 구성하였다. 두 집단 모두 2014년 7월부터 한 학기동안 5개 주제의 논의기반 탐구 과학수업을 교육 과정에 따라 순차적으로 진행하였다. 실험집단은 동료평가활동과 자기평가활동으로 구성된 학생 평가활동을 논의기반 탐구 과학수업에 적용하여 진행하였으며, 비교집단은 학생평가활동을 생략하고 논의기반 탐구 과학수업을 진행하였다.

### 학생 평가활동을 적용한 논의기반 탐구 과학수업(ABI) 프로그램 및 적용

논의기반 탐구 과학수업(Argument-Based Inquiry, ABI)은 논의와 글쓰기를 활용하여 학생들의 탐구능력과 과학개념의 이해를 향상시킬 뿐만 아니라 메타인지를 촉진시키기 위해 개발된 프로그램으로, Keys 등<sup>6</sup>이 개발한 탐구적 과학 글쓰기(Science Writing Heuristic, SWH) 수업전략을 우리나라의 교육실정에 맞게 재구성한 프로그램이다. 논의기반 탐구 과학수업은 의문만들기, 실험설계 및 수행, 관찰, 주장과 증거, 읽기, 반성하기의 6단계로 구성되어 있으며,<sup>6,7</sup> 이 연구에서는 주장과 증거 단계에 동료평가활동과 자기평가활동을 포함하는 학생 평가활동을 실시하였다. 1개의 주제에 대하여 주장과 증거 단계까지 수행하는데 평균 2차시(50분)가 소요되었으며 읽기와 반성 단계는 과제로 수행하였다. 따라서 5개의 주제의 프로그램을 수행하는데 총 10차시가 소요되었으며 수업은 일주일에 1차시씩 진행하였다.

프로그램은 실험 수행 가능 여부에 따라 논의기반 탐구 과학수업의 모든 단계를 수행하는 A유형과, 제시된 재료로 분자의 구조를 제작하는 과정을 실험 수행단계로 대체한 B유형으로 나누어 개발하였다. 프로그램에 적용할 활동 주제는 고등학교 1학년 학생들이 과학 교과에서 학습하는 내용을 바탕으로 분자의 구조와 극성과 관련한 2개의 주제와 화학반응속도와 관련한 3개의 주제를 선정하였다(Table 1). 개발된 프로그램은 과학 교육 전문가 1명, 과학 교육 박사과정 3명, 과학 교육 석사과정 3명에게 내용타당도를 검증받았다.

학생 평가활동을 적용한 논의기반 탐구 과학수업은 기존의 논의기반 탐구 과학수업에서 주장과 증거 단계에 동료평가와 자기평가를 포함하는 학생 평가활동을 추가하였다. 실험집단 학생들은 주장과 증거 단계에서 다른 모둠의 주장과 증거에 대해 평가하는 동료평가활동과 자기 모둠의 주장과 증거를 평가하는 자기평가활동을 수행하였다.

학생 평가활동은 ‘등급 정하기’, ‘평가 결과 작성하기’, ‘피드백 제공하기’의 세 가지 과정으로 구성하였다. 주장과 증거 단계에서 모둠의 주장과 증거를 A3 용지에 작성하여 칠판에 게시하면 학생들은 다른 모둠의 주장과 증거를 읽고 이 중에서 가장 우수한 모둠과 가장 부족하다고 생각되는 모둠을 한 모둠씩 선정하는데 이 과정이 첫 번째 단계인 ‘등급 정하기’ 과정이다. ‘평가 근거 작성하기’ 과정에서는 학생들이 활동보고서에 자신이 가장 우수한 모둠과 가장 부족한 모둠을 선정한 근거를 작성하는 과정으로, 주장 및 증거자료의 적절성과 타당성에 대한 평가 결과를 작성하였다. 또한 가장 우수하다고 선택한 모둠과 가장 부족하다고 선택한 모둠과 비교했을 때 자신의 모둠은 이 두 모둠과 어떻게 비교되는지 자기 모둠의 주장과 증거자료를 평가하였다. ‘피드백 제공하기’ 과정은 평가 자료를 공유하기 위한 과정이다. 우선 학급 논의과정에서 실험결과를 바탕으로 작성한 모둠의 주장과 증거에 대해서 한 팀이 발표를 하면 실험결과에서 오차나 의문사항에 대해 질의·응답하는 과정을 통해 즉각적인 피드백을 제공한다. 또한 학생들이 활동보고서에 작성한 평가 결과를 교사가 모두 수집하고 정리하여 다음 활동을 시작하기 전 학생들에게 배부하는 지연 피드백을 제공하였으며, 평가 결과를 반영하여 이전 활동에서 지적 받은 문제점들을 개선할 수 있도록 지도하였다. 실험집단에서는 동료평가 및 자기평가를 포함하는 평가활동을 실시하였으며, 비교집단에서는 평가 단계만 생략하고 나머지 모든 단계는 실험집단과 동일하게 실시하였다.

**Table 1.** Topics of Argument-based inquiry activities

Topic	Type	Period
Polarity of molecular	B	2hr
Structure of molecular	B	2hr
Effect of temperature on reaction rates	A	2hr
Concentration and reaction rates	A	2hr
Surface area and reaction rates	A	2hr

가장 우수한 모둠과 가장 부족하다고 생각되는 모둠을 한 모둠씩 선정하는데 이 과정이 첫 번째 단계인 ‘등급 정하기’ 과정이다. ‘평가 근거 작성하기’ 과정에서는 학생들이 활동보고서에 자신이 가장 우수한 모둠과 가장 부족한 모둠을 선정한 근거를 작성하는 과정으로, 주장 및 증거자료의 적절성과 타당성에 대한 평가 결과를 작성하였다. 또한 가장 우수하다고 선택한 모둠과 가장 부족하다고 선택한 모둠과 비교했을 때 자신의 모둠은 이 두 모둠과 어떻게 비교되는지 자기 모둠의 주장과 증거자료를 평가하였다. ‘피드백 제공하기’ 과정은 평가 자료를 공유하기 위한 과정이다. 우선 학급 논의과정에서 실험결과를 바탕으로 작성한 모둠의 주장과 증거에 대해서 한 팀이 발표를 하면 실험결과에서 오차나 의문사항에 대해 질의·응답하는 과정을 통해 즉각적인 피드백을 제공한다. 또한 학생들이 활동보고서에 작성한 평가 결과를 교사가 모두 수집하고 정리하여 다음 활동을 시작하기 전 학생들에게 배부하는 지연 피드백을 제공하였으며, 평가 결과를 반영하여 이전 활동에서 지적 받은 문제점들을 개선할 수 있도록 지도하였다. 실험집단에서는 동료평가 및 자기평가를 포함하는 평가활동을 실시하였으며, 비교집단에서는 평가 단계만 생략하고 나머지 모든 단계는 실험집단과 동일하게 실시하였다.

### 자료 수집

학생 평가활동을 적용한 논의기반 탐구 과학수업에서 평가활동이 학생들의 주장과 증거 형성에 미치는 영향을 알아보기 위해 학생들이 작성한 모둠의 주장과 증거와, 학생 평가활동의 결과 자료를 수집하였다. 또한 평가활동에 대한 학생들의 인식을 알아보고자 프로그램이 끝난 후 실험집단 학생들을 대상으로 설문조사와 인터뷰를 실시하였다.

### 분석 방법

논의기반 탐구 과학수업에서 학생 평가활동이 학생들의 주장과 증거 형성에 미치는 영향을 알아보기 위해 활동보고서에 작성한 모둠의 주장과 증거 및 학생 평가활동의 결과 자료를 분석하였다. 평가들은 Sampson 등<sup>32</sup>의 ADI (Argument-Driven inquiry) 수업모형 분석틀과 장경화<sup>13</sup>의 탐구적 과학 글쓰기 활동 평가틀을 수정·보완하여 논의기반 탐구 과학수업(ABI)에서 모둠의 주장과 증거에 대한 평가틀과 학생들이 작성한 평가 결과에 대한 평가틀을 개발하였다. 두 개의 평가틀은 과학교육 전문가 1명, 과학교육 박사과정 3명, 과학교육 석사과정 1명이 지속적인 논의를 통해 개발하였다.

논의기반 탐구 과학수업에서 학생들이 작성한 활동 보

고의 주장과 증거를 분석하기 위한 평가들은 주장과 증거의 두 개 영역으로 구분하여 사용하였다. 주장에 대한 세부요소로는 적절성과 정확성을 추출하였으며, 증거에 대한 세부요소로는 적절성, 충분성, 정확성으로 추출하였다. 추출된 각 요소들은 학생들에게 기대되는 수행수준에 따라 0점, 1점, 2점으로 점수화하였다.

주장은 실험을 통해 수집된 자료를 분석하여 학급 의문에 대한 주장을 만드는 것으로, 세부요소 중 적절성은 주제에 대한 주장을 펼칠 때 변인과의 관계를 제대로 진술하였는지를 평가하는 요소이다. 변인과의 관계를 정확하게 제시한 경우 2점, 각 변인에 대한 진술만 한 경우 1점, 변인에 관한 언급이 없는 경우 0점을 부여하였다. 정확성은 주제에 대한 과학적 개념을 모두 사용하여 개념과의 관계가 명확하게 진술되는지를 평가하는 요소이다. 관련된 개념간의 관계를 정확히 설명하는 경우 2점, 관련된 개념간의 관계를 정확히 설명하지 않는 경우 1점, 관련된 개념간의 관계가 없는 경우 0점을 부여하였다.

증거는 주장을 진술할 때 이를 뒷받침 할 수 있는 증거를 제시하는 것으로, 증거의 세부요소는 적절성, 충분성, 정확성의 세 가지로 구성되어 있다. 적절성은 수집된 증거를 바탕으로 주장을 뒷받침할 수 있는지를 평가하는 요소이다. 주장을 뒷받침할 수 있는 증거를 진술하는 경우 2점, 주장과 관련이 없는 증거가 일부 포함되게 진술하는 경우 1점, 주장과 관련이 없는 데이터만 나열하여 진술하는 경우 0점을 부여하였다. 충분성은 다양한 자료로 증거를 제시하는지 평가하는 요소이다. 증거의 개수가 많을수록 주장에 대한 신뢰도는 높아지며, 수집된 자료를 바탕으로 정리된 실험결과를 단순히 글로만 나타내는 것이 아니라 그래프, 표, 그림 등으로 변환하여 이해하기 쉽게 설명할 수 있어야 한다. 따라서 실험결과를 그래프, 표, 그림 등을 사용하여 3개 이상의 증거를 제시하는 경우 2점, 2개 제시하는 경우 1점, 1개 제시하는 경우 0점을 부여하였다. 정확성은 주장에 대한 증거 자료가 정확한지 평가하는 요소이다. 내용을 정확히 진술하였을 뿐만 아니라 오차 분석도 제시하는 경우 2점, 내용을 일부 정확하지 않게 진술하거나 오차를 제시하지 않는 경우 1점, 주장을 뒷받침하지 않거나 내용을 정확하게 진술하지 않은 경우 0점을 부여하였다.

평가 결과는 모둠 별로 작성한 주장과 증거 자료를 칠판에 게시하면 이를 바탕으로 동료평가활동가 자기평가 활동을 실시하는데 이 때 가장 우수한 모둠과 가장 부족한 모둠이라고 선정한 평가 근거와, 선정한 두 모둠과 비교하여 자신이 속한 모둠의 주장과 증거에 대해 어떻게 평가하는지를 작성한 결과를 말한다. 설문조사에서 다른 모둠의 주장 및 증거에 대한 평가 기준을 제시하라는 설

문에 학생들이 서술한 증거의 적절성, 증거의 충분성, 증거의 정확성, 증거제시 방법의 4가지 범주로 분류할 수 있었고, 이 기준에 따라 학생 평가활동 결과를 분석하였다. 평가 결과에 대한 분석은 평가들에 제시한 평가 기준이 나타나는 빈도수를 측정하였다.

논의기반 탐구 과학수업에서 학생 평가활동에 대한 결과가 타당하다고 인식하는지 알아보기 위해 실험집단 학생들을 대상으로 실시한 설문조사와 인터뷰 자료를 분석하였다. 설문지는 동료평가결과와 자기평가결과가 타당하다고 생각되는지, 또한 평가 결과를 피드백으로 활용하여 자신의 학습에 도움을 주었는지에 대한 Likert 척도 문항과 사례를 서술하는 자유서술형 문항으로 구성되었다. 추가적으로 동료평가결과에 대해 수궁할 수 없었던 사례를 자유서술형으로 작성하도록 하여 분석하였다. Likert 5점 척도 문항은 전혀 그렇지 않다 1점, 그렇지 않다 2점, 보통이다 3점, 그렇다 4점, 매우 그렇다 5점으로 하여 평균점수를 구하였다. 자유서술형 문항에 대해서는 학생들의 응답을 유형별로 범주화하여 세부요소를 추출하였다. 모든 인터뷰 내용은 전사하여 분석하였으며 학생 응답을 유형별로 범주화하여 세부요소를 추출하였다. 범주화된 요소들은 과학 교육 전문가 1명에게 타당성을 검증받았으며, 최종적으로 검증된 요소들을 바탕으로 설문지를 다시 분석하여 세부요소의 빈도수를 결정하였다.

모든 평가들과 검사는 과학교육 전문가 1명과 과학 교육 박사과정 3명, 과학교육 석사과정 3명으로부터 내용 타당도를 검증받았다. 평가들의 신뢰도와 타당도를 높이기 위해 4인의 평가자는 각각 2개 반을 무작위로 선정하여 2차례에 걸쳐 채점하였으며, 채점자 간에 점수 차이가 발생한 경우 합의에 이를 때까지 지속적으로 협의하고 조정하여 분석의 신뢰도를 높였다.

이 연구에서 수집한 자료는 SPSS WIN 20.0 통계패키지 프로그램을 사용하여 정량분석을 하였장과 증거 형성에 효과가 있는지를 알아보기 위해 1차 활동보고서와 5차 활동보고서에서 학생들이 제시한 모둠의 주장과 증거를 분석하였다. 두 집단의 동질성을 확인하기 위해 1차 활동보고서의 주장과 증거에 대한 t-test를 실시하여 두 집단의 평균 차이를 검증하였다. 5차 활동보고서의 주장과 증거는 1차 반성 글쓰기 값을 공변량(covariate)으로 하여 일원 공변량 분석(one-way ANCOVA)을 하였다.

## 연구 결과

### 동질성 검사 결과

논의기반 탐구 과학수업에서 학생 평가활동을 적용한 실험집단과 적용하지 않은 비교집단 간의 동질성을 확인

**Table 2.** Claim and Evidence Analysis in the 1<sup>st</sup> report

	Experimental groups		Comparative groups		t	p
	M	SD	M	SD		
Claim	1.46	1.18	2.08	1.74	-0.27	0.79
Evidence	1.38	0.97	2.00	1.50	-1.77	0.86

하기 위해 평가 결과를 피드백으로 제공받기 전인 1차 활동에서 학생들이 작성한 모둠의 주장과 증거 점수를 SPSS 20 프로그램을 이용하여 독립표본 t-test로 분석하였다.

1차 활동보고서 분석 결과, 실험집단의 주장 평균은 1.46점, 비교집단의 주장 평균은 1.38점으로 실험집단이 비교집단에 비해 높았으나 통계적으로 유의미한 차이는 나타나지 않았다( $t = -0.27, p = 0.79$ ). 실험집단의 증거 평균은 2.08점, 비교집단의 증거 평균은 2.00점으로 실험집단이 비교집단보다 높게 나타났으나 통계적으로 유의미한 차이는 나타나지 않았다( $t = -1.77, p = 0.86$ ). 따라서 두 집단을 동질집단으로 간주하였다(Table 2).

#### 논의기반 탐구 과학수업에서 학생 평가활동이 모둠의 주장과 증거 형성에 미치는 효과

논의기반 탐구 과학수업에서 학생 평가활동이 모둠의 주장과 형성에 효과가 있는지 알아보기 위해 프로그램의 마지막 활동인 5차 활동보고서에 작성한 모둠의 주장과 증거를 분석하였다. 1차 활동보고서에서 두 집단의 주장과 증거 점수에 유의미한 차이가 나타나지 않아 동질집단으로 간주하였지만 평균값에 차이가 있었기 때문에 1차 활동보고서의 점수를 공변량으로 통제하여 일원 공변량 분석(one-way ANCOVA)을 통해 5차 활동보고서를 분석하였다.

**학생 평가활동이 주장 형성에 미치는 효과:** 프로그램 처치 후 실험집단과 비교집단 학생들이 5차 활동보고서에

작성한 모둠의 주장을 분석한 결과는 Table 3에 제시하였다. 분석 결과, 실험집단의 주장 평균은 3.75점, 비교집단의 주장 평균도 3.75점으로 두 집단 모두 1차 활동보고서에서 보다 높은 점수를 보였다. 그러나 두 집단 간의 유의미한 차이는 나타나지 않았다( $t = 0.00, p = 1.00$ ). 주장에 대한 세부요소별 분석 결과, 주장의 적절성 요소에서는 실험집단의 평균은 2.00점, 비교집단의 평균은 2.00점으로 통계적으로 유의미한 차이가 나타나지 않았다( $t = 0.00, p = 1.00$ ). 주장의 정확성 요소에서도 실험집단의 평균 1.75점, 비교집단의 평균 1.75점으로 통계적으로 유의미한 차이가 나타나지 않았다( $t = 0.00, p = 1.00$ ). 이를 종합해 보면, 두 집단 모두 1차 활동보고서에서 작성한 주장 점수보다 5차 활동보고서에서 작성한 주장 점수가 향상된 것을 통해 논의기반 탐구 과학수업이 학생들의 주장 형성에 효과적인 프로그램이나, 학생 평가활동을 적용한 실험집단과 적용하지 않은 비교집단 간의 유의미한 차이는 나타나지 않은 것을 알 수 있다.

**학생 평가활동이 증거 형성에 미치는 효과:** 논의기반 탐구 과학수업 활동에서 학생 평가활동이 모둠의 증거 형성에 효과가 있는지 알아보기 위해서 5차 활동보고서에서 학생들이 작성한 모둠의 증거를 적절성, 충분성, 정확성의 요소로 구분하여 분석하였다(Table 4).

분석 결과, 실험집단의 증거 평균은 4.54점, 비교집단의 평균은 3.75점으로 두 집단 모두 1차 활동에서보다 높은 점수를 보였으며, 실험집단의 평균이 비교집단보다 통계

**Table 3.** Claim Analysis in the 5<sup>th</sup> report

	Experimental groups		Comparative groups		t	p
	M	SD	M	SD		
Validity	2.00	0.00	2.00	0.00	0.00	1.00
Accuracy	1.75	0.44	1.75	0.53	0.00	1.00
Total	3.75	0.53	3.75	0.44	0.00	1.00

**Table 4.** Evidence Analysis in the 5<sup>th</sup> report

	Experimental groups		Comparative groups		t	p
	M	SD	M	SD		
Validity	1.83	0.48	1.71	0.55	-0.84	0.41
Sufficiency	1.58	0.58	1.08	0.65	-2.80	0.008*
Accuracy	1.13	0.45	0.96	0.20	-1.66	0.03*
Total	4.54	1.10	3.75	1.15	-0.29	0.04*

\* $p < .05$

적으로 유의미하게 높게 나타났다( $t = -2.09, p = 0.04^*$ ).

증거에 대한 세부요소별 분석 결과, 적절성 요소에서는 실험집단의 평균이 1.83점으로 비교집단의 평균 1.71점보다 높게 나타났으나 두 집단 간의 유의미한 차이가 나지 않았다( $t = -0.84, p = 0.41$ ). 충분성 요소에서는 실험집단의 평균은 1.58점, 비교집단의 평균은 1.08점으로 실험집단의 평균이 비교집단보다 통계적으로 유의미하게 높게 나타났다( $t = -2.80, p = 0.008^*$ ). 정확성 요소에서는 실험집단의 평균은 1.13점, 비교집단의 평균은 0.96점으로 실험집단의 평균이 비교집단보다 통계적으로 유의미하게 높게 나타났다( $t = -1.66, p = 0.03^*$ ).

5차 활동보고서에서 학생들이 작성한 모둠의 증거를 분석해 보면, 학생 평가활동을 실시한 실험집단 학생들이 평가활동을 실시하지 않은 비교집단 학생들보다 논의활동에서 증거를 형성하는 능력이 더 높다는 것을 알 수 있다. 또한 이 차이는 실험집단 학생들이 비교집단 학생들보다 적절한 증거를 제시하는 능력과 그래프, 표, 그림, 설명을 추가하여 충분한 증거를 제시하는 능력이 더 높기 때문에 나타나는 것으로 보인다.

다음은 5차 활동보고서에서 실험집단과 비교집단의 학생들이 작성한 모둠의 주장과 증거(Fig. 1, 2)에 대한 분석 내용이다.

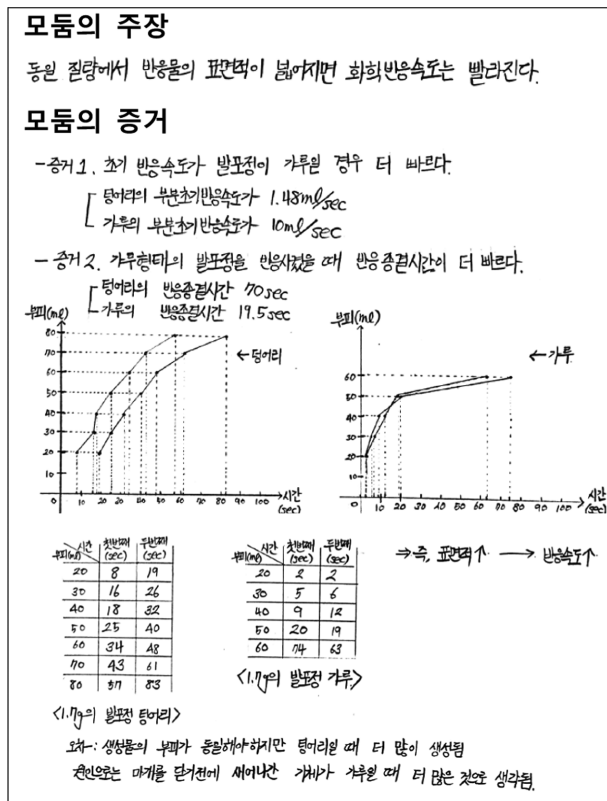


Figure 1. Experimental group student' writing sample.

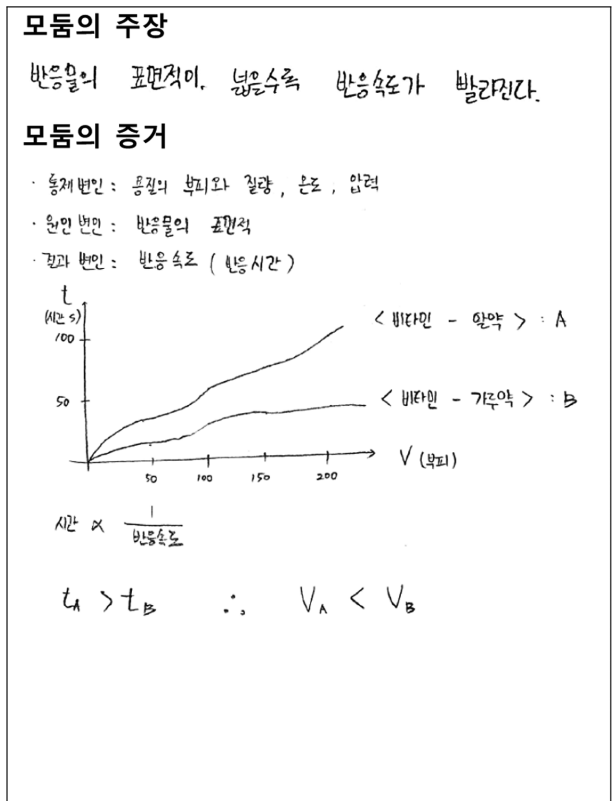


Figure 2. Comparative group student' writing sample.

실험집단 학생(Fig. 1)의 경우, 모둠의 주장에서 표면적과 화학반응속도 간의 관계를 구체적이고 정확하게 언급하고 있어 주장의 적절성과 정확성에서 각각 최고점인 2점을 부여하였다. 모둠의 증거에서 가루발포정기 알약발포정보다 같은 시간동안 더 많은 기포를 발생시켰다는 실험결과를 정리한 자료가 주장을 뒷받침하기에 적절하였으므로 증거의 적절성에서 2점, 실험결과를 설명과 그래프, 표를 사용하여 3개 이상 제시하였으므로 증거의 충분성에서 2점, 실험 결과에서 표와 그래프에서의 데이터도 정확하게 표시하였으며, 오차의 원인도 진술하고 있으므로, 증거의 정확성에서 2점을 부여하였다.

비교집단 학생(Fig. 2)의 경우, 모둠의 주장에서 표면적과 화학반응속도 간의 관계를 구체적이고 정확하게 언급하고 있어 주장의 적절성과 정확성에서 각각 최고점인 2점을 부여하였다. 모둠의 증거에서 통제변인, 원인변인, 결과변인을 작성하고 가루발포정기 알약발포정에 비해 같은 양의 기체를 발생시키는 시간이 더 짧다고 정리한 실험결과가 주장을 뒷받침하기에 적절하였으므로 증거의 적절성에서 2점을 부여하였다. 또한 설명과 그래프만 제시하였으므로 증거의 충분성에서 중간점인 1점, 그래프를 그렸으나 데이터를 정확한 수치로 제시하지 않았고 오차의

**Table 5.** Frequency of evaluation criteria in peer-assessment activities

	Fit with evidence	Sufficiency of evidence	Accuracy of evidence	Aesthetic criterion	Total
Topic 1	49(36.0%)	63(46.3%)	0(0.0%)	24(17.6%)	136(100%)
Topic 2	65(34.6%)	71(37.8%)	9(4.8%)	43(22.9%)	188(100%)
Topic 3	44(21.0%)	94(44.8%)	41(19.5%)	31(14.8%)	210(100%)
Topic 4	38(16.3%)	108(46.4%)	56(24.0%)	31(13.3%)	233(100%)
Topic 5	31(11.1%)	144(51.6%)	77(27.6%)	27(9.7%)	279(100%)

원인을 제시하지 않았으므로 최하점인 0점을 부여하였다.

### 논의기반 탐구 과학수업에서 학생 평가활동이 모둠의 주장과 증거 형성에 미치는 영향

학생 평가활동이 구체적으로 모둠의 주장과 증거 형성에 어떻게 영향을 미쳤는지 알아보기 위해 학생들이 활동보고서에서 작성한 동료평가 결과와 자기평가 결과를 분석하였다.

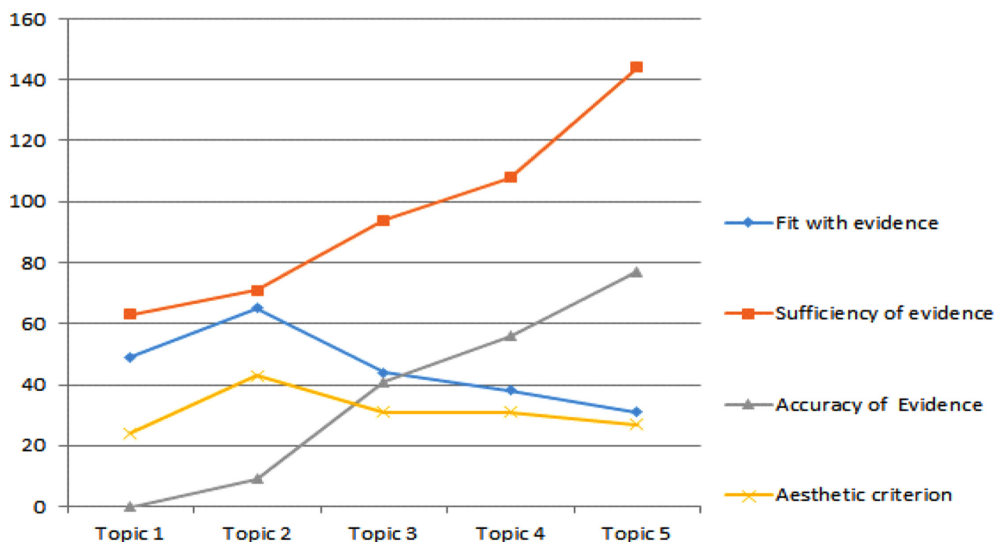
활동보고서에 학생들이 작성한 동료평가 결과는 실험 집단 학생들을 대상으로 실시한 설문조사에서 학생들이 평가 기준으로 제시한 증거의 적절성, 증거의 충분성, 증거의 정확성, 증거제시 방법의 4가지 기준에 따라 분석하였다.

동료평가 결과에서 나타난 세부요소별 빈도는 다음과 같다(Table 5, Fig. 3). 동료평가활동에서 나타난 동료평가의 빈도는 활동 1에서 136회, 활동 2에서 188회, 활동 3에서 210회, 활동 4에서 233회, 활동 5에서 279회로 활동이 진행될수록 증가하는 경향을 보였다.

증거의 적절성을 기준으로 평가한 빈도는 활동이 진행될수록 49회(36.0%), 65회(34.6%), 44회(21.0%), 38회(16.3%), 31회(11.1%)로 증가하다가 감소하는 경향을 보이고 있으

나, 한 활동에 대한 백분율로 분석한 결과로는 활동이 진행됨에 따라 감소하는 경향을 보였다. 증거의 충분성을 기준으로 평가한 빈도는 활동이 진행될수록 63회(46.3%), 71회(37.8%), 94회(44.8%), 108회(46.4%), 144회(51.6%) 순으로 대체적으로 증가하는 경향을 보였다. 증거의 정확성을 기준으로 평가한 빈도는 0회(0.0%), 9회(4.8%), 41회(19.5%), 56회(24.0%), 77회(27.6%) 순으로 빈도와 백분율 모두 증가하는 경향을 보였다. 또한 증거제시 방법을 기준으로 평가한 빈도는 24회(17.6%), 43회(22.9%), 31회(14.8%), 31회(13.3%), 27회(9.7%) 순으로 평가 빈도와 백분율 모두 증가하다가 감소하는 경향을 보였다(Fig. 4).

학생들은 활동 초기 동료평가활동에서는 모둠의 주장에 대한 증거가 적절한지, 그리고 증거를 충분히 제시하였는지를 평가 기준으로 많이 활용하였음을 알 수 있었다. 그러나 활동이 진행됨에 따라 주장에 대한 증거에 오차의 원인도 함께 제시하여 자료를 정확하게 제시했는지, 표와 그래프에 제시된 데이터를 정확하게 기록하였는지에 대해 평가하는 것을 알 수 있었다. 이는 논의기반 탐구 과학수업을 통해 활동이 진행될수록 학생들은 프로그램에 익숙해지면서 주장을 뒷받침할 수 있는 적절한 증거를 제시할 수 있게 되어 잘못된 증거를 제시하는 경우는 감소

**Figure 3.** Frequency of evaluation criteria in peer-assessment activities.

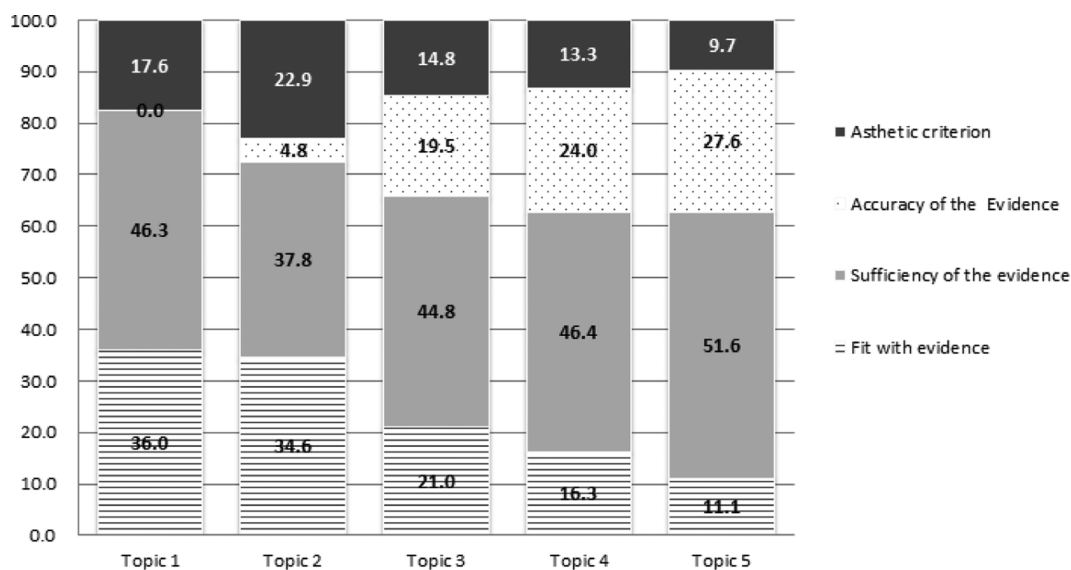


Figure 4. Percentage of evaluation criteria in peer-assessment activities.

Table 6. Frequency of evaluation criteria in self-assessment activities

	Fit with evidence	Sufficiency of evidence	Accuracy of evidence	Aesthetic criterion	Total
Topic 1	8(21.6%)	26(70.3%)	0(0.0%)	3(8.1%)	37(100%)
Topic 2	20(20.0%)	60(60.0%)	9(9.0%)	11(11.0%)	100(100%)
Topic 3	9(8.6%)	63(60.0%)	20(19.0%)	13(12.4%)	105(100%)
Topic 4	6(5.7%)	51(48.1%)	29(27.4%)	20(18.9%)	106(100%)
Topic 5	4(3.1%)	56(43.8%)	44(34.4%)	24(18.8%)	128(100%)

하였다. 이에 따라 증거를 평가할 때의 기준으로 표와 그래프를 활용하여 많은 증거를 활용하였는지, 제시된 자료에 오류 없이 정확한지를 평가 기준으로 활용하는 빈도가 늘어난 것으로 보인다.

자기평가 결과에서 나타난 세부요소별 빈도는 다음과 같다(Table 6, Fig. 5). 자기평가활동에서 나타난 평가 빈도는 활동 1에서 37회, 활동 2에서 100회, 활동 3에서 105회, 활동 4에서 106회, 활동 5에서 128회로 활동이 진행될

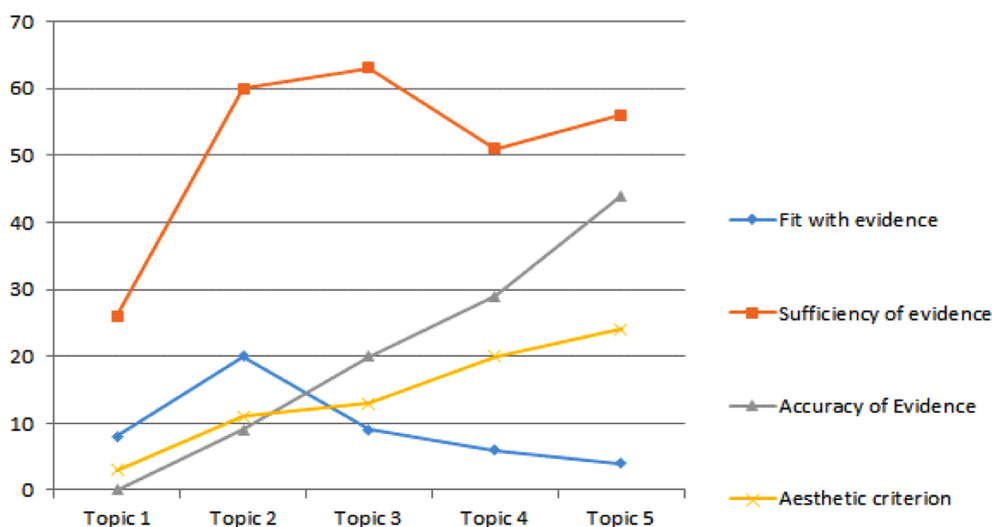


Figure 5. Frequency of evaluation criteria in self-assessment activities.



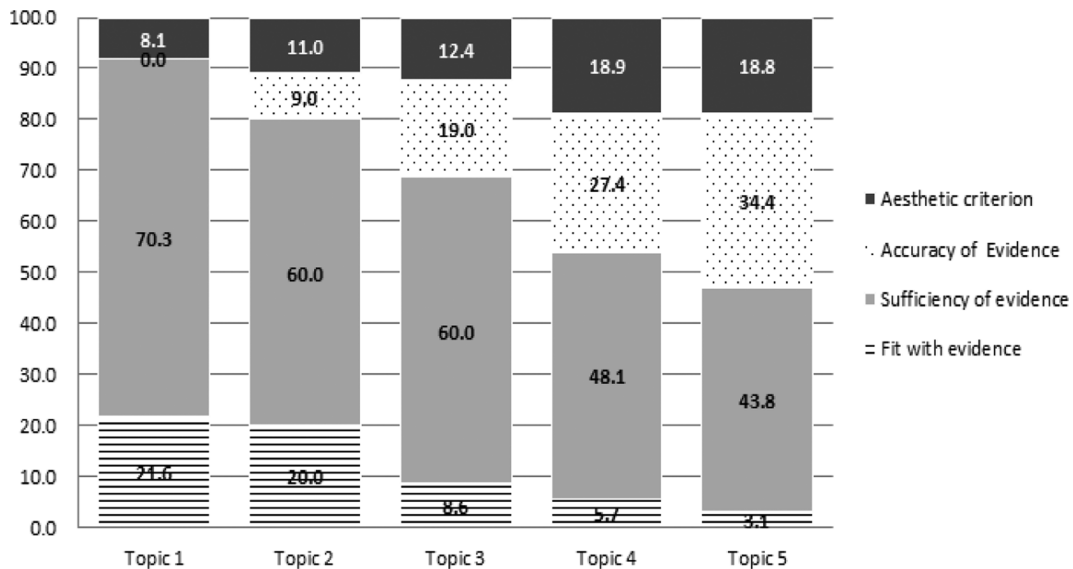


Figure 6. Percentage of evaluation criteria in self-assessment activities.

수록 증가하는 경향을 보였다.

증거의 적절성을 기준으로 평가한 빈도는 활동이 진행될수록 8회(21.6%), 20회(20.0%), 9회(8.6%), 6회(5.7%), 4회(3.1%)로 감소하는 경향을 보이고 있다. 증거의 충분성을 기준으로 평가한 빈도는 활동이 진행될수록 26회(70.3%), 60회(60.0%), 63회(60.0%), 51회(48.1%), 56회(43.8%)로 경향성이 나타나지 않으나 백분율로 비교한 결과로는 감소하는 경향을 보였다. 하지만 모든 활동에서 전체 평가 기준 중 가장 높은 비중을 차지하고 있었다. 증거의 정확성을 기준으로 평가한 빈도는 0회(0.0%), 9회(9.0%), 20회(19.0%), 29회(27.4%), 44회(34.4%) 순으로 평가 빈도와 백분율 모두 증가하는 경향을 보였다. 또한 증거제시 방법을 기준으로 평가한 빈도는 3회(8.1%), 11회(11.0%), 13회(12.4%), 20회(18.9%), 24회(18.8%) 순으로 대체적으로 증가하는 경향을 보였다(Fig. 6).

자기평가의 결과는 증거의 적절성에 대해 평가하는 빈도는 감소하고 증거의 정확성을 평가하는 빈도는 증가하는 것이 동료평가 결과와 유사한 경향성을 나타내고 있다. 증거의 충분성에 대한 경향성은 동료평가 결과와 다르게 나타나지만, 모든 활동에서 가장 빈도가 높게 나타난 점은 동료평가 결과와 같다. 이를 토대로 학생들은 표와 그림, 그래프나 다른 모둠에서 제시하지 않았던 증거의 수를 보강했는지 증거의 충분성에 대해서 꾸준히 평가하고 있으며, 오차의 원인과 변인통제, 제시된 그래프를 정확하게 그렸는지 등의 자료의 정확성 측면에 대해 평가하는 빈도는 증가하고 있음을 알 수 있다.

증거 제시방법에 대한 평가 기준에 대해서는 동료평가 결과와 다르게 대체적으로 증가하는 경향이 나타났다. 이

는 학생들이 자신의 모둠과 증거를 제시할 때, 여러 데이터를 하나의 표로 그렸는지, 다양한 색을 사용하여 중요한 부분을 강조하였는지 등 자료를 보는 다른 모둠의 학생들이 이해하기 쉽게 작성하였는지를 지속적으로 평가하고 있었음을 알 수 있다.

학생 평가활동의 타당성과 활용정도에 대한 학생들의 인식

논의기반 탐구 과학수업에서 학생 평가활동의 타당성을 점검하기 위해 다른 모둠의 주장과 증거에 대해 평가한 기준이 무엇인지 질문하였으며, 학생들이 서술한 응답은 주장과 증거의 적절성, 증거의 충분성, 증거의 정확성, 증거제시 방법의 4가지 범주로 분류할 수 있었다. 설문조사와 인터뷰 결과로부터, 대부분의 학생들은 학생 평가활동을 실시할 때 한 가지의 기준으로만 평가한 것이 아니라 여러 기준에 의거하여 평가했음을 알 수 있었다. 학생들은 주장과 증거가 적절한지를 먼저 평가한 이후에, 이를 만족시키면 증거가 충분한지를 평가했으며, 증거를 제시할 때 한 눈에 알아보기 쉽거나 이해가 쉽게 제시하였는지를 평가했다고 응답하였다. 주장과 증거를 제시할 때 어떻게 제시해야 한 눈에 알아보기 쉽거나 이해되기 쉬웠다고 생각하는지에 대한 추가 질문에는 단순히 글만 나열하는 것이 아니라 그림과 그래프, 표와 같은 다중표상 자료를 활용하는 경우라고 응답하였다. 따라서 학생들은 증거를 제시할 때 다중표상을 활용했는지를 평가의 중요한 요소로 활용하였음을 알 수 있다. 또한 활동이 진행될수록 증거의 정확성을 높이기 위해 오차의 원인을 제시하거나 그래프의 데이터를 정확하게 표기하는 등 증거 자료가 정확한지 점검한 것을 알 수 있었다.

다른 모둠의 학생들이 평가한 동료평가 결과는 타당하다고 생각하는지에 대한 Likert 5점 척도문항의 응답 결과 평균 3.89점의 점수를 보였으며, 학생들은 동료평가의 결과가 대체적으로 타당하다고 인식하고 있음을 알 수 있었다. 동료평가 결과 중 수공이 되지 않았던 구체적인 사례를 제시하여 보라는 설문에 '없다'라고 응답한 17명(21.6%)의 학생을 제외하고, 학생들이 서술한 응답 결과는 근거가 없는 평가와 증거제시 방법에 대한 평가의 두 가지 범주로 분류하였다. 이 중 31명(39.2%)의 학생들은 평가 근거가 명확하게 제시되어 있지 않은 결과를 수궁할 수 없다고 서술하였으며, 31명(39.2%)의 학생들은 증거제시 방법에 대한 의견 차이로 인하여 수궁할 수 없었다고 서술하였다.

동료평가 결과를 다음 활동에서 어떻게 활용되었는지에 대한 질문에 학생들은 증거를 보강하여 충분한 증거를 제시하는 데에 활용하였다고 응답한 학생이 45명(65.2%)으로 가장 높은 비율을 보였다. 또한 증거를 보강하는 방법으로 다중표상을 사용하고, 실험뿐만 아니라 교과서를 활용하여 증거의 개수가 증가하였다고 서술하였다. 증거제시 방법 개선에 활용하였다고 서술한 학생은 13명(18.8%) 나타났다. 기타 의견으로는 증거의 정확성을 점검하고, 수업 태도를 개선할 수 있었으며, 주장과 증거의 적절성을 점검하는 데에도 활용하였다고 응답하였다. 이를 통해 학생들은 동료평가 결과를 다음 활동을 위한 피드백으로 활용하고 있음을 알 수 있었다.

자기평가 결과는 타당하다고 생각하는지에 대한 Likert 5점 척도문항의 응답 결과 평균 3.96점의 점수를 보였으며, 동료평가 결과와 유사하게 대체적으로 타당하다고 인식하고 있음을 알 수 있다. 자기평가 결과를 다음 활동에서 어떻게 활용하였는지에 대한 질문에 수업 태도 개선에 활용되었다고 응답한 학생이 29명(45.3%)으로 가장 높은 빈도를 보였다. 증거를 보강하는데 활용하였다는 학생은 16명(25.0%), 증거의 정확성을 점검하는데 활용하였다는 학생은 15명(23.4%) 나타났다. 기타 의견으로 주장과 증거의 적절성을 점검하고 증거제시 방법을 점검하는데 활용하였다고 응답하였다. 학생들은 자기평가 결과를 증거의 정확성을 점검하고 증거를 보강하며 주장과 증거의 적절성을 점검하는데 활용하였음을 알 수 있다.

## 결 론

이 연구는 학생 평가활동을 적용한 논의기반 탐구 과학 수업이 고등학생들의 주장과 증거 형성에 미치는 영향에 대해 알아보고자 하였다. 이를 위해 학생 평가활동을 적용한 실험집단과 적용하지 않은 비교집단 학생들이 작성

한 모둠의 주장과 증거를 분석하여 프로그램의 효과를 알아보았으며, 실험집단 학생들이 작성한 동료평가 및 자기평가 결과를 분석하여 평가활동이 주장과 증거 형성에 미치는 영향을 알아보았다. 마지막으로 인터뷰와 설문조사를 통해 학생 평가활동의 타당성과 활용 정도를 알아보았다.

연구 결과, 학생들이 작성한 주장에 대해서는 두 집단 간의 유의미한 차이는 없었으나( $p>0.5$ ), 증거에 있어서는 실험집단이 비교집단보다 유의미하게 높게 나타났다( $p<0.5$ ). 특히 증거의 세부요소 중 충분성과 정확성 요소에서 유의미한 차이가 나타났다. 학생들이 작성한 동료평가와 자기평가결과를 분석한 결과, 활동이 진행될수록 증거의 적절성에 대한 평가 빈도는 모두 감소하는 경향이 나타났으며, 증거의 정확성에 대한 평가 빈도는 증가하는 경향이 나타났다. 또한 증거의 충분성에 대한 평가 빈도는 여러 평가 기준 중 가장 높은 비율로 나타났다. 논의기반 탐구 과학수업에서 활동이 진행됨에 따라 실험집단과 비교집단 학생 모두 적절한 주장을 제시하고, 주장을 뒷받침할 수 있는 타당한 근거를 제시하는 능력이 향상되었음을 알 수 있다. 하지만 학생 평가활동을 통해서 증거의 개수, 다중표상의 사용, 제시 방법의 다양성, 그리고 자료의 정확성에 대해 꾸준히 피드백을 받은 실험집단 학생들이 증거의 충분성과 증거의 정확성 요소에서 유의미하게 높은 점수를 나타낸 것으로 보인다.

학생들은 학생 평가활동에서 평가 기준으로 주장과 증거의 적절성, 증거의 충분성, 증거의 정확성, 증거제시 방법을 평가하였다고 응답했으며, 학생 평가활동에서 학생들은 동료평가와 자기평가의 결과가 타당하다고 인식하고 있었다. 동료평가의 경우 근거가 없는 평가나 글씨 크기나 글씨의 색깔 등의 증거제시 방법에 대한 평가의 경우 수공이 되지 않았다고 응답하였다. 학생들은 평가 결과를 다음 활동에 반영하여 활용하고 있다고 응답하였으며, 다중표상을 사용하고 증거의 개수를 늘려 증거를 보강하거나, 증거의 정확성을 점검하고, 증거제시 방법 개선, 수업 태도 개선, 주장과 증거의 적절성 점검에 활용하였다. 피드백으로써 제공되는 평가 결과를 바탕으로 학생들은 증거를 보강하며 증거의 정확성과 적절성, 증거의 제시 방법을 개선할 뿐만 아니라 수업에 조금 더 적극적으로 참여하게 되었다고 하였다.

이를 통해 논의기반 탐구 과학수업에서 학생 평가활동이 학생들의 주장과 증거를 형성할 때 주장의 타당성과 정당성을 높이는 증거 형성에 효과적이라는 것을 알 수 있다. 또한 동료평가 및 자기평가활동은 학생을 평가의 주체자로서 학습에 능동적으로 참여시켜 평가활동 자체가 학습의 일부분으로 통합되어 수업에 참여도를 높이고

논의능력을 향상시키는 등 학생들의 학습에 긍정적인 영향을 주었음을 알 수 있다.

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

## REFERENCES

1. Ministry of Education, *2015 Revised National Curriculum*. Notice No. 2015-74; Ministry of Education, 2015.
2. Kelly, G. J.; Chen, C.; Prothero, W. *Journal of Research in Science Teaching* **2000**, *37*, 691.
3. Newton, P.; Driver, P.; Osborne, J. *International Journal of Science Education* **1999**, *21*, 553.
4. Osborne, J. *Cambridge Journal of Education* **2002**, *32*, 203.
5. Sampson, V.; Clark, D. B. *Science Education* **2008**, *92*, 447.
6. Keys, C. W.; Hand, B.; Prian, V.; Collins, S. *Journal of Research in Science Teaching* **1999**, *36*, 1065.
7. Nam, J.; Kwak, K.; Jang, K.; Hand, B. *Journal of the Korean Association for Science Education* **2008**, *8*, 922.
8. Kwak, K. H.; Nam, J. H. *Journal of the Korean Association for Research in Science Education* **2009**, *29*, 400.
9. Nam, J. H.; Lee, D. W.; Cho, H. S. *Journal of the Korean Association for Research in Science Education* **2011**, *31*, 931.
10. Sung, H.; Nam, J. *Journal of the Korean Association for Research in Science Education* **2013**, *33*, 249.
11. Hand, B.; Hohenshell, L.; Prain, V. *Instructional Science* **2007**, *35*, 343.
12. Meyer, K.; Woodruff, E. *Science Education*. **1997**, *81*, 173.
13. Jang, K.; Nam, J. *Journal of the Korean Chemical Society* **2016**, *60*, 39.
14. Fiske, S.; Taylor, S. *Social cognition (2nd ed.)*; McGraw-Hill: New York, 1991.
15. Purchase, H. *Assessment and Evaluation in Higher Education* **2000**, *27*, 341.
16. Rust, C.; Price, M.; O'Donovan, B. *Assessment and Evaluation in Higher Education* **2003**, *28*, 147.
17. Smyth, K. *Assessment and Evaluation in Higher Education* **2004**, *29*, 369.
18. Boud, D.; Falchikov, N. *Rethinking Assessment in Higher Education*; Kogan Page: London, 2007.
19. Andrade, H.; Du, Y. *Assessment and Evaluation in Higher Education* **2007**, *32*, 159.
20. Nam, H.; Ryu, S. *Research Science Mathematics Education* **2000**, *23*, 55.
21. Li, L.; Liu, X.; Steckelberg, A. *British Journal of Educational Technology* **2010**, *41*, 525.
22. Liu, E. Z. F.; Lin, S. S. J.; Chiu, C. H.; Yuan, S. M. *IEEE Transactions on Education* **2001**, *44*, 246.
23. Xiao, Y.; Lucking, R. *The Internet and Higher Education* **2008**, *11*, 186.
24. Lee, S.; Bak, D.; Nam, J. *Journal of the Korean Association for Science Education* **2015**, *35*, 353.
25. Lee, S.; Nam, J. *Journal of the Korean Association for Science Education* **2016**, *36*, 347.
26. Biggs, J. B.; Moore, P. J. *The Process of Learning (3rd ed.)*; Prentice Hall: New York, 1993.
27. National Research Council. *Taking Science to School: Learning and Teaching Science in Grades K-8*; Duschl, R., Schweingruber, H., Shouse, A. Eds.; The National Academies Press: Washington, DC, 2007.
28. Sandoval, W.; Millwood, K. *Cognition and Instruction* **2005**, *23*, 23.
29. Driver, R.; Newton, P.; Osborne, J. *Science Education* **2000**, *84*, 287.
30. Maloney, J.; & Simon, S. *International Journal of Science Education* **2006**, *28*, 1817.
31. Nussbaum, E. *Educational Psychologist*. **2011**, *46*, 84.
32. Sampson, V.; Grooms, J.; Walker, J. *Science Education* **2010**, *95*, 217.