



초등학생 과학 탐구과제 수행 특성 분석 및 채점기준 개발*

함은혜¹, 이유경², 박소영², 박혜진³, 이성혜^{3*}
¹공주대학교, ²숙명여자대학교, ³한국과학기술원

Analysis on the Characteristics and Criteria Development in Performing Science Inquiry Tasks for Elementary School Students

Eun Hye Ham¹, You-kyung Lee², So-Young Park², Hyejin Park³, Sunghye Lee^{3*}
¹Kongju National University, ²Sookmyung Women's University, ³KAIST

ARTICLE INFO

Article history:

Received 7 March 2022
Received in revised form
4 April 2022
19 April 2022
Accepted 21 April 2022

Keywords:

Science-inquiry tasks,
Science-inquiry process
component, Science-inquiry
competence component,
Many-faceted Rasch model,
Performance scoring criteria

ABSTRACT

This study aims to develop performance criteria based on characteristics observed in science inquiry tasks for elementary school students. First, the performance characteristics by observing 70 fifth-grade elementary school students' science inquiry activity report are listed. Second, the checklist-type scoring criteria in connection with the theoretical framework of scientific inquiry process and relevant competencies are developed. Third, with the developed scoring criteria, 11 raters participate in scoring 350 students' reports. The main findings are as follow: first, the scoring data are well-fitted for the many-faceted Rasch model, and 22 scoring criteria are reasonably-well differentiated for various levels of proficiency. Second, at low performance level, observable characteristics are to answer questions explicitly required by the task or to observe objects or phenomena using pre-learned scientific concepts, while at high performance level, to explore additional data other than given data or to reflect on one's experimental process. Based on the results, the usefulness of analyzing students' performance characteristics for developing the scoring criteria, and further research directions are discussed.

1. 서론

과학 탐구활동은 학생들이 과학을 수행하는 실제적인 과정을 경험하게 함으로써, 과학적 지식의 본성을 이해하고 과학에 대한 긍정적인 태도를 가지게 하는 과학 교수-학습의 핵심적인 활동이다(Lee, 2005; Lee, Lee, & Lee, 2017; Lou, Blanchard & Kenndedy, 2015). 2015 개정 과학과 교육과정에서도 탐구중심 학습을 통해 교과외의 핵심 개념에 대한 이해와 탐구 경험을 통합하고, 과학적 사고력, 과학적 탐구 능력, 과학적 문제해결력, 과학적 의사소통 능력 등의 핵심역량을 함양할 것을 강조하였다(Ministry of Education, 2015). 이에 따라 학교를 포함한 교육 현장에서는 과제연구(프로젝트), 토의 및 토론, 실험 등의 학생 수행 중심의 수업 활동과 수행 관찰, 탐구보고서, 서술형 문항, 과학글쓰기를 통해 탐구 수행 과정이나 결과를 평가하는 방법의 활용이 꾸준히 확대되어 왔다(Baek *et al.*, 2020; Koh & Jeong, 2014). 그러나 여전히 이와 같은 학생 수행 특히, 탐구중심 수업과 평가를 운영하는 데는 여러 어려움이 있는 것으로 알려졌으며, 교사들은 평가 시간이 많이 소요된다, 평가기준이 모호하다, 평가항목을 만들기 어렵다, 평가 방법을 잘 모르겠다 등 주로 평가와 관련된 어려움을 호소하고 있다(Baek *et al.*, 2020; Koh & Jeong, 2014).

이에 과학 탐구활동 평가를 위한 평가 요소를 도출하고 이를 바탕으로 평가도구를 개발하는 연구들이 지속적으로 이루어지고 있다.

다수의 선행연구들은 과학 탐구활동 또는 탐구과제를 평가하기 위해 수행의 과정 또는 절차적 요소를 도출하고, 도출된 각 요소를 평가 요소로 간주하여 그 내용을 상세화하는 방식으로 평가기준을 개발하였다(Arnold *et al.*, 2018; Hwang, Lee, & Kwak, 2010; Lou, Blanchard & Kenndedy, 2015; Kim & Kim, 2012; Timmerman *et al.*, 2011). 과학 탐구의 과정 혹은 절차적 요소들은 평가의 범위 혹은 과제의 특성에 따라 다양하게 제시되어 왔는데, 대체로 가설 생성, 실험 설계, 실험 수행 및 자료 해석, 결론 도출 및 평가를 포함한다. 이러한 연구들은 과학 탐구과제를 설계하거나 탐구활동을 평가할 때, 어떤 과정 또는 절차적 요소들을 포함해야 하는지 나아가 어떤 내용이 평가의 영역이 될 수 있는지를 안내해준다. 그러나 과정 요소의 개념 자체가 실제 학생들의 과제를 다양한 수행 수준에 따라 분류하거나 변별하는 데 유용하게 활용되기는 어렵다.

또 다른 연구들은 이론적으로 도출한 평가 요소와 그에 근거하여 상세한 평가기준이 신뢰로운 채점결과를 산출하는지를 경험적으로 검토하기도 하였다(Choi, Lee, & Chae, 2017; Lee & Jeong, 2013; Lee, Lee, & Lee, 2017; Park & Kang, 2012; Song, Lee & Lim, 2015). 이러한 연구들은 도출된 평가 요소별 성취 수준을 단계별(상, 중, 하 또는 우수, 보통, 미흡)로 구분하여 진술하고, 그에 따라 점수를 부여하는 방식(분석적 루브리 활용)을 채택하거나, 특정 평가요소에서의 성취 수준을 채점자가 임의로 판단하여 상, 중, 하 또는 우수, 보통,

* 이 논문은 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2020S1A3A2A02095447).

교신저자: 이성혜 (slee45@kaist.ac.kr)

<http://dx.doi.org/10.14697/jkase.2022.42.2.239>

미흡 등의 차등 점수를 부여하는 방식을 채택하였다. 전자가 후자에 비해 채점자 간 일관성이 높을 뿐만 아니라, 학생들에게 수행 수준별로 구체적인 피드백을 제공하는데 유용한 것으로 알려졌지만, Baek et al.(2020)의 연구에 따르면 실제 과학탐구 수업에서 교사들이 평가 요소별 성취 수준이나 특성을 단계별로 구분, 진술하여 평가에 활용한 비율은 매우 낮았다. 이는 평가 요소별×단계별 성취 수준에 대한 서술자 개발에는 상당한 전문성과 노력이 요구되기 때문이기도 하며, 평가 요소별×단계별 성취 수준에 대한 서술자가 제공된다 하더라도 학생들의 실제 수행을 변별하기 어려울 정도로 모호한 경우도 많기 때문이다.

이러한 문제의식에 기반하여, 이 연구에서는 과학 탐구과제에서의 수행 수준을 효과적으로 변별할 뿐만 아니라 개별 학생들의 수행에 대해서 체계적이면서도 즉각적인 진단과 피드백을 제공할 수 있는 채점기준을 개발하고자 하였다. 이 연구에서 개발, 제안하는 채점기준의 특징과 강점을 선행연구와의 차별성을 중심으로 기술하면 다음과 같다.

첫째, 선행연구에서는 과학탐구 내용 혹은 과정 요소로부터 채점기준을 논리·연역적으로 도출하고자 했던 반면, 이 연구에서는 학생들의 실제 수행의 특징을 분석하여 과학탐구 과정 요소들과 연계하는 경험·귀납적 접근을 취하였다. 이를 위해 학생들의 탐구활동보고서를 관찰하여 다양한 수행 수준의 차이를 변별할 수 있는 특징들을 중심으로 목록화하였으며, 그 특징들이 선행연구에서 도출된 과학탐구 과정 요소 및 역량 요소들과 어떻게 연계되는지 검토하였다.

둘째, 다수의 선행연구에서는 채점기준을 루브리 형태로 개발하였으나, 이 연구에서는 채점의 효율성과 일관성을 높이기 위하여 체크리스트형 형태로 개발하였다. 체크리스트형 평가는 평가 대상이 핵심적인 요소를 포함하고 있는지에 중점을 두고 평가하고자 할 때 적합한 형태로(Popham, 1995; Ruiz-Primo, Shavelson, 1996; Schafer, 1997), 채점기준이 비교적 명확하여 채점자 간의 일치도가 높은 장점이 있다(Cho, Mun, & Kim, 2010). 이에 채점자의 인지 부하를 최소화하면서 탐구과정의 핵심적인 수행의 유무를 판단하여 전체적인 수행 수준을 평가할 수 있도록 체크리스트형 채점기준을 개발하고자 하였다.

셋째, 채점데이터 분석에 다국면 Rasch모형을 활용하여 학생의 능력(숙달도)과 채점항목의 난이도, 채점자의 엄격성을 단일 척도에서 직접 비교 가능하도록 추정하였다. 이를 통해 채점항목으로 구체화된 다양한 과제 수행의 특징들이 다양한 능력 수준의 학생들을 변별하는데 어떻게 기능하는지를 구체적으로 이해하고자 하였다.

요약하면, 이 연구에서는 과학 영역에서의 비교적 비구조화된 과제를 통해 학생들이 실험을 진행하고 결과를 기술하도록 한 탐구활동보고서를 분석하여 보고서에 나타난 탐구과제 수행의 특징을 분석하고 선행연구를 바탕으로 도출된 과학탐구 과정 요소별로 분류하여, 각 요소별 채점기준을 체크리스트 형태로 개발하고자 한다. 또한, 각 채점기준이 어떠한 역량 요소를 반영하고 있는지 함께 검토함으로써 학생들의 탐구 과정 요소별 수행을 입체적으로 이해하고 진단하고자 한다(Park et al., 2021). 구체적인 연구문제는 다음과 같다.

첫째, 개발된 체크리스트형 채점기준표를 활용하여 생성한 채점데이터가 피험자, 채점자, 채점기준 국면을 고려한 다국면 Rasch모형에 적합한가? 개발된 채점기준은 학생들의 수행수준을 변별하는 데 적절하게 기능하는가?

둘째, 과학 탐구과제 과정 요소별 채점기준의 난이도 분포는 어떠한가? 학생 수행 수준에 따라 과제 수행의 특징이 어떻게 다른가?

II. 선행연구 검토

1. 과학 탐구과제의 평가

과학 탐구과제 수행의 특징을 분석하기 위해서는 탐구문제 선정 후 실험을 설계하고 가설 검증을 위한 데이터를 수집하여 결과를 해석, 일반화하는 과정을 포함하는 탐구과정 요소를 도출하고 각 요소의 평가기준을 구체화할 필요가 있다(Choi, Lee, & Chae, 2017; Lee, Lee, & Lee, 2017). 탐구과정의 요소를 제시한 선행연구들을 구체적으로 살펴보면, Choi, Lee, & Chae(2017)의 경우 과학 탐구과제의 평가기준 중 하나로 ‘탐구능력’을 제시하고 문제인식, 가설설정, 실험설계, 자료수집 및 해석, 그리고 결론 도출이라는 다섯 가지 하위요소를 제시하였다. Kim & Kim(2012)는 과학 탐구 사고력 측정을 위한 서술형 평가도구 개발을 진행하였는데 평가 요소를 좀 더 세분화하여 탐구문제 인식(탐구문제가 무엇인지 추론함), 변인 인식(통제변인, 독립변인, 종속변인을 판별함), 가설 설정(실험 설계가 주어졌을 때 가설을 추론하거나 주어진 실험 상황에 대해 가설을 설정함), 실험 설계(가설에 따라 통제변인, 독립변인, 종속변인을 포함하여 실험을 설계함), 자료 변환(데이터를 그래프로 나타냄), 자료 해석(그래프 자료를 통해 변인의 관계를 해석함), 결론 도출(자료를 종합하여 알맞은 결론을 진술함), 일반화(실험 결과를 통해 공통된 원리를 찾아냄), 평가(실험 설계에서 부적합한 부분을 찾아내어 올바르게 수정함)인 9가지를 제안하였다.

Kim & Kim(2012)와 유사하게 과정 요소를 도출하였지만 각 요소에 대해 좀 더 세부적인 요소를 제시함으로써 평가목표를 구체화한 경우도 있다. Hwang, Lee, & Kwak(2010)의 평가도구가 이에 해당하는데, 이들은 과학 과정기능으로 7개 범주를 제시하고 각 범주에 대해 2-5개의 요소를 구체화하여 총 27개의 세부 과학 과정기능 요소를 설정하였다. 7개 범주와 27개 과정기능 요소는 문제 및 가설 설정(문제 설정, 가설 설정), 변인확인 및 실험통제(독립변인 확인, 종속변인 확인, 통제변인 확인, 실험 통제), 문제해결 방법설계(관찰 설계, 실험 설계, 표 설계, 그래프 설계, 결과 예상), 수행(분류, 정성적 관찰/측정, 정량적 관찰/측정, 실험기구 조작, 결과 기록/기술, 실험기법 해석), 결과분석(실험결과를 표로 요약, 산술적 계산, 자료를 그래프로 그리기, 정량적 관계 결정), 결과해석(결론 도출, 추리, 일반화), 적용(결과에 근거한 예상, 자연현상에 대해 실험결과 적용, 동일문제를 다른 예에 적용)이다.

정리하면 과학 탐구과제의 탐구과정 요소를 평가한 도구들은 과정 요소의 상위범주를 대체로 10개 이내로 제시하되, 평가의 범위 및 목표에 따라 하위범주를 좀 더 구체화하여 제시하는 경우가 있었다. 이 연구에서는 선행 평가도구의 과정 요소들을 종합하는 동시에 Hwang, Lee, & Kwak(2010)의 평가도구와 같이 세분화된 채점항목을 제시하되, 연구 목적에 맞게 체크리스트형 채점기준에 부합할 수 있는 항목들을 개발하고자 한다. 체크리스트형 문항은 수행 수준에 대한 세밀한 측정에 제한이 있지만 평가자의 주관적 해석을 최소화하

Table 1. Science-inquiry process components

과학탐구 과정 요소		평가지표
요소명	하위 내용요소	
가설 생성	종속변인 규명	종속변인을 판별하고 실험 전 상태를 관찰함
	독립변인 탐색	실험 설계를 위해 필요한 실험 조건을 탐색함
	가설 정당화	검증 가능한 가설을 설정하고 근거를 제시함
실험 설계 및 수행	독립변인 규명	가설 검증을 위해 필요한 독립변인을 찾아냄
	혼재변인 통제	주어진 실험 상황에서 가설을 검증하기 위해 필요한 혼재변인(통제변인)을 찾아냄
	실험 수행	가설을 바탕으로 독립변인을 포함한 실험조건을 조정하여 실험을 수행함
자료 분석 및 해석	가설 검증	실험결과를 바탕으로 가설 검증 결과를 도출함
	자료 해석	실험을 통해 얻은 자료를 해석함
결론 도출 및 평가	평가	실험결과들을 종합하여 결론을 도출하고 실험과정을 평가함
	학습과정 성찰	실험 수행을 통해 배운 것을 확인함

고 평가항목이 다소 많더라도 해당 수행의 유무만을 통해 전체적인 수행 수준을 평가할 수 있다는 객관성 및 실용성 측면에서의 장점이 있다. 이러한 체크리스트형 채점항목 개발을 위해 선행연구들을 종합하고 이 연구목적에 맞게 수정·보완한 과학 탐구과제의 과정 요소와 평가지표는 Table 1과 같다.

또한 실제 이러한 평가 요소를 적용하여 평가하고 평정자 간의 일관성을 높이기 위한 채점 기준을 개발하는 연구들이 다수 이루어져 왔다. 이러한 연구들은 과학 탐구 활동 평가 요소별로 수행 수준과 단계에 구분을 두어 채점할 수 있도록 제시한 특징이 있다. 예를 들어 Lee & Jeong(2013) 고등학교 생명 과학 수업의 맥락에서 과학 글쓰기를 활용한 과학적 사고력 평가도구를 개발하였는데, 이 연구에서는 과학적 사고의 하위 영역을 귀납적 사고력, 연역적 사고력, 비판적 사고력, 창의적 사고력으로 구분하고 각 영역에 대한 평가 준거를 제시한 후 학생의 수행 수준에 따라 이를 상, 중, 하 그리고 0점으로 구분하여 평가할 수 있는 루브릭을 제시하였다. Choi, Lee, & Chae(2017)의 경우에도 과학 탐구 과정에서 요구되는 사고역량을 지식, 논리분석적 사고, 비판적 사고, 탐구능력, 창의적 사고, 문제해결 능력으로 제시하고 각각을 4단계로 나누어 평가할 수 있는 루브릭을 개발하였다. Lee, Lee, & Lee(2017)는 중학생의 과학 탐구 과정 평가를 위한 루브릭을 개발하였는데, 탐구 과정의 평가 요소를 기초 탐구(관찰, 분류, 측정, 예산, 추리)와 통합 탐구 과정(가설 설정, 변인 통제, 자료 해석, 결론 도출, 일반화)로 나누어 도출하고 각 평가 요소에 대해 ‘우수’, ‘보통’, ‘미흡’ 수준의 수행 특성을 기술하여 채점할 수 있도록 하였다. Song, Kil, & Shim(2015)은 초등 과학 영재학생들의 과학 탐구 능력을 평가하기 위한 평가문항을 개발하고 이를 평가에 적용하였다. 이들이 개발한 평가도구는 과학 탐구 능력의 평가 영역을 관찰 및 추리, 실험 설계, 실험 결과로 구분하고 각 영역에 대해 다시 평가 준거를 제시하였다. 특히, 이 평가도구는 학생들의 서술형 답안에 대해 채점 기준을 차등해서 점수를 부여할 수 있도록 구성하였는데(0점~3점), 모범답안과 이를 기준으로 부분점수를 부여할 수 있는 기준을 제시하였다.

이러한 평가도구들은 탐구과제 평가에서 수준에 따라 채점 기준을 달리 제시하여 차등 점수를 주는 방식에서 대동소이하나, 수준에 따른 채점 기준 또는 행동 특성의 구체성에서 차이가 나타난다. 일부 평가도구는 수준에 따른 구체적인 채점 기준 없이 평가자가 차등 점

수를 부여하는 방식으로 제시되기도 한다(Park & Kang, 2012; Baek et al., 2020). 실제 과학탐구 수업에서 교사들이 활용한 평가도구를 분석한 Baek et al.(2020)의 연구에 따르면, 평가도구를 구성할 때 세부 평가 요소들을 수준별로 구분하여 단계를 진술한 비율은 매우 낮은 것으로 나타났으며, 특히 수준에 따라 구체적인 성취 수준이나 특성을 제시한 평가도구는 매우 소수로 파악되었다. 평가 요소별로 배점을 제시하고, 수준을 구분하여 차등 배점하도록 하는 경우에도 교사들은 ‘평가기준이 모호’하다고 인식하며 보다 세부적인 평가기준이 필요함을 강조하였다.

Timmerman 외(2011)는 과학수업에서 학생들의 수행 결과를 효율적이고 일관성 있게 평가할 수 있으면서, 학생들의 수행의 수준을 파악하고 비교할 수 있는 평가도구의 필요성을 강조한다. 또한 평가로부터 도출된 점수가 학생의 수행 수준에 대한 의미있는 정보를 제공할 수 있어야 한다고 언급하였다. Baek et al.(2020) 역시 교사가 평가결과로부터 즉각적으로 피드백을 제공할 수 있는 도구를 개발하는 것이 중요하며, 상위 수준의 행동 특성과 현재의 행동 특성의 차이를 확인하고 이에 근거하여 전략과 방안을 피드백 할 수 있어야 한다고 제안하였다. 그러나 선행연구에서 개발된 평가도구를 적용하여 탐구 수행 수준에 따른 행동 특성 차이를 분석한 연구는 거의 수행되지 않았다.

이와 같이 탐구과제 평가도구는 각 평가 요소에 대해 단계별로 구분하여 성취 수준을 진술하고 차등 점수를 부여하는 방식 또는 채점 기준 내에서 수준에 따라 상, 중, 하 또는 우수, 보통, 미흡 등으로 평정하도록 하는 방식으로 구성되었으며, 일부 연구에서 이를 적용하여 학생들의 과학 탐구 수행을 분석하고자 하였다. 그러나 학습자의 수행 수준에 따라 과학탐구 행동이 다르게 관찰되는지 확인한 연구는 없는 상황이다. 따라서 이 연구에서는 과학 탐구과제 채점 기준을 개발하고, 이를 적용하여 수행 수준에 따라 탐구와 관련된 행동이 다르게 나타나는지 확인하여 우수한 수행을 보이는 학습자의 행동 특성을 파악하고자 한다.

2. 과학 탐구과제 수행 관련 역량

Table 1에서 제시된 각 하위요소의 평가지표를 바탕으로 제시된 채점항목들이 과제 수행 특성을 체계적으로 반영하도록 하기 위해서

는 해당 과제의 수행을 위해 어떠한 능력 또는 지식이 요구되는지를 파악하여야 한다. 과학 탐구과제 평가 문헌들에 따르면 과학 분야에서의 탐구활동이란 자연 현상에서 도출된 인과적 질문을 과학적 ‘지식’과 과학적 ‘방법’을 활용하여 논리적으로 해결함을 의미한다(Park & Kang, 2012). 즉, 과학적 탐구활동을 위해서는 기본적으로 과학적 지식이 바탕이 되어야 하고, 이에 더불어 문제 해결을 위해 필요한 방법을 고안하고 수행해나갈 수 있는 탐구적 사고력이 필요하다는 것이다. 이렇듯 선행연구들은 과학 탐구과제 수행에 요구되는 역량을 내용적 지식과 탐구적 사고력 두 가지를 공통적으로 제시하고 있다(Choi, Lee, & Chae, 2017; Lee & Jeong, 2013; Park & Kang, 2012; Son, 2006).

과학적 지식이란 자연 현상이나 사실에 대한 체계적인 관찰, 실험, 분석 등을 통해 검증된 체계화된 정보를 의미하며, 과학적 탐구활동 중 가설을 설정할 때 중요한 역할을 한다. 가설을 연역적으로 도출할 때에는 기존의 과학적 지식을 바탕으로 하기 때문이다(Lee & Jeong, 2013). 가설을 검증한 후에 결론을 도출하고 일반화하는 귀납적 문제 해결과정에서도 일련의 과학적 지식이 필요하다는 점에서 과학적 탐구활동에서 과학적 지식은 필수적인 요인이라고 볼 수 있다.

이렇듯 과학 탐구과제의 수행을 위해서는 자연 현상, 현실을 이해하고 설명할 수 있는 기본적인 지식이 요구되지만 이러한 지식이 적절하게 기능할 수 있도록 사고할 수 있는 능력 또한 중요한 한 축이라고 할 수 있다(Park & Kang, 2012; Shim & Yoo, 2020). 특히 과학적 지식은 절대불변의 특성의 것이 아니라 계속적으로 변해가는 것이므로 어떠한 대상 및 현상과 관련된 문제에 대해 논리분석적 사고를 통해 자료를 분석하고 종합할 수 있는 능력은 탐구활동에서 필수적인 역량이라고 할 수 있다(Kim, 1993). 논리분석적 사고는 가설을 검증하는 과정, 현상에 대한 원인을 설명하고 예측하는 과정, 근거를 바탕으로 결론을 유도하는 과정을 포함한 과학 탐구활동의 전반적 과정에 적용된다(Lee & Jeong, 2013; Son, 2006). 논리분석적 사고를 구성하는 세부 능력에 대해 국내 과학 탐구과제 연구들을 분석한 한 메타분석에서는 보존, 비례, 조합, 확률, 상관, 변인통제, 종합적 비교를 논리분석적 사고의 구성 요인들로 보고하기도 하였다(Kim & Kim, 2009).

과학 탐구과제의 성공적인 수행을 위해서는 과학적 지식과 과학적 사고뿐만 아니라 탐구과제를 수행해나가는 태도 또는 자세 또한 중요한 요인이라고 볼 수 있다. 특히, 최근 과학교육과정은 과학 교과에서는 지식의 단순한 암기를 넘어서서 직접적인 활동을 통한 능동적인 학습을 강조하기 때문에 학습자 중심의 탐구 활동이 매우 중요하다(Ministry of Education, 2015; Lee, Lee, & Lee, 2017). 따라서 이 연구에서는 학생 스스로 탐구 활동에 필요한 추가 자료를 탐색하고 본인의 탐구 활동에 대해 평가하고 성찰하는 탐구적 태도를 과학 탐구과제 수행을 구성하는 하나의 주요 요인으로 포함하였다. ‘태도’라는 용어가 다소 포괄적인 의미를 가질 수 있기 때문에 본 연구에서는 과학적 태도를 주어진 탐구과제의 범위를 넘어서서 능동적으로 과제 외의 과학적 개념 혹은 지식을 연결하거나 새로운 자료를 탐색하는 확장적 활동, 그리고 과제를 통해 배운 것을 스스로 반추하는 반성적 활동으로 제한하여 정의하였다.

이와 더불어 비구조화된 과제에 대해 자유롭게 기술하도록 하는 이 연구에서의 평가도구의 특성상 의사소통능력 또한 과학 탐구과제 수행을 위한 요인으로 추가하였다. 실제로 2015 개정 교육과정뿐만

아니라 일부 연구에서는 탐구 수행 결과를 보고하는 의사소통 능력을 과학 탐구과제의 한 과정(기능) 요소로 제시하기도 하였다(Baek *et al.*, 2020; Ha *et al.*, 2018; Shim & Yoo, 2020). 구체적으로 2015 개정 교육과정에서는 과학과 교과역량으로 과학적 사고력, 과학적 탐구능력, 과학적 문제 해결력, 과학적 참여와 평생학습 능력과 더불어 과학적 의사소통 능력을 도입하고, 이를 ‘과학적 문제 해결 과정과 결과를 공유하고 발전시키기 위해서 자신의 생각을 주장하고 타인의 생각을 이해하며 조정하는 능력’으로 칭하였다. 이 능력에는 다양한 양식과 매체를 통하여 과학적 정보를 표현하고 증거에 근거하여 논증을 하는 능력을 포함함을 명시하였다(Ha *et al.*, 2018). 이 연구에서는 기술의 조직화, 명료성, 풍부성을 바탕으로 과학적 의사소통 능력을 평가하고자 하였다.

III. 연구방법

1. 연구 맥락 및 연구 대상

연구를 위하여 A대학 온라인 교육프로그램을 수강한 초등학교 5학년 350명 학생(남학생 231명, 여학생 119명)이 제출한 탐구활동보고서를 활용하였다. 학생들은 온라인 교육프로그램에 참여를 위해 시스템에 회원가입을 하면서 학생 본인 및 법정 대리인의 개인정보 수집, 보호 및 연구 목적의 데이터 활용에 대해 자발적 동의를 하고 참여하였다. A대학 온라인 교육프로그램은 수학, 과학 분야에 흥미가 높은 전국 초·중·고등학생을 대상으로 하며, 학생들은 별도의 선발 과정 없이 관심있는 프로그램을 선택하여 수강할 수 있다. 프로그램은 실생활과 연계된 과학 지식을 학습하고 관련 문제를 탐구함으로써 과학적 사고력과 문제해결력을 신장하도록 설계되었다. 첫째, 개념학습 단계에서 학습자는 국가 교육과정에서 다루고 있는 실생활과 관련된 개념, 문제해결에 필요한 개념을 학습관리시스템(LMS)에 탑재된 e-Book 콘텐츠를 통해 학습한다. 둘째, 문제해결 단계에서는 사전학습 단계에서 배운 개념이나 원리를 적용·확장·심화하여 새로운 현상을 탐구하거나 새로운 아이디어를 제안하는 과제를 수행한다. 특히, 탐구과제는 가설생성, 실험설계, 자료 분석 및 해석, 결과 도출 및 개선 방안 제안 등과 같은 절차에 따라 실험은 수행하고 그 과정과 결과를 기록하여 탐구활동보고서를 제출하게 된다. 탐구과제는 두 개 혹은 세 개의 미션으로 되어있으며, 각 미션에는 1~3개의 하위 문항이 포함되어 있다.

이 연구에서 활용한 탐구과제 주제는 ‘치즈는 왜 맛이 다를까?’로 개념학습에서는 응고, 발효, 효소와 같은 과학적 개념과 함께 치즈의 역사, 종류, 치즈 원료 및 분류, 치즈에 포함된 영양소, 치즈를 활용한 요리를 학습하게 된다. 주요 개념학습 내용은 2015 교육과정 초등학교 5~6학년 과학 교과의 균류, 세균의 특징과 기술가정 교과의 식생활, 식품과 영양 등에 기반하였으며, 그 외 추가적인 내용은 개념학습에 설명과 예시를 제공하여 학습이 이루어지도록 하였다. 탐구과제에서는 개념학습 부분에서 사전에 학습한 내용을 바탕으로 리코타 치즈를 만들어 보고 레시피에서 조작 변인을 설정하여 분석하고, 리코타 치즈를 직접 만들어 본 후 변화시킨 조건에 따라 결과가 어떻게 달라지는지, 치즈에 나타나는 현상은 무엇이고 치즈의 특징은 어떠한지 결과를 분석·정리하여 보고서로 제출한다. 학습자는 탐구중심 과제

Table 2. Contents of scientific inquiry activity report

과제 단계	내용
과제 안내	[미션명] 리코타 치즈레시피를 해부해보자! 지금까지 우리는 치즈가 무엇인지, 치즈는 어떻게 만들어지는지, 다양한 치즈의 종류와 함유된 영양소가 무엇인지 배웠습니다. 앞서 배운 내용을 바탕으로 본격적으로 치즈를 만들어보면서 미션을 해결해봅시다. 이번 미션에서는 기본 리코타 치즈의 레시피에서 조자 변인을 설정하여 분석하고, 여러분만의 특별한 치즈를 만들어 봅시다.
	[리코타 치즈 레시피] (치즈 재료와 조리 절차를 5개 단계로 안내함. 구체적인 내용은 생략)
미션1. 치즈의 특징 관찰	[지시문] 앞서 나온 레시피를 따라 ‘기본 리코타 치즈’를 만들어 보고 특징을 적어봅시다. [도움말] 여러분은 기본 리코타 치즈 레시피를 충실히 따랐나요? 이 기본 레시피는 어떻게 만들어진 것일까요? 끓이는 시간, 불의 세기, 우유/생크림의 비율, 레몬즙의 양, 물기를 짜주는 정도 등의 조건들은 치즈의 특성에 어떠한 영향을 미칠까요?
미션2. 치즈의 특성을 변화시키는 조건 탐색	[지시문] 리코타 치즈 레시피에서 치즈의 특성을 다르게 변화시킬 수 있는 조건을 적어도 2개 이상 생각해봅시다. 변화시킨 조건에 따라 치즈에 어떤 결과가 기대되는지 역시 설명해 주세요. (조건 ex. 끓이는 시간)
미션3. 실험 수행 과정 및 결과 보고	[지시문] 미션 2에서 제시했던 조건 중 하나를 선정하여, 그 조건에 변화를 주며(ex. 물을 덜 넣는다 / 더 넣는다) 치즈를 직접 만들어 보고 결과를 분석해 봅시다.

를 수행하면서 기초 탐구과정인 관찰, 측정, 예상, 추리와 통합 탐구과정인 가설 설정, 변인 통제, 자료 해석, 결론 도출과 같은 탐구와 문제 해결 절차를 경험할 것으로 기대되었다. 탐구활동보고서의 구성 내용이 Table 2에 소개되었으며, 실제 활용한 보고서 양식 일부가 [부록]에 제시되었다.

2. 채점기준 개발 절차

첫째, 학생들이 제출한 탐구활동보고서 중 일부를 활용하여 수준별 보고서에서 관찰되는 수행의 특징을 목록화하였다. 먼저, 75개의 탐구활동보고서를 간단한 채점기준을 활용하여 상, 중, 하의 3개 그룹으로 분류하였는데, 이 때, 다음의 3가지 채점기준이 활용되었다: ① 제시된 레시피를 따라 리코타 치즈를 만들었고, 리코타 치즈의 특징에 대한 서술이 개념학습에서 학습한 내용에 근거하였는가? ② 치즈의 특성을 변화시킬 수 있는 레시피 조건을 제시하였고, 그 이유가 타당한가? ③ 직접 만든 치즈에 대한 현상과 특징, 분석 결과를 타당하게 제시하였는가? 다음으로, 연구자 5인이 수준별로 분류된 보고서의 특징을 분석하였다. 일차적으로 5인의 연구자가 독립적으로 도출한 서술자(descriptor)는 총 107개였으나, 서로 중복되는 특징들을 통합하여 진술하고, 과학 탐구역량과는 관계가 없는 특징들은 삭제하여, 최종 30여 개의 특징으로 최종 요약하였다.

둘째, 과학 탐구과제 혹은 과학 탐구역량 관련 선행연구에서 제시하는 탐구과정 및 역량 요소를 검토하고, 앞서 도출된 탐구활동보고서의 수준별 특징들이 어떠한 과학탐구 과정 및 역량 요소들을 대표하는지를 매칭하였다. 이 때, 선행연구 분석에 근거하여, 과학탐구 과정 요소는 가설생성, 실험설계, 실험 수행 및 자료 해석, 결론 도출 및 평가, 의사소통의 5개로 구분하였으며, 도출된 특징들이 어떠한 과정 요소에 해당하는지 매칭하였다. 이 후 항목들 간의 상호보완성과 명료성을 검토하면서 2-3차례 수정·보완을 거쳐, 최종 25개 항목을 포함하는 체크리스트형 채점기준표를 완성하였다. 다만, Table 3에서 확인할 수 있듯이, 채점항목 C6, C12, C19는 각각 동일한 내용의 채점항목 2개를 통합한 것인데, 과제 지시에 따라 해당 채점요소의 포함 여부를 2회 확인할 필요가 있었던 항목들이다. 구체적으로, C6

은 지시문에 조건을 적어도 2개 이상 제시하도록 했기 때문에 그에 따라 동일한 내용을 2회 확인한 것이고, C12는 ‘결과물의 특성을 (미션 1에서 만든) 기본 치즈와 비교하여 기술하는가?’와 ‘서로 다른 조건 변화에 따른 결과물의 특성을 비교하여 기술하는가?’에서의 점수를 통합한 것이다. 또한, C19는 ‘조건을 변화시키는 과정에서 계량적 접근이 관찰되는가?’와 ‘조건을 변화시키는 과정 외에 계량적 접근이 관찰되는가?’의 두 개 채점항목을 통합한 것이다. 최종적으로 22개의 고유한 채점항목을 포함하는 채점기준표가 완성되었다.

셋째, 완성된 채점기준표의 채점항목들이 선행연구에서 도출된 3가지 역량 요소 즉, 과학적 지식, 논리분석적 사고, 탐구적 태도 중 어떤 역량 요소를 반영하는지 검토하였다. 개별 채점항목들이 어떤 과학탐구 과정 요소를 내용적으로 반영하였는지와 관계없이, 개별 채점항목에 점수를 얻기 위해서 피험자가 발휘해야 하는 능력의 유형에 초점을 맞추어 분류하고자 하였다. 일차적으로 연구자 5인이 독립적으로 각 채점항목이 어떤 역량 요소를 반영하고 있는지를 판단하여 판단의 일치도를 검토하였으며, 2인 이상의 판단의 불일치가 발생한 항목에 대해서는 역량 요소의 개념에 대한 논의를 거쳐 Table 3의 마지막 열에 제시된 바와 같이 최종 분류하였다.

넷째, 완성된 채점기준표를 활용하여 채점자 11인이 350명 학생들이 작성한 탐구활동보고서를 채점하였다. 채점자는 교수설계, 교육평가, 교육심리, 과학교육 전공 분야의 연구 경력이 있는 경력 10년 이상의 교육학 박사학위 소지자 3명과 박사과정생 8명이 포함되었다. 학생 25명(보고서 25개)씩 1개 군집으로 묶어 개별 채점자에게 2~3개 군집을 무선배정하되, 채점자 엄격성 모수 추정을 위하여 채점자 간 채점대상 군집이 일부 교차되도록 하였다. 보고서 1부당 2인의 채점자가 교차 채점하였으며, 최종적으로 채점자 1인당 총 50~75부의 보고서를 채점하였다. 30분 내외의 채점자 연수 시행 후 독립적으로 채점을 시행하였으며, 채점자의 요청에 따라 채점항목에 대한 세부적인 이해나 판단의 방식에 대한 논의가 2-3회 이루어졌다. 1차 채점 후, 다국면 Rasch 모형을 활용하여 채점자 적합도를 모니터링하였으며, 채점자 간 불일치가 크게 나타난 사례에 대해서 2차 채점을 시행하여 최종 데이터를 산출하였다.

Table 3. Scoring checklist by science-inquiry process and competence components

과학탐구 과정 요소		채점항목(0/1 혹은 0/1/2)		역량 요소
가설 생성	종속변인 규명	C1	해당 차시에서 학습한 과학적 개념을 활용하여 치즈의 특징을 분석하는가?	과학적 지식
		C2	치즈의 특징을 분석하기 위하여 기타 사전 지식(해당 차시 학습 이외)을 활용하는가?	과학적 지식
	독립변인 탐색	C3	치즈의 특징을 변화시킬 수 있을 것으로 제시되는 조건을 2개 이상 명료하게 제시하는가?	논리분석적 사고
		C4	각 조건의 변화가 치즈의 어떤 특징을 어떻게 변화시킬지를 예상(예측)하여 진술하는가?	논리분석적 사고
		C5	기타 자료를 참고하거나 스스로의 사고를 통해 고려할만한 조건을 추가적으로 탐색하였는가?	탐구적 태도
	가설 정당화	C6	각 조건의 변화가 왜 혹은 어떻게 치즈의 특징을 변화시키게 될지를 사전에 학습한 과학적 개념이나 원리를 활용하여 설명하는가? (0/1/2)	논리분석적 사고
		C7	각 조건의 변화가 왜 혹은 어떻게 치즈의 특징을 변화시키게 될지를 기타 과학적 개념이나 원리를 활용하여 설명하는가?	탐구적 태도
실험 설계 및 수행	독립변인 규명	C8	제시된(선정된) 조건이 앞에서(사전 학습자료 혹은 개별 추가 학습자료) 학습한 과학적 개념이나 원리와 관련이 있는가?	과학적 지식
		C9	실험을 위해 변화시킨 조건이 명확한가?	논리분석적 사고
	실험 수행	C10	선택한 조건을 변화시킬 수 있는 방법으로 치즈 제작 과정의 조건을 변화시켰는가?	논리분석적 사고
	혼재변인 통제	C11	자신이 선택한 조건 이외의 조건을 고려하여 통제하는가?	논리분석적 사고
자료 분석 및 해석	가설 검증	C12	가설을 바탕으로 결과를 분석하는가?: 조건 변화에 따른 결과물의 특성을 비교하여 기술하는가? (0/1/2)	논리분석적 사고
		C13	가설을 바탕으로 결과를 분석하는가?: 분석결과를 본인의 가설(예상했던 결과)과 비교하여 기술하는가?	논리분석적 사고
	자료 해석	C14	과학적 원리와 개념에 대한 사전지식을 활용하여 결과를 해석하는가?	과학적 지식
		C15	분석 결과를 이해(해석)하기 위해 추가 자료를 탐색하는가?	탐구적 태도
		평가	C16	분석 결과에 근거하여 특정 조건의 적절성이나 유용성을 평가하는가?
결론 도출 및 평가	학습과정 성찰	C17	실험을 통해 자신이 무엇을 배웠는지 기술하는가?	탐구적 태도
		C18	본인이 수행한 실험의 강점이나 보완할 점 등을 기술하거나 향후 탐구과제를 제시하는가?	탐구적 태도
		계량적 접근	C19	조건을 변화시키는 과정 등에서 계량적 접근이 관찰되는가? (0/1/2)
전반	문서기반 의사소통	C20	치즈의 특징(관찰 결과)에 대한 언어적 기술이 풍부한가?	의사소통
		C21	참고한 자료의 출처를 제시하는가?	의사소통
		C22	진술이 명료하고, 문단을 구조화하는 등 독자의 가독성을 고려하는가?	의사소통

3. 다국면 Rasch 모형 분석

개발된 채점기준표를 활용한 채점결과가 측정모형에 적합한지를 검토하기 위하여 다국면 Rasch모형을 활용하였다. 다국면 Rasch모형은 채점자의 엄격성이나 채점항목의 난이도, 과제의 유형 등 피험자 점수의 분산을 설명하는 다양한 조건(국면)들을 모형화하여, 조건들과는 독립적인 피험자의 숙달도 수준을 추정한다. 이 연구에서는 피험자와 채점항목, 그리고 채점자를 포함한 3개 국면을 고려하였으며, 구체적으로 총 350명의 피험자, 22개의 채점항목, 11명의 채점자의 채점결과 데이터를 분석하였다. 22개 채점항목 중 19개 채점항목은 0 혹은 1로 채점한 이분(dichotomous)변수로 기본적인 다국면 Rasch모형을 적용한 반면, 3개 채점항목(C6, C12, C19)은 0, 1, 혹은 2로 채점한 다분(polytomous)변수로 Andrich 평정척도모형(Andrich's rating scale model)을 적용하였다(Andrich & Marais, 2019; Eckes, 2015). 아래 식은 Andrich 평정척도모형으로, 수준의 피험자 n이 채점자 j로부터 채점항목 i에 k-1점 받을 확률 대비 k점을 받을 확률을 모형화한 것이다. 분석에는 FACET 프로그램(Linarce, 2019)이 활용되었다.

$$\text{Log}[P_{nik}/P_{nik-1}] = \theta_n - \beta_i - \gamma_j - \tau_k$$

where P_{nik} 피험자 n이 채점항목 i에서 채점자 j로부터 k점을 받을 확률

θ_n 피험자 n의 숙달도

β_i 채점항목 i의 난이도(총 22개 채점항목)

γ_j 채점자 j의 엄격도(총 11명의 채점자)

τ_k 채점항목 i에서 k-1점과 k점을 받을 확률이 동일한 지점 (threshold)의 난이도

IV. 연구결과

1. 다국면 Rasch모형 분석 결과

다국면 Rasch모형 적합 결과, 피험자, 채점자, 채점항목(기준)간 점수 분산의 52.9%가 측정모형에 의해 설명되는 것으로 나타나, 과학 탐구과제를 수행하는 데 필요한 잠재능력의 일차원성에 대한 가정을 지지하였다. 각 국면에서 separation 지수는 모두 1 이상으로 각 국면의 추정치들이 서로 변별되는 것으로 나타났으며, 신뢰도 또한 최소 .90이상으로 양호하였다(Table 4 참고).

피험자 국면에서는 $\chi^2=3045.4$, $df=349$ ($p<.001$)로, 피험자 350명의 숙달도 추정치 즉, 과제 수행 수준이 뚜렷하게 변별되는 것으로 나타났다. 피험자 숙달도 추정치는 평균 -1.28 logits이었으며, 극단치를 제외하고 최소 -5.72 logits부터 최대 3.12 logits까지 분포하였다.

Table 4. Summary of Many-faceted Rasch model fit

		피험자(n=350)	채점자(k=11)	채점항목(i=22)
Separation		3.02	7.28	15.78
Separation Reliability		.90	.96	1.00
Chi-square(df)		3045.4(349)***	314.5(10)***	5268.2(21)***
추정치 평균(표준편차)		-1.28(1.02)	.00(.40)	.00(2.22)
내적적합도 평균제곱(MS)	평균(표준편차)	1.02(0.32)	1.00(.11)	.99(.13)
	범위	(.38, 2.47)	(.78, 1.21)	(.79, 1.31)
외적적합도 평균제곱(MS)	평균(표준편차)	1.02(1.03)	1.04(.43)	1.09(.49)
	범위	(0.18, 9.00)	(.52, 1.88)	(.51, 2.87)

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

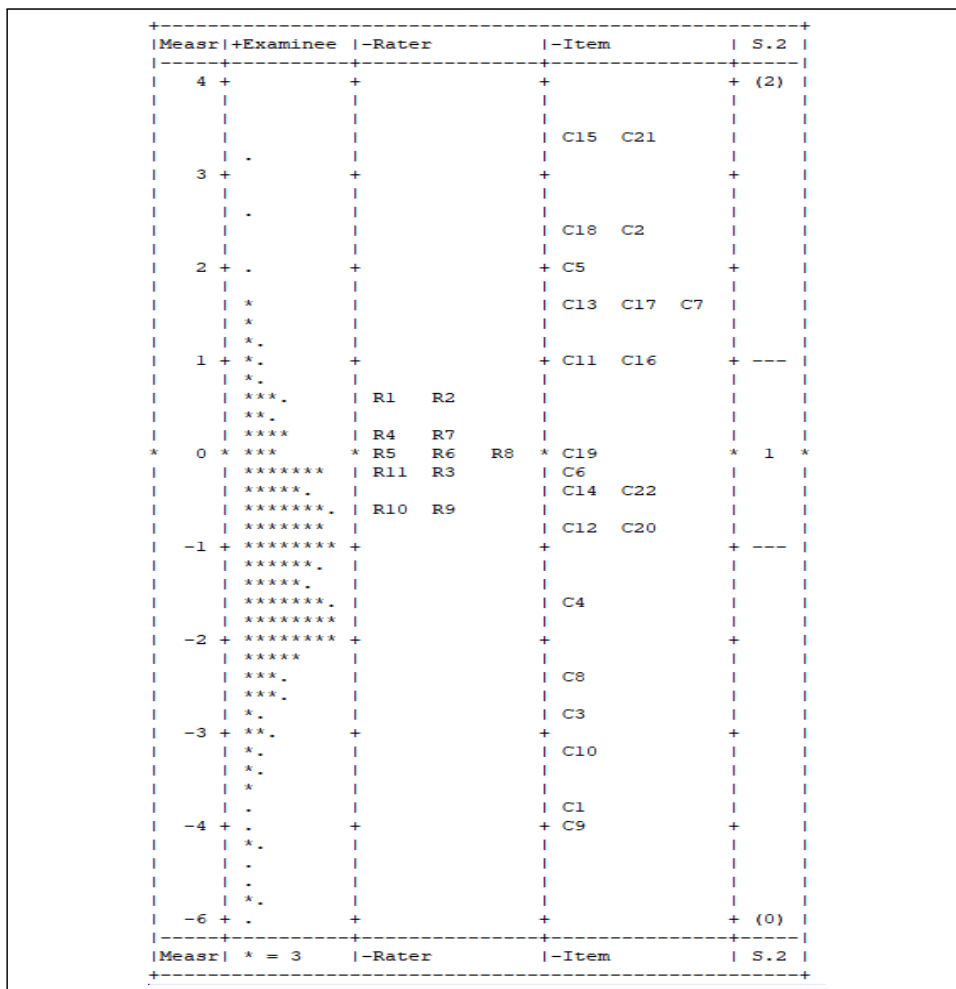


Figure 1. Item x person map(Wright map)

Figure 1의 문항-피험자 지도에서 확인할 수 있듯이, 전체적인 피험자들의 숙달도 분포는 채점항목들의 난이도와 비교하여 낮았다. 피험자 숙달도 추정치의 내적적합도 및 외적적합도 평균은 1.02로 1에 가까웠으며, 일부 적합도가 낮은 사례들이 있었으나, 90%이상의 사례가 내적적합도 평균제곱 기준 .56이상 1.62이하의 범위에 포함되어 있었다.

채점자 국면에서는 $\chi^2=314.5$, $df=10$ ($p<.001$)로, 채점자 11인의 엄격성 추정치에 유의한 개인차가 관찰되었다. 구체적으로 엄격성 추정치는 최소 -.63logits부터 최대 .61logits까지 분포하였으며, 채점자 5, 6, 8은 평균 수준의 엄격성을 보인 반면, 채점자 1과 2가 가장

엄격하고, 채점자 9와 10은 가장 관대한 것으로 나타났다. 채점자들의 내적적합도 평균은 1.00으로, 최소 .78에서 최대 1.21까지 분포하여 적합하였다.

채점항목 국면에서는 $\chi^2=5268.2$, $df=21$ ($p<.001$)로, 22개 채점항목 간 난이도 추정치가 뚜렷하게 변별되는 것으로 나타났다. 채점항목별 난이도는 최소 -3.99logits에서 최대 3.41 logits까지 고르게 분포하여, 다양한 수준의 피험자들을 적절하게 변별할 수 있었다. Figure 1과 Table 5에서 확인할 수 있듯이, 난이도가 가장 낮은 채점항목은 ‘실험을 위해 변화시킨 조건이 명확한가(C9, -3.99logit)’와 ‘해당 차

Table 5. Estimated item difficulty parameters and fit indices

과학탐구 과정 요소	항목번호	채점항목	난이도 (logits)	표준 오차	내적 적합도 MS	외적 적합도 MS
가설 생성	C1	해당 차시에서 학습한 과학적 개념을 활용하여 치즈의 특징을 분석하는가?	-3.86	0.13	1.23	1.31
	C2	치즈의 특징을 분석하기 위하여 기타 사전 지식(해당 차시 학습 이외)을 활용하는가?	2.41	0.18	1.00	1.18
	C3	치즈의 특징을 변화시킬 수 있을 것으로 제시되는 조건을 2개 이상 명료하게 제시하는가?	-2.90	0.11	1.12	1.09
	C4	각 조건의 변화가 치즈의 어떤 특징을 어떻게 변화시킬지를 예상(예측)하여 진술하는가?	-1.56	0.09	0.98	0.99
	C5	기타 자료를 참고하거나 스스로의 사고를 통해 고려할만한 조건을 추가적으로 탐색하였는가?	2.04	0.16	0.98	1.25
	C6	각 조건의 변화가 왜 혹은 어떻게 치즈의 특징을 변화시키게 될지를 사전에 학습한 과학적 개념이나 원리를 활용하여 설명하는가? (0/1/2)	-0.13	0.07	1.11	1.10
	C7	각 조건의 변화가 왜 혹은 어떻게 치즈의 특징을 변화시키게 될지를 기타 과학적 개념이나 원리를 활용하여 설명하는가?	1.57	0.14	0.84	0.51
실험 설계 및 수행	C8	제시된 조건이 앞에서 학습한 과학적 개념이나 원리와 관련이 있는가?	-2.49	0.1	0.96	1.01
	C9	실험을 위해 변화시킨 조건이 명확한가?	-3.99	0.14	0.79	0.67
	C10	선택한 조건을 변화시킬 수 있는 방법으로 치즈 제작 과정의 조건을 변화시켰는가?	-3.22	0.11	0.89	0.93
	C11	자신이 선택한 조건 이외의 조건을 고려하여 통제하는가?	1.03	0.12	1.31	2.02
자료 분석 및 해석	C12	가설을 바탕으로 결과를 분석하는가?: 조건 변화에 따른 결과물의 특성을 비교하여 기술하는가? (0/1/2)	-0.74	0.06	1.22	1.21
	C13	가설을 바탕으로 결과를 분석하는가?: 분석 결과를 본인의 가설(예상했던 결과)과 비교하여 기술하는가?	1.68	0.14	0.96	0.66
	C14	과학적 원리와 개념에 대한 사전지식을 활용하여 결과를 해석하는가?	-0.44	0.09	0.90	0.87
결론 도출 및 평가	C15	분석 결과를 이해(해석)하기 위해 추가 자료를 탐색하는가?	3.34	0.26	0.88	0.81
	C16	분석 결과에 근거하여 특정 조건의 적절성이나 유용성을 평가하는가?	1.00	0.12	1.02	1.10
	C17	실험을 통해 자신이 무엇을 배웠는지 기술하는가?	1.57	0.14	0.9	2.87
	C18	본인이 수행한 실험의 강점이나 보완할 점 등을 기술하거나 향후 탐구과제를 제시하는가?	2.45	0.18	0.94	0.83
전반	C19	조건을 변화시키는 과정 등에서 계량적 접근이 관찰되는가? (0/1/2)	0.02	0.07	0.91	0.97
	C20	관찰 결과에 대한 언어적 기술이 풍부한가?	-0.79	0.09	0.97	0.94
	C21	참고한 자료의 출처를 제시하는가?	3.41	0.27	0.94	0.91
	C22	진술이 명료하고, 문단을 구조화하는 등 독자의 가독성을 고려하는가?	-0.42	0.09	0.84	0.73

시에서 학습한 과학적 개념을 활용하여 치즈의 특성을 분석하는가(C1, -3.86logits)’였으며, 반대로 난이도가 가장 높은 항목은 분석 결과를 이해하기 위해 추가 자료를 탐색하는가(C15, 3.34logits)’, ‘참고한 자료의 출처를 제시하는가(C21, 3.41logits)’였다. 피험자 숙달도에 비해 채점항목들의 난이도가 전반적으로 높았으며, 특히, C2, C4, C15, C18, C21의 난이도는 2logits 이상으로 상위 3%의 학생들의 수행에서만 관찰가능하였다. 그러나 그 외17개 채점항목의 난이도는 다양하게 넓게 분포하여 학생들의 능력 수준을 고루 변별할 수 있는 기준들이었다. 내적적합도 평균은 .99로, 최소 .79에서 최대 1.31사이 에 분포하여 적합하였다.

2. 과학탐구 과정 요소별 × 수준별 학생 수행의 특성

Rasch모형 분석의 가장 큰 장점은 각 문항별 난이도와 피험자 숙달도를 직접 비교할 수 있다는 점이다. 과학탐구 과정 요소별 학생 수행의 특성이 어떻게 다른지를 구체적으로 살펴보기 위하여 과학탐구 과정 요소별로 채점항목의 난이도를 Table 6와 같이 분류하였다. 피

험자 숙달도 평균을 기준으로 -2logits 미만의 난이도에 해당하는 채점항목은 ‘기초 수준’, 피험자 숙달도 평균+1SD 범위(-2logits 이상, -0.26logits 미만)의 난이도에 해당하는 채점항목은 ‘평균 수준’, 피험자 숙달도 평균+1SD 이상 2logits 미만(-0.26logits 이상, 2logits 미만)의 난이도에 해당하는 채점항목은 ‘상 수준’, 2logits 이상의 난이도에 해당하는 채점항목은 ‘최상 수준’으로 구분하였다. 다만, 이 연구에서 ‘치즈는 왜 맛이 다를까?’의 탐구주제에 맞추어 진술된 채점항목들을 다른 다양한 탐구주제의 맥락에서 활용할 수 있도록 채점항목을 일반화하여 일부 재진술하였으며, 재진술된 부분은 밑줄로 표시하였다. 관찰된 내용을 정리하면 다음과 같다.

첫째, 기초 수준에 해당하는 채점기준들은 피험자의 약 98%가 도달가능한 수준으로, 사전 학습자료에서 학습한 내용을 가설생성에 활용하는지(C1, C8), 실제 과제의 지시문에서 피험자에게 요구한 사항들이 답안에 명시되어 있는지(C3, C9), 독립변인과 실험 조건의 관계가 적절한지(C10)를 포함하였다. 기초 수준의 피험자들은 과제 지시사항에 적절하게 반응하고, 사전학습 자료의 내용을 독립변인 선정에 활용할 수 있는 수준으로 이해될 수 있다. 흥미롭게도, ‘자료

Table 6. Performance characteristics by science inquiry process components × performance levels (the underlined were restated for potential generalization)

	기초(-2logits 미만)	중(-2logit 이상 평균+1SD 미만)	상(평균+1SD 이상 2logits 미만)	최상(2logits 이상)
가설 생성	[C1] 사전 학습자료에서 학습한 과학적 개념을 활용하여 대상의 특징/현상을 분석하는가? (과학적 지식) [C3] 대상의 특징/현상을 변화시킬 수 있을 것으로 제시되는 조건을 명료하게 제시하는가? (논리분석적 사고)	[C4] 각 조건의 변화가 대상의 특징/현상을 어떻게 변화시킬지를 예상(예측)하여 진술하는가? (논리분석적 사고)	[C6] 각 조건의 변화가 왜 혹은 어떻게 대상의 특징/현상을 변화시키게 될지 설명하기 위하여 사전학습 자료를 활용하는가? (논리분석적 사고) [C7] 각 조건의 변화가 왜 혹은 어떻게 대상의 특징/현상을 변화시키게 될지 설명하기 위하여 기타 자료를 활용하는가? (탐구적 태도)	[C2] 대상의 특징/현상을 분석하기 위하여 기타 사전 지식(해당 차시 학습 이외)을 활용하는가? (과학적 지식) [C5] 기타 자료를 참고하거나 스스로의 사고를 통해) 고려할만한 조건을 추가적으로 탐색하였는가? (탐구적 태도)
실험설계 및 수행	[C8] 제시된(선정된) 조건이 앞서 학습한 과학적 개념이나 원리와 관련이 있는가? (과학적 지식) [C9] 실험을 위해 변화시킨 조건이 명확한가? (논리분석적 사고) [C10] 선택한 조건을 변화시킬 수 있는 방법으로 실험 조건을 변화시켰는가? (논리분석적 사고)		[C11] 자신이 선택한 조건 이외의 조건을 고려하여 통제하는가? (논리분석적 사고)	
자료 분석 및 해석		[C12] 조건 변화에 따른 결과물의 특성을 비교하여 기술하는가? (논리분석적 사고) [C14] 과학적 원리와 개념에 대한 사전지식을 활용하여 결과를 해석하는가? (과학적 지식)	[C13] 분석결과를 본인의 가설(예상했던 결과)과 비교하여 기술하는가? (논리분석적 사고)	[C15] 분석 결과를 이해(해석)하기 위해 추가자료를 탐색하는가? (탐구적 태도)
결론 도출 및 평가			[C16] 분석 결과에 근거하여 특정 조건의 적절성이나 유용성을 평가하는가? (탐구적 태도) [C17] 실험을 통해 자신이 무엇을 배웠는지 기술하는가? (탐구적 태도)	[C18] 본인이 수행한 실험의 강점이나 보완할 점 등을 기술하거나 향후 탐구과제를 제시하는가? (탐구적 태도)
전반 -의사소통		[C20] 대상의 특징/현상에 대한 언어적 기술이 풍부한가? (의사소통) [C22] 진술이 명료하고, 문단을 구조화하는 등 독자의 가독성을 고려하는가? (의사소통)	[C19] 조건을 변화시키는 과정 등에서 계량적 접근이 관찰되는가? (의사소통)	[C21] 참고한 자료의 출처를 제시하는가? (의사소통)

분석 및 해석', '결론 도출 및 평가'와 '전반적인 문서기반 의사소통'과 관련된 기준들은 기초 수준의 학습자에게서 관찰되기 어려운 것으로 나타났다.

둘째, 평균 수준에 해당하는 채점기준들은 피험자의 약 80%가 도달 가능한 수준으로, 독립변인에 따른 종속변인의 변화를 예상하고(C4), 실험 결과를 구체적이고 명료하게 기술하고(C12, C20, C22), 과학적 원리나 개념을 활용하여 해석(C14)할 수 있는 것이다. 기초 수준에서는 사전지식을 실험을 위한 독립변인 선정에만 활용하였다면, 평균 수준에서는 사전지식을 실험의 결과를 해석하는 데도 활용하였다. 한편, 탐구보고서 진술의 명료성이나 언어적 유창성이 평균 수준의 학생들에게서 나타나는 특징이라고 볼 수 있다.

셋째, 상 수준에 해당하는 채점기준들은 피험자의 약 15%가 도달 가능한 수준으로, 가설 생성에서 독립변인에 따른 종속변인의 변화를 예상할 뿐만 아니라 그것이 왜 혹은 어떻게 일어나는지를 설명하여 가설을 정당화하고(C6), 이를 위해 다른 자료들을 탐색하기도 하였다(C7). 실험 설계시에는 자신이 선택한 조건 이외의 조건을 고려하여 통제하기도 하였으며(C11), 분석 결과를 본인의 가설을 검증하는 방식으로 해석할 수 있었다(C13). 흥미롭게도 상 수준부터 '결론 도출 및 평가'와 관련된 채점항목들이 관찰되는데, 구체적으로 분석 결과에 근거하여 특정 조건의 유용성을 평가하거나(C16), 실험을 통해 본인이 무엇을 배웠는지를 기술하였다(C17). 실험 과정에 계량적인

접근이 관찰되는 것 또한(C19) 두드러진 특징이다.

넷째, 최상 수준에 해당하는 채점기준들은 피험자의 5%만 도달 가능한 수준으로, 독립변인을 탐색할 때(C2, C5) 뿐만 아니라 분석 결과를 해석하기 위해서(C15) 다른 자료를 다양하게 탐색하고, 탐색하여 참고한 자료의 출처를 명확히 하는 수행(C21)이 관찰되었다. 한편, 본인이 수행한 실험을 스스로 평가하고 향후 실험 방향을 제안하기도 하였다(C18).

한편, 채점기준의 난이도 간 편차가 가장 크게 나타난 '가설 생성' 영역에서 학생들 간 수행의 편차가 실질적으로 어떻게 관찰되는지를 확인하기 위하여 '탐구활동보고서의 미션 2(Table 2 참조)'에 대한 응답 사례를 Table 7에 제시하였다. 기초~중 수준에 해당하는 응답의 경우, 치즈의 특징을 변화시킬 수 있을 것으로 제시되는 조건을 명확하게 제시하였으며(C3), 각 조건의 변화가 어떻게 치즈의 특징을 변화시킬 것인지를 예상하고 있으나(C4), 그 변화가 왜 혹은 어떻게 일어나는지에 대해서는 설명하지 않았다. 한편, 중~상 수준의 응답은 사전학습 자료에 제시된 내용(치즈 생성 과정이 단백질 응고 과정임을 설명)을 활용하여 해당 조건이 왜 그리고 어떻게 치즈의 특징을 변화시킬지에 대한 설명을 포함하였다(C6). 상 수준의 응답에서는 한 단계 더 나아가, 사전학습 자료에서 제시되지 않았던 조건들을 추가적인 참고자료를 활용하여 탐색한 것을 확인할 수 있었다(C5).

Table 7. Samples of student actual responses by performance levels

전반적인 수행 수준	<미션 2. 치즈의 특성을 변화시키는 조건 탐색>에 대한 학생 응답 사례
기초~중	<ol style="list-style-type: none"> 1. 끓이는 시간 <ul style="list-style-type: none"> -많이 끓이면 촉감은 부드러워질 것 같다. -많이 끓이면 맛의 세기는 약해질 것 같다. -많이 끓이면 강도는 부드러워질 것 같다. -맛도 달라질 것 같다. 2. 물기를 짜주는 정도 <ul style="list-style-type: none"> -많이 짜면 촉감은 딱딱해질 것 같다. -많이 짜면 맛의 세기는 진해질 것 같다. -많이 짜면 강도는 딱딱해질 것 같다. -맛도 달라질 것 같다.
중~상	<ol style="list-style-type: none"> 1. 물기를 짜주는 정도: 반죽을 하고 난 뒤, 면포에 넣어서, 더 많이 짜면, 물기가 없어져서, 발효된 치즈의 촉감이 딱딱할 것 같고, 반죽을 면포에 넣어서 덜 짜주면, 물기가 많아 촉촉하거나 물렁물렁하고, 또한 부드러기도 할 것이다. 2. 레몬즙의 양: 레몬즙이 치즈를 만드는데 사용되는 이유는 치즈반죽을 응고시키기 위해서이다. 따라서, 레몬즙이 더 많이 들어간 치즈는 반죽이 더 많이 응고되고, 그로 인해 냉장보관 후, 많은 양의 치즈로 변할 것이다. 또, 레몬즙이 더 적게 들어간 치즈는 반죽이 덜 응고되어, 냉장보관 후, 적은 양의 치즈로 변할 것이다
상~최상	<ol style="list-style-type: none"> 1. 소금과 레몬즙의 양을 다르게 한다: 소금을 더 많이 넣으면 짭짤한 맛이 강해지고 덜 넣으면 짠맛이 약해질 것이다. 같은 원리로 레몬즙의 양에 따라 치즈의 시큼한 맛의 강도가 달라질 것이다. 또한, 레몬즙의 산성은 단백질의 응고와 관련이 있으므로, 레몬즙을 많이 넣으면 치즈가 더 단단해질 것 같다. 레몬즙 말고도 식초나 유자초 같이 산성을 가진 다른 재료를 넣어줘도 치즈의 맛과 향이 달라질 것으로 예상된다. 2. 생크림의 양을 다르게 한다: 생크림은 고소하고도 약간 느끼한 맛이 있는데, 생크림이 가진 지방 때문인 것 같다. 생크림을 더 많이 넣으면 치즈의 맛도 더 고소하고 풍부한 맛이 나고 치즈의 질감도 부드러워질 것이고 더 적게 넣으면 덜 고소하고 질감도 푸석푸석할 것 같다. 3. 재료들을 끓이는 시간과 불의 세기를 다르게 한다: 불을 더 세게 하거나 끓이는 시간이 길어지면 재료 속의 수분이 증발하여 치즈의 강도를 더 단단하게 하고 질감을 푸석푸석하게 할 것 같다. 4. 냉장고에 넣기 전 물기를 짜내는 정도를 다르게 한다: 물기를 많이 짜내면 치즈 속의 수분이 적어지므로 치즈의 질감이 더 거칠고 건조해질 것 같고 물기를 적게 짜내면 더 부드럽고 촉촉한 치즈가 될 것 같다.
상~최상	<ol style="list-style-type: none"> 1. 끓이는 시간을 바꾸어서 치즈를 다르게 제조할 수도 있다. 만약 끓이는 시간을 30분이 아닌 1시간으로 한다면 결과는 다르게 나올 가능성이 높다. 김치를 더 오랫동안 발효시키면 그 맛이 다르게 나오는 것과 같이 치즈도 더 오랫동안 발효시키면 그 맛이 변화할 수도 있을 것 같다. 2. 레몬 대신 식초를 넣어서 치즈를 만들 수도 있다. 레몬은 ph 2.2 이고, 식초는 ph 4 인데, 산도가 높을수록 응고가 더 잘 되기 때문에 이론적으로 식초를 넣으면 응고가 이루어지는 속도가 더 빠를 것으로 예상된다. 3. 숙성 시간을 바꾸어도 치즈의 맛이 변화할 수 있다. 숙성이 되는 시간에 따라 맛의 진하거나 종류가 더 진해지거나 변화할 수 있다. 위에 있는 사진을 보면 발효숙성기간(Ripening Period)에 따라 효모, 곰팡이, 그리고 생균의 수가 변화하는 것을 볼 수 있다(발효숙성 기간에 따른 효모, 곰팡이, 생균의 수 변화에 대한 그래프 자료와 출처를 제시함). 이렇게 때문에 숙성 시간이 달라지면 효모, 곰팡이, 세균의 수 등이 변화해 만들어지는 리코타 치즈의 맛을 변화시킬 수 있다. 4. 우유와 생크림의 비율도 치즈의 맛에 영향을 줄 수가 있다. 우유와 생크림의 비율을 기본 레시피대로 2:1로 넣은 것과, 우유와 생크림의 비율을 1:1로 해서 넣은 것을 비교해보는 경우에는 어떤 것은 우유의 맛이 조금 강하고, 어떤 것은 맛이 우유와 생크림이 잘 조화롭게 이루어진 맛으로 맛이 나오게 나올 수 있다. 이렇게 우유와 생크림의 비율도 치즈의 맛을 변화시킬 수 있다

V. 논의 및 결론

이 연구는 초등학교 과학 탐구활동보고서에서 관찰되는 수행의 특징을 분석하고, 그 특징을 과학탐구 과정 요소와 역량 요소와 연계하여 과학탐구과제 채점기준을 개발하기 위하여 수행되었다. 이를 위하여 초등학교 5학년 학생 75명의 비구조화된 과학 탐구활동보고서의 특징을 분석하여 채점기준표를 개발하였으며, 이를 활용하여 채점자 11인이 350여 개 과제를 교차채점하여 데이터를 생성하였다. 주요 연구결과와 논의를 정리하면 다음과 같다.

첫째, 개발된 체크리스트형 채점기준표를 활용하여 생성한 채점데이터는 피험자, 채점자, 채점항목의 3개 국면을 고려한 다국면 Rasch 모형에 적합하였으며, 22개 채점항목들의 적합도가 양호하였다. 구체적으로, 난이도가 -2logits 이하로 낮은 항목, 난이도가 +2logits 이하로 높은 항목이 각각 5개였으며, 이를 제외한 12개 항목이 -2logits에서 2logits 사이에 적절하게 분포하여 서로 다른 수행수준을 변별하는데 적절하게 기능하는 것으로 나타났다.

둘째, 학생들의 숙달도 평균은 -1.28logits으로 채점항목의 난이도와 채점자의 엄격성과 비교하여 평균적으로 낮았다. 특히, 주어진 자료 이외의 추가자료를 탐색하거나(C5, C7, C15, C21), 자신의 실험과정을 평가하고 성찰하는(C17, C18) 특성은 숙달도가 매우 높은 소수의 학생들에게서 관찰되었다. 반면, 과제 안내문에 명시된 요구사항에 대한 응답(C3, C9)은 대다수의 학생들에게서 관찰되었다.

셋째, 기초 숙달 수준의 학생들은 ‘가설 생성’과 ‘실험 설계 및 수행’과 관련된 수행 중 일부는 상대적으로 쉽게 달성할 수 있는 반면, ‘자료 분석 및 해석’, ‘결론 도출 및 평가’단계와 ‘문서기반 의사소통’ 관련 전반적인 특징 측면에서는 적절한 수행 수준에 도달하기 어려운 것으로 나타났다. ‘가설 생성’과 ‘실험 설계 및 수행’ 단계의 채점기준들은 사전 학습자료에서 학습한 내용을 바탕으로 치즈의 특징을 변화시킬 조건을 찾고, 그 조건을 반영하여 독립변인을 조작하도록 요구하는데, 이 활동은 사전 학습자료에 등장하는 개념들을 적절히 활용하면 충분히 수행할 수 있다. 반면, ‘자료 분석 및 해석’단계는 실험 결과를 본인의 가설과 연계하여 분석·해석하는 것으로, 사전

학습자료에서 제시된 혹은 기타 자료를 통해 학습한 원리를 좀 더 적극적으로 이해하고 활용할 것을 요구하기 때문이다.

넷째, 과학탐구 과정 요소 중 결론 도출 및 평가 단계를 제외한 모든 과정 요소와 관련된 수행은 학생들의 숙달도 수준에 따라 다르게 나타났으며, 특히 숙달도 수준이 높아질수록 탐구적 태도가 더 자주 관찰되었다. 구체적으로, 가설 생성의 경우, ‘기초 수준’에서는 독립변인 즉, 종속변인을 변화시킬 것으로 기대되는 조건을 명확하게 제시하였지만, 선정된 독립변인이 종속변인을 어떻게 변화시킬 것인지에 대한 예측, 즉 가설을 진술하는 것은 ‘중 수준’에서 관찰되었다. 나아가 ‘상 수준’에서는 독립변인이 종속변인에 일으킬 변화에 대한 자신의 예측을 사전 학습자료 혹은 그 외 자신이 탐색한 자료를 활용하여 설명함으로써 가설을 정당화하는 수행을 보일 수 있었다. 한편, ‘실험 설계 및 수행’에서 독립변인의 조작은 ‘기초 수준’에서 도달 가능하였지만, 통제변인을 고려하는 수행은 ‘상 수준’ 이상에서 도달 가능하였다.

또한, ‘자료 분석 및 해석’에서는, ‘중 수준’에서는 실험 결과로 대상의 특징이 어떻게 변화했는지를 기술하고, 그 변화를 사전지식을 활용하여 해석할 수 있었던 반면, 그 결과를 실험 설계 당시 본인이 세웠던 가설에 비추어서 검토하고 해석하는 것은 ‘상 수준’에서 관찰되었다. 한편, ‘최상 수준’에서는 실험 결과에 대한 이해를 확장하기 위하여 추가자료를 탐색하는 행동이 관찰되었다. 문서기반 의사소통 측면에서, 보고서 진술의 명료성과 유창성은 ‘중 수준’ 학생들에게서도 관찰될 수 있는 특징이었으나, 계량적 접근을 활용하거나, 자신이 추가적으로 탐색한 자료의 출처를 기록하는 수행은 더 높은 수준에서 달성 가능한 것으로 나타났다. 일반적으로 선행연구에서는 과학적 태도 혹은 과학자적 특성은 적극성, 객관성, 비판성 등의 정의적 영역으로 구분하고, 이러한 태도가 과학탐구능력과 상관이 있다고 보고하였다(Laforgia, 1988; Yang *et al.*, 2003). 그러나 이 연구에서는 탐구적 태도를 정의적 영역으로 수행과 구분하지 않고 과학탐구 수행능력의 지표로 간주하여, 그것이 과제 수행과정에서 어떻게 관찰되는지를 구체적으로 분석하였다. 그 결과, 탐구적 태도는 전반적으로 과학탐구능력이 높은 학생들에게서 관찰되는, 즉 난이도가 높은 수행지표이자, 과제 수행과정 전반에서 고성취 학생과 일반성취 학생을 변별하는 역량인 것으로 나타났다. 이는 고성취 학생과 일반성취 학생 간 탐구적 태도가 평균적으로 유의하게 차이를 보이지만, 탐구적 태도와 과학탐구능력 간의 상관이 고성취 학생들에게서는 유의하게 나타나지 않았다는 선행연구 결과(Yang *et al.*, 2003)와 일관된 결과이다.

다섯째, ‘결론 도출 및 평가’의 채점기준들은 ‘상 수준’ 이상에서 관찰가능하였다. 이러한 결과는 초등 영재학생들의 과학 탐구 수행의 특징을 분석한 Song, Kil, & Shim 외(2015)의 연구에서도 보고된 바 있으며, 중학생의 과학 창의적 문제 해결 능력과 관련한 Park & Kang(2012)의 연구에서도 해결책 도출능력이 현저히 낮게 나타난 결과와도 유사하다. 결론 도출은 탐구 과정에서 얻어진 결과를 종합하고 해석하여 설정한 가설 또는 실험의 옳고 그름을 판단하는 활동이며, 평가는 실험 및 탐구 과정에서 적절한 전략을 사용하였는지 점검하고 성찰하는 활동이다(Yang, 2006). 이렇게 탐구 과정 및 결과를 판단, 점검, 성찰하기 위해서는 학습자의 메타인지, 즉 자신의 인지 활동을 모니터링할 수 있는 능력(Flavell, 1979)이 요구된다. 메타인지는 모든 영역에서 학습을 효율적으로 수행하는데 필수적인 기능으로

강조되며, 과학탐구 활동 역시 전반적으로 메타인지를 요구하는 활동이다(Kuhn, 2000; White, Frederiksen, & Collins, 2009; Shim, 1994). 예컨대, Kim & Jeong(2015)은 메타인지가 고등학생의 과학탐구 능력 및 과학 학업성취도에 직접적으로 영향을 미친다는 결과를 보고한 바 있다. 본 연구에서 개발한 채점 요소 중 특히 ‘결론 도출 및 평가’는 판단, 점검, 성찰과 같은 높은 메타인지를 요구하는 활동이기 때문에 주로 ‘상 수준’의 학생들에게서 이 요소들이 관찰된 것으로 보인다. 이는 메타인지 전략의 활용이 일반적으로 성취도가 높은 학생들에게서 나타나는 특성이 있다는 연구결과로 뒷받침할 수 있다(Zimmerman & Schunk, 2001).

이 연구는 초등학생의 탐구과제 수행 결과를 분석하여 보고서에 나타난 탐구수행의 특징을 분석하고 선행연구를 바탕으로 탐구과정 요소를 도출, 각 요소의 채점 기준을 체크리스트 형태로 개발하였다. 또한 개발된 채점 기준을 실제 채점에 적용하고 채점 데이터를 분석하여 통계적으로 적합성을 검증하고자 하였다. 이러한 접근은 기존 연구들이 이론에 기반하여 평가 요소 및 채점 기준을 개념적으로 도출하여 구분한 것과 달리(Lee & Jeong, 2013; Choi, Lee, & Chae, 2017), 실제 채점 데이터를 기반으로 각 채점 기준의 적합도를 확인한 점에서 의미를 가진다. 또한 각 채점 문항의 난이도를 분류하여 학습자들이 각 문항에 해당하는 수행에 어느 정도 도달할 수 있는지 파악하였다. 이러한 결과는 과학탐구 과정 요소별로 학생 수행의 특성이 어떻게 다른지 확인하여 개별 학생을 진단하고 피드백하는데 활용될 수 있다. 또한 이를 통해 전반적인 학생의 수행 수준을 파악하고 과학탐구 수업을 개선하는데 도움이 될 수 있을 것이다.

과학 탐구과제와 과학적 탐구역량과의 관련성에 대해, 본 연구는 과학 탐구과제가 선행연구에서 관련 있는 것으로 밝혀온 논리분석적 사고력, 의사소통능력, 과학적 태도뿐만 아니라 과학적 지식 등의 역량 요소로 이루어져 있음을 밝히고(Lee & Jeong, 2013; Son, 2006), 이들 요소 중 과학적 지식, 논리분석적 사고력, 의사소통능력 등은 과학탐구과제 전반에 걸쳐 필요한 능력인 것임을 확인하였다. 이는 과학탐구과제의 수행을 통해 이들 역량을 신장시킬 수 있는 가능성이 높다는 것을 의미한다. 특히, 이 연구에서는 그동안 과학 탐구역량의 측정에서 상대적으로 소홀하였던 과학적 지식을 포함함으로써 과학적 지식의 역할을 중요성을 확인할 수 있었다. Park & Kang(2012)의 연구에서 과학적 지식과 과학적 방법이 과학적 탐구역량에서 필요하다고 언급한 바와 같이 이 연구에서 과학적 지식은 하 수준 학생과 상 수준 학생에서 모두 나타난 것으로 보인다. 다만 주어진 탐구과제의 범위를 확장하여 추가적인 과학적 지식을 활용하거나, 새로운 자료를 탐색하는 등 보다 능동적, 적극적으로 과학적 지식을 탐구하고 활용하는 역량은 상 수준 학생들에게서만 나타난 특징이었다. 과학적 지식은 과학 탐구 과정에서 기본적으로 요구되는 역량이지만, 궁극적으로는 관련 개념을 탐색하여 지식을 확장하고 개념들을 연결하여 탐구하고자 하는 역량이 우수한 수행을 보이는 학생들에게서 핵심적으로 나타나는 역량임을 보여주는 결과이다. 이러한 수행의 특성이 과학 탐구과제의 평가도구에 반영되어 학생들이 탐구과제 수행에 있어 핵심적인 요소로 인식한다면, 학생들이 보다 높은 수준의 과학탐구 역량을 개발하는데 도움이 될 수 있을 것이다.

이 연구에서 개발된 체크리스트 형태의 채점 기준은 여러 연구자들이 언급했듯이(Popham, 1995; Ruiz-Primo, Shavelson, 1996; Schafer,

1997)이 탐구과정에서 학습자가 수행하는 핵심적인 행동을 명확하게 포함하고 있으며, 이에 대해 채점자는 수행 여부만을 고려하여 채점할 수 있는 장점이 있다. 이는 과학 탐구과제의 채점에서 그간 채점자들이 호소하던 평가의 모호성을 줄이고 채점자의 주관적 해석을 최소화하여 보다 객관적인 평가가 가능하도록 할 수 있을 것이다. 향후 이 연구에서 개발된 채점 기준에 대한 평가자의 인식을 추가적으로 살펴보고 기존 채점 도구의 한계를 보완할 수 있는지 확인해 볼 필요가 있다.

이 연구에서는 특정 탐구과제를 대상으로 채점 기준을 도출, 적용하고, 이를 일반화된 채점항목으로 제시하였는데, 이에 본 채점 기준을 다양한 탐구과제에 적용할 수 있는지 확인해 볼 필요가 있다. 이에 후속연구에서는 본 채점 기준의 일반화 가능성에 대해 보다 면밀히 검토해 보고자 한다. 또한, 본 채점기준에서 다루지 않았지만 과학과 핵심역량에서 중요하게 다루어질 수 있는 다른 요소들의 채점 기준 개발에 대한 연구도 확장적으로 이루어지기를 기대해본다. 마지막으로, 이 연구에서 온라인 수업에 참여한 학생들은 비교적 과학에 흥미가 높아 해당 프로그램에 자발적으로 참여하여 과제를 수행한 학생들이었다는 점에서 이 연구의 채점 기준을 보다 다양한 학생을 대상으로 확대하여 적용해볼 필요가 있다.

국문요약

이 연구는 초등학교 과학 탐구과제에서 관찰되는 수행의 특징을 분석하여 채점기준을 개발하기 위하여 수행되었다. 이를 위하여 초등학교 5학년 학생 75명의 과학 탐구활동보고서의 특징을 분석하고, 그 특징을 과학탐구 과정 요소와 역량 요소에 대한 이론 및 개념과 연계하여 체크리스트형 채점기준을 개발하였다. 개발된 채점기준을 활용하여 채점자 11인이 350여 개 과제를 교차채점하여 데이터를 생성하였으며, 그 결과는 다음과 같다. 첫째, 채점데이터는 피험자, 채점자, 채점항목의 3개 국면을 고려한 다국면 Rasch모형에 적합하였으며, 22개 채점항목들의 적합도가 양호하였다. 둘째, 낮은 수행수준에서 관찰가능한 특성은 과제에서 명시적으로 요구하는 질문에 답하거나, 사전에 학습한 과학적 개념을 활용하여 대상 혹은 현상을 관찰하는 것인 반면, 높은 수행수준에서는 주어진 자료 이외의 추가자료를 탐색하거나 자신의 실험과정을 성찰하는 것이었다. 학생 수행 특성 분석에 근거한 채점기준표의 유용성과 활용을 위한 향후 과제가 논의되었다.

주제어 : 과학 탐구과제, 과학탐구 과정 요소, 과학탐구 역량 요소, 다국면 Rasch모형, 채점기준

References

- Andrich, D. & Marais, I. (2019). A course in Rasch measurement theory: measuring in the educational, social, and health sciences. Springer: Singapore.
- Arnold, J. C., Boone, W. J., Kremer, K., & Mayer, J. (2018). Assessment of competencies in scientific inquiry through the application of Rasch measurement Techniques. *Education Sciences*, 8, 184.
- Baek J., Byun T., Lee D., & Shim H. (2020). An investigation on the assessment tool and status of assessment in the 'Scientific Inquiry Experiment' of the 2015 revised curriculum. *Journal of The Korean Association For Science Education*, 40(5), 515-529.
- Cho M., Mun K., & Kim S. (2010). The development and application of evaluating standards for creative problem solving items. *The Journal of Curriculum and Evaluation*, 13(2), 309-333.
- Choi K., Lee S., & Chae Y. (2017). Development of evaluation criteria for online problem-based science learning. *Journal of the Korean Association for Science Education*, 37(5), 879-889.
- Eckes, T. (2015). Introduction to many-facet Rasch measurement: analyzing and evaluating rater-mediated assessments. Peter Lang Edition.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911.
- Ha, M., Park, H., Kim, Y., Kang, N., Oh, P. S., Kim, M., Min, J., ... & Son, M. (2018). Developing and applying the questionnaire to measure science core competencies based on the 2015 Revised National Science Curriculum. *Journal of the Korean Association for Science Education*, 38(4), 495-504.
- Hwang J., Lee h., & Kwak D. (2010). An assessment of high school students' performance on science process skills in biology. *Journal of Research in Curriculum Instruction*, 14(1), 67-84.
- Kim H. (1993). Analysis on logical thinking levels and science process skills of high school student. *Journal of The Korean Association for Science Education*, 14(4), 424-431.
- Kim, S. & Chung, Y. (2015). Structural relationships among the epistemological beliefs, metacognition, science inquiry skills, and science achievement of high school students. *Journal of the Korean Association for Science Education*, 35(6), 931-938.
- Kim Y., & Kim S. (2009). A meta-analysis on logical thinking ability of Korean middle-school students: meta-analysis of the researches between 1980 and 2000. *Journal of The Korean Association for Science Education*, 29(4), 437-449.
- 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.
- Ko, E., & Choi A. (2019). Analysis of achievement standards, activities, and assessment items in Integrated science, chemistry I, chemistry II textbooks on science core competency: focusing on acid · base · neutralization and oxidation · reduction. *Journal of the Korean Chemical Society*, 63(6), 486-504.
- Koh E., & Jeong D. (2014). Study on Korean science teachers' perception in accordance with the trends of core competencies in science education worldwide. *Journal of the Korean Association for Science Education*, 34(6), 535-547.
- Kuhn, D. (2000). Metacognitive development. *Psychological Science*, 9(5), 178-181.
- Laforgia, J. (1988). The affective domain related to science education and its evaluation. *Science Education*, 72(4), 407-421.
- Lee, B. (2005). Analysis of inquiry standards in Foreign national science curricula. *Journal of the Korean Association for Science Education*, 25(7), 873-884.
- Lee H., Lee Y., & Lee H. (2017). Development and application of a rubric for assessing scientific inquiry process. *Secondary Education Research*, 65(1), 145-171.
- Lee J., & Jeong E. (2013). Development of an evaluation tool for assessing scientific thinking ability using science writing. *Teacher Education Research*, 52(3), 575-588.
- Lee K. (2010). Development of test of biology inquiry skills for middle and high school students (Unpublished master's thesis). Graduate School of Korea National University of Education, Chungbuk.
- Linacre, J. M. (2019). Facets computer program for many-facet Rasch measurement, version 3.81.2. Oregon: Winsteps.com.
- Lou, Y., Blanchard, P. & Kennedy, E. (2015). Development and validation of a science inquiry skills assessment. *Journal of Geoscience Education*, 63(1), 73-85.
- Ministry of Education (2015). National Science Curriculum. No. 2015-74. Sejong: Ministry of Education.
- Park I., & Kang S. (2012). The development of assessment tools to measure scientific creative problem solving ability for middle school students. *Journal of The Korean Association for Science Education*, 32(2), 210-235.
- Park, S., Shin, Y., Lee, Y., Min, J., & Kim, J. H. (2021). A research trend analysis on students' core competencies. *Asian Journal of Education*, 22(4), 769-799.
- Popham, W. J. (1995). Classroom Assessment: What Teachers Need to Know. Boston: Allyn & Bacon.
- Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Rhetoric and reality in science assessment: An update. *Journal of Research in Science Teaching*, 33(10), 1045-1064.
- Schafer, W. D. (1997) Classroom assessment. In G. D. Phye (ed.), Handbook

- of Academic Learning: Construction of Knowledge. San Diego, CA: Academic Press, 513-547.
- Shim, J. (1994). The exploration and effectiveness analysis of metacognitive instruction model. Doctoral dissertation, Korea University.
- Shim B., & Yoo M. (2020). An analysis of science core competencies reflected in the performance assessment of 2015 revised curriculum integrated science and scientific inquiry experiments. *School Science Journal*, 14(4), 481-500.
- Son J. (2006). A science writing teaching method based on scientific thinking for improving scientific essay writing ability. *The Journal of Curriculum & Evaluation*, 9(2), 333-355.
- Song K., Lee H., & Lim C. (2004). Development of a test of science inquiry skills for elementary school fifth and sixth graders. *Journal of the Korean Association for Science Education*, 24(6), 1245-1255.
- Song S., & Shim K. (2018). Analysis of science key competencies in inquiry activity of integrated science textbooks for high school students. *Biology Education*, 46(2), 222-236.
- Song S., Kil J., & Shim K. (2015). A case study on the evaluation of scientific inquiry ability of elementary scientifically gifted students : observing and inferring, designing an experiment, and concluding. *Journal of Science Education*, 39(3), 376-388.
- Timmerman B. E. C., Strickland, D. C., Johnson, R. L. & Payne, J. R. (2011). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, 36(5), 509-547.
- White, B., Frederiksen, J., & Collins, A. (2009) The interplay of scientific inquiry and metacognition: More than a marriage of convenience. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 175-205). Routledge.
- Yang, I. (2006). Guidance materials for each element of the school science inquiry process. Science Education Research Center, Korea National University of Education.
- Yang, T., Bae, M., Han, K., & Park, I. (2003). Scientifically gifted students' science related attitudes and its relationships with intelligence and science process skills. *Journal of the Korean Association for Science Education*, 23(5), 531-543.
- Yun D., & Choi A. (2019). Analysis of achievement standards, activities, and assessment items in the 2015 revised science curriculum and grade 7 science textbooks: focusing on science core competencies. *Journal of the Korean Chemical Society*, 63(6), 196-208.
- Zimmerman, B, & Schunk, D. (Eds.) (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Mahwah, NJ: Erlbaum.

저자정보

함은혜(공주대학교 부교수)
 이유경(숙명여자대학교 조교수)
 박소영(숙명여자대학교 교수)
 박혜진(한국과학기술원 선임연구원)
 이성혜(한국과학기술원 연구교수)

[부록] 과학탐구활동보고서 구성(일부)

문제해결

치즈는 왜 맛이 다를까?

💡 지금까지 우리는 치즈가 무엇인지, 치즈는 어떻게 만들어지는지, 다양한 치즈의 종류와 함유된 영양소가 무엇인지 배웠습니다. 앞서 배운 내용을 바탕으로 본격적으로 치즈를 만들어 보면서 미션을 해결해 봅시다.

미션명

리코타 치즈 레시피를 해부해 보자!

이번 미션에서는 기본 리코타 치즈의 레시피에서 조작 변인을 설정하여 분석하고, 여러분만의 특별한 치즈를 만들어 봅시다.

만드는 방법

- ① 우유 1000mL, 생크림 500mL, 소금 1/3 큰술, 레몬즙 5 큰술을 넣고 섞어 줍니다.
- ② 모든 재료를 섞은 후에 중불에서 한번 팔팔 끓여줍니다. 끓일 때 재료들을 절대로 휘젓지 않도록 합니다. 끓어서 거품이 많이 나면 약불로 줄입니다.
- ③ 한번 끓인 후에 약불로 순두부같은 덩어리가 생길 때까지 약 30분 정도 끓여줍니다. 만약 응고가 잘 되지 않는다면 레몬즙을 2 큰술 정도 더 넣어줍니다.
- ④ 덩어리를 면포에 올리고 면포를 짜줍니다. 물기를 짜 주는 정도에 따라 치즈의 맛이나 굳기가 달라집니다.
- ⑤ 물기를 다 뺐으면 면포에 감싼 채로 그릇에 두어 냉장고에 6시간 정도 보관합니다.

[출처 : <http://amyzzung.tistory.com/438>]

※ 불과 같이 위험한 도구를 사용할 때는 꼭! 부모님과 함께 해 주세요.

- 리코타 치즈 제조법은 사람마다 다릅니다. 레몬즙을 1스푼 넣을 수도 있고, 3스푼 넣을 수도 있습니다. 중불에서 30분, 혹은 1시간을 끓일 수도 있습니다. 이러한 이유는 각 레시피마다 치즈의 사용 용도와 추구하는 맛이 다르기 때문입니다.

문제해결

치즈는 왜 맛이 다를까?

1. 앞서 나온 레시피를 따라 '기본 리코타 치즈'를 만들어 보고 특징을 적어 봅시다.

사진	
특징	ex) 맛의 종류 : 단맛, 시큼한 맛 등 ex) 맛의 세기 : 진하다, 약하다 등 ex) 강도 : 딱딱함, 부드러움 등 ex) 촉감 : 알갱이가 있다, 부드럽다 등

여러분은 기본 리코타 치즈 레시피를 충실히 따랐나요?
이 기본 레시피는 어떻게 만들어진 것일까요? 끓이는 시간, 불의 세기, 우유/생크림의 비율, 레몬즙의 양, 물기를 짜주는 정도 등의 조건들은 치즈의 특성에 어떠한 영향을 미칠까요?

2. 리코타 치즈 레시피에서 치즈의 특성을 다르게 변화시킬 수 있는 조건을 적어도 두 개 이상 생각해 봅시다. 변화시킨 조건에 따라 치즈에 어떤 결과가 기대되는지 역시 설명해 봅시다. (조건 ex. 끓이는 시간)

조건1	
조건2	