2015 개정 교육과정에 따른 초등학교 3학년 과학과 교과서의 물질 영역에 나타난 자기 평가 분석

김민환 · 노준혜 · 노태희†

An Analysis of the Self-Assessments in the Matter Units of Elementary Science Textbooks for 3rd Grade Developed under the 2015 Revised National Curriculum

Kim, Minhwan · Roh, Junhye · Noh, Taehee†

국문 초록

본 연구에서는 2015 개정 교육과정에 따른 초등학교 3학년 과학과 교과서의 물질 영역에 나타난 자기 평가를 분석하였다. 선행연구를 바탕으로 '위치', '평가 대상 및 내용', '응답 형태', '채점 기준', '레퍼런스', '평가 준거의 제작', '피드백'으로 구성된 분석 기준을 개발하였다. 7개의 출판사에서 발행한 과학 교과서와 실험관찰 교과서 총 14권을 모두 분석하였다. 분석 결과, 교과서에서 자기 평가는 주로 '마무리' 영역에서 이루어졌으며, 이는 전통적인 결과 중심적 평가와 유사한 양상이었다. 다만, 일부 교과서는 '본문'과 '도입'에서도 자기 평가를 제시하였다. 자기 평가의 평가 대상은 대부분 학생이었으며, 학생의 '지식'과 '역량', '태도'를 상대적으로 고르게 평가하고 있었다. 자기 평가에 대한 '응답 형태'는 '평정'이 가장 많았으나, 교과서별로는 '평정'이나 '체크리스트' 중 하나로 응답 형태를 단일화하는 경향이 있었다. 모든 교과서가 자기 평가에 대한 '채점 기준'과 '레퍼런스'는 제시하지 않으며, '피드백'은 일부 교과서에서만 제한된 방식으로 활용하였다. 마지막으로 교육과정에서 강조하고 있는 자기 평가를 위한 '평가 준거의 제작'도 전혀 제시되지 않았다. 이 연구 결과를 토대로 과학과 교과서에 제시된 자기 평가의 개선 방안과 효과적인 자기 평가 활용을 위한 시사점을 논의하였다.

주제어: 초등학교 과학과 교과서, 자기 평가, 2015 개정 교육과정, 과정중심평가, 자기주도학습

ABSTRACT

In this study, we analyzed the self-assessments presented in the matter units of third grade science textbooks, according to the 2015 Revised National Curriculum. The analytical framework, comprising the 'position,' 'subject and content,' 'response form,' 'scoring criteria,' 'reference,' 'creation of assessment criteria,' and 'feedback,' was developed based on previous studies. We analyzed the science textbooks from seven publishers. The results revealed that self-assessments were primarily conducted in the 'closing' section of textbooks, showing a similar pattern to traditional result-oriented assessments. However, some textbooks presented the self-assessments in the 'body' and 'introduction' sections. The subjects of the self-assessments were mostly students, with a relatively balanced assessment of students' 'knowledge,' 'competency,' and 'attitude.' The most common 'response form' was 'rating,' although there was a tendency for the textbooks to standardize the response forms to either 'rating' or 'checklist.' Furthermore, none of the textbooks provided 'scoring criteria' or 'references' for the self-assessments. Only some textbooks used 'feedback' in a limited manner. Additionally, the 'creation of assessment criteria,' which was emphasized in the curriculum, was not presented at all. Based on the results of this study, we discussed strategies to improve the self-assessments presented in the science textbooks and implications for the effective use of self-assessments.

이 논문은 2023년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2023S1A5B5A17089236)
2024.07.19(접수), 2024.08.02.(1심통과), 2024.08.09(2심통과), 2024.08.13(최종통과)
E-mail: nohth@snu.ac.kr(노태희)

Key words: elementary science textbook, self-assessment, 2015 revised national curriculum, process-based assessment, self-directed learning

I. 서 론

그동안 과학교육 분야에서는 구성주의적 관점에 따라 과학 수업의 주체를 교사에서 학생으로 옮겨 학생 중심 수업을 실천하기 위한 노력이 이루어져 왔다. 이때 학생들은 학습 목표를 설정하고 학습을 점검하며 평가를 수행하는 모든 과정에 능동적으로 참여해야 하므로(Fiske & Taylor, 1991), 궁극적으로는 교수학습뿐 아니라 평가에서도 학생이 주체가 되어야 할 필요가 있다.

자기 평가(self-assessment)는 학생이 평가 결과와 피드백 등을 수동적으로 수용하는 것이 아니라 주어진 평가 준거를 이해하고 이를 직접 자신의 학습에 적용하는 학생 주도의 평가라고 할 수 있다(Boud, 2003; Brown & Harris, 2013). 때로는 주어진 평가 준거를 단순히 적용하기만 하는 것이 아니라 동료 혹은 교사와의 논의를 거쳐 평가 준거를 직접 생성하고 이를 적용할 수도 있다. 그러므로 구성주의 관점에서 자기 평가는 가장 이상적인 학생 중심의 평가 방법 중 하나라고 할 수 있다(남정희 등, 2004).

학생은 자기 평가의 과정을 통해 학습 목표를 명확히 이해하고 학습 과정을 점검함으로써 학습 과정 전반에 능동적으로 참여하게 된다. 따라서 자기 평가는 인지・정의적 측면에서 긍정적인 교육적 효과를 갖는다(Karaman, 2021; Panadero *et al*., 2017). 또 자기 평가는 학습 목표에 비추어 자신의 학습 과정과 성과를 스스로 점검하는 활동이므로(Boud & Falchikov, 2007), 메타 인지 능력을 향상할 수 있으며 자기조절능력의 향상에도 도움을 준다(Andrade, 2010; Ramdass & Zimmerman, 2008). 평가의 타당도와 신뢰도에 있어서도 자기 평가의 결과는 교사 평가의 결과와 크게 다르지 않다는 것을 여러 연구에서 밝히고 있다(황성원 등, 2001; Panadero *et al*., 2016).

최근에는 우리나라 교육과정에서도 자기 평가를 강조하고 있다. 2015 개정 교육과정의 총론 해설서(교육부, 2017)에서는 성취기준에 근거한 평가를 설명하며 “학생들을 평가의 과정에 포함시킨다. 예를 들면 수행의 과정과 결과물을 평가하는 준거를 교사와 학생이 함께 만들어 자기 평가하도록 한다(p. 107).”라고 서술하고 있다. 이처럼 2015 개정 교육과정에서 자기 평가를 강조하기 시작하면서, 이전 교육과정에 따라 개발된 교과서와는 달리 2015 개정 교육과정에 따라 개발된 모든 초・중・고등학교 과학 교과서에는 자기 평가가 포함되어 있다. 최근 발표된 2022 개정 교육과정에서는 해설서가 아닌 총론에서 평가 시 고려해야 할 원칙과 유의사항으로 ‘학생이 자신의 학습 과정과 결과를 스스로 평가할 수 있는 기회를 제공하는 것’을 명시함으로써 자기 평가를 더욱 강조하고 있다(교육부, 2022).

한편 구성주의 관점에서 학생 중심 평가의 다양한 방법 중 하나로 자기 평가가 널리 알려져 있음에도 불구하고 자기 평가와 관련된 연구는 아직 많이 부족한 실정이다. 우선 자기 평가와 관련한 연구는 국외에서도 대학생 이상을 대상으로 한 고등교육에서 많이 이루어졌다(Kearney, 2013; Taras, 2010; van Zundert *et al.*, 2010; Wanner & Palmer, 2018). 초・중등 교육에서는 자기 평가의 실태를 조사한 기초 연구마저 제한적인데 범교과적으로만 자기 평가의 실태를 조사한 연구가 있었다(박정과 이진주, 2015; Noonan & Duncan, 2005). 과학 교과에서는 통합과학(강남화와 김민지, 2020)이나 과학탐구실험(백종호 등, 2020)과 같은 특정 과목, 탐구학습 과정(Grob *et al.*, 2021)이나 자유학기제(김유라와 최애란, 2019a,b) 등과 같은 특정 교수 맥락에서 평가 실태를 조사할 때 자기 평가의 실태를 일부 포함하여 조사하는 수준으로, 자기 평가에 초점을 두고 실태를 조사한 연구는 찾아보기 어렵다.

이외에 국내 과학교육 분야에서 이루어진 자기 평가 관련 연구는 중・고등학교 학생을 대상으로 한 연구가 대부분이었다. 먼저 다양한 맥락에서 자기 평가를 적용하고 그 효과를 정량적으로 조사하는 연구가 많았다(남정희 등, 2002; 남정희 등, 2004; 이선우와 남정희, 2016; 조은비 등, 2021). 이외에 학생 자기 평가의 신뢰성을 조사하기 위하여 과학적 탐구과정기능에 대한 자기 평가를 교사 평가와 비교

한 연구도 있었다(황성원 등, 2001). 초등 과학교육 분야에서 이루어진 일부 연구는 과학 창의성을 효과적으로 평가하기 위한 방법을 탐색하기 위하여 과학 창의성에 대한 자기 평가의 결과를 동료 평가의 결과나 교사 중심의 평가 결과와 비교한 연구였다(김민주와 임채성, 2018; 김민주와 임채성, 2019). 따라서 초등 과학교육 분야에서 자기 평가를 적극적으로 활용하기 위한 다양한 연구가 활발히 이루어질 필요가 있다.

교과서는 학생들이 가장 많이 접하는 교수학습 자료로서 학생의 평가에도 중요한 역할을 한다. 따라서 교과서에서 제시된 자기 평가를 분석하고 이를 더욱 나은 방법으로 활용하는 방안을 탐색할 필요가 있다. 특히, 초등교사는 과학 수업에서 교과서를 중심으로 수업을 진행하여 교과서에 크게 의존하는 경향이 있다(김현정, 2020). 이에 이 연구에서는 초등학교 3학년 과학과 교과서의 물질 영역을 중심으로 교과서에서 제시된 자기 평가를 분석하였다.

II. 연구 방법

1. 분석 대상

이 연구에서는 2015 개정 교육과정에 따른 초등학교 과학과 교과서에 제시된 자기 평가를 분석하였다. 분석 대상은 금성 교과서, 김영사, 두산동아, 비상교육, 아이스크림 교과서, 지학사, 천재 교과서 등 7개 출판사에서 발행한 초등학교 3학년 과학과 검정 교과서이다. 초등학교 과학 교과서는 '과학'과 '실험관찰'로 구분되어 발행되기 때문에, 각각 7권씩 총 14권을 분석하였다. 분석 대상 단원은 3학년의 물질 영역으로 3학년 1학기 '물질의 성질'과 3학년 2학기 '물질의 상태'이다. 결과를 제시할 때는 7개 출판사의 교과서에 A부터 G까지 임의의 이름을 부여하였다.

선행 연구(남형채와 류성림, 2000; Boud, 2003; Brown & Harris, 2013; Panadero *et al*., 2016)에서 제시한 자기 평가의 정의를 바탕으로 자기 평가를 '학생이 능동적으로 자신의 학습 과정과 결과를 반성하고 이를 특정한 형태로 표현하는 과정'으로 정의하였다. 자기 평가 자체가 하나의 독립된 활동으로 제시되거나 활동 중 일부로 자기 평가가 제시되었을 때 하나의 자기 평가로 분석하였다.

초등학교 과학 수업에서 과학 교과서와 실험관찰 교과서는 수업 중에 함께 활용되며, 본 연구의 분석 결과 과학 교과서와 실험관찰 교과서가 동일한 자기 평가를 중복하여 제시하는 사례가 많았다. 따라서 동일한 자기 평가가 제시될 경우 이를 별도로 나타내었다.

2. 분석 기준

자기 평가와 관련된 연구(Boud, 1999; Brown & Harris, 2014; Harris & Brown, 2013; Panadero *et al.*, 2012; Panadero *et al.*, 2016)를 참고하여 예비 분석 기준을 개발하였다. 이 기준을 바탕으로 무작위로 선택한 일부 교과서를 분석하였고, 3명의 연구자가 반복적인 논의와 수정을 통해 최종 분석 기준을 Table 1과 같이 확정하였다.

1) 위치

교과서에서 자기 평가를 제시하고 있는 위치를

Table 1. Analytic framework for components and sub-components of self-assessment

범주	하위 영역	
위치	도입	
	본문	
	마무리	
	창의융합	
평가 대상 및 내용	학생	지식
		역량
		태도
	산출물	
	기타	
응답 형태	체크리스트	
	평정	
	기술	항목 제시
		항목 비제시
채점 기준	제시	
	없음	
레퍼런스	명시	
	비명시	
평가 준거의 제작	제시	
	제작	
피드백	제시	
	없음	

‘도입’, ‘본문’, ‘마무리’, ‘창의융합’의 네 가지로 분류하였다. ‘도입’은 각 소단원이나 대단원의 시작 부분에서 핵심 개념을 배우기 전에 핵심 개념과 관련된 활동을 통해 흥미, 호기심을 유발하고 앞으로 학습할 내용을 개괄적으로 소개하여 학습 방향을 제시하는 영역을 의미한다. ‘본문’은 핵심 개념을 문장이나 그림, 도표 등의 형태로 설명하고 교육과정에서 제시한 탐구 활동을 중심으로 다양한 탐구 과정을 제시하는 영역으로 설정하였다. ‘마무리’는 소단원이나 대단원이 마무리될 때 단원 전체에 관한 내용을 마인드맵이나 평가 문항 등의 형태로 정리하는 영역으로 정의하였다. 또한 모든 교과서에서 단원 전체의 학습 내용을 바탕으로 실생활 문제를 창의·융합적으로 탐구하는 과제를 구별하여 제시하고 있어 이를 ‘창의융합’으로 분석하였다.

2) 평가 대상 및 내용

평가 대상에 따라 자기 평가를 ‘학생’, ‘산출물’, ‘기타’의 세 가지로 분류하였다. ‘학생’은 “물체의 기능과 물질의 성질을 연결 지어 설명할 수 있어요.”나 “모둠원과 함께 여러 가지 물질로 이루어진 물체를 관찰했어요.”와 같이 학생 본인이나 본인이 속한 모둠의 인지적, 정의적 측면 등을 평가하는 자기 평가로 정의하였다. ‘산출물’은 “내가 만든 빨대 로켓의 좋은 점과 보완할 점을 써 봅시다.”와 같이 학습자들이 수행한 활동을 통해 만들어 낸 결과물의 기능, 편리성 등을 평가하는 것을 의미한다. ‘기타’는 “‘물질의 성질’을 되돌아보고 무엇을 알게 되었는지 스스로 생각하면서 써 봅시다.”와 같이 활동 후 느낀 점이나 배운 점을 기술하는 등 ‘학생’과 ‘산출물’을 종합적으로 평가하거나 둘 모두에 해당하지 않는 경우로 분류하였다.

평가 대상이 ‘학생’인 경우 평가 내용을 ‘지식’, ‘역량’, ‘태도’로 세분하였다. 이를 위해 과학 교과에서 평가와 관련된 선행 연구(김민환 등, 2022; 백종호 등, 2020; 윤도운 등, 2018)를 참고하였다. ‘지식’은 각 단원의 핵심 개념을 이해하고 이를 어떻게 적용하는지와 관련된 내용을 포함하였다. ‘역량’은 2015 개정 교육과정에서 제시된 과학과 핵심 역량 다섯 가지를 포함하며, 각 역량의 구체적 정의는 윤도운 등(2018)의 정의를 따랐다. ‘지식’과 ‘역량’에 포함되지 않은 과학에 대한 흥미와 가치 인식, 과학 학습 참여의 적극성 등의 정의적 영역을 ‘태도’로 분석하였다.

3) 응답 형태

자기 평가와 관련된 연구(Brown & Harris, 2014; Harris & Brown, 2013)를 참고하여 자기 평가에 대한 응답 형태를 ‘체크리스트(checklist)’, ‘평정(rating)’, ‘기술(description)’의 세 가지로 구분하였다. ‘체크리스트’는 학습자가 주어진 평가 준거에 대해 ‘예’ 또는 ‘아니오’와 같이 2단계 척도로 응답하거나 해당 준거를 만족할 경우 빈칸에 표시하는 방식이다. ‘평정’은 평가 결과를 3단계 이상의 척도로 표현하는 방법으로, 교과서에서는 ‘상, 중, 하’ 혹은 별의 개수 등으로 제시되며 학생들은 여러 단계의 척도 중 자신이 해당하는 수준을 선택한다. ‘기술’은 학습자가 자신의 학습 과정과 결과에 대한 평가를 글로 서술하는 방식을 말한다. 이때 학생들이 단순히 활동에 대한 느낌이나 감상을 서술하는 수준에 그칠 가능성이 있다(홍소영, 2018). 그러므로 “우리 모둠이 만든 새집의 잘된 점과 개선할 점을 물질의 성질과 관련지어 써 봅시다.”라는 준거에 대해 학생들이 응답하는 항목을 “잘된 점”과 “개선할 점”으로 나누어 제공하는 등 기술해야 하는 내용을 구체적인 하위 항목을 통해 제시하는 ‘하위 항목 제시’와 “느낀 점을 써 봅시다.”와 같이 기술해야 하는 내용을 추상적이고 포괄적으로 제시한 ‘하위 항목 비제시’로 ‘기술’ 항목을 세분하였다.

4) 채점 기준

자신의 학습 과정이나 결과를 평가할 때 명확한 기준에 따라 이를 수행하는 것이 바람직하다. 그러나 학생들은 스스로 이러한 기준을 설정하는 데 어려움을 겪을 수 있다. 그러므로 효과적인 자기 평가를 위해서는 구체적이고 명확한 채점 기준을 제공할 필요가 있다(Kostons *et al.*, 2012; Lindblom-Ylanne *et al.*, 2006; Panadero *et al.*, 2012). 예를 들어 “서로 다른 물질을 섞으면 물질의 성질이 어떻게 되는지 설명할 수 있어요.”라는 평가 준거에 따라 성취 수준을 세 단계로 나누고 세 단계에 대한 채점 기준을 “자신있게 설명할 수 있다.”, “알지만 설명하기 힘들다.”, “이해하기 어려웠다.”와 같이 제시할 수 있다. 이러한 맥락에 따라 자기 평가를 위한 채점 기준을

명시한 경우 '제시'로, 채점 기준을 제시하지 않은 경우 '없음'으로 분류하였다. 단, 응답 형태가 '기술'인 자기 평가는 명확한 채점 기준을 제시하기 어려우므로, 이 범주의 분석에서 제외하였다.

5) 레퍼런스

학습에 대한 자신의 인식이나 감각에만 의존하기보다는 외부의 레퍼런스를 통해 자신의 학습에 대한 추가적인 정보가 주어질 때 더욱 정확한 자기 평가가 가능하다(Boud, 1999; Panadero *et al.*, 2016). 그러므로 교사나 동료의 피드백, 과제 수행에 대한 점수 등과 같은 레퍼런스에 근거하여 자기 평가를 수행하는 것이 바람직하다(Joughin *et al.*, 2019; Yan & Brown 2017; Yan & Carless, 2022). 예를 들어 학생들에게 '물질의 상태' 내용을 학습하고 이를 친구에게 설명하는 활동을 통해 자신의 이해도를 확인한 후, 친구가 평가한 점수를 바탕으로 자기 평가를 진행하여 자신의 학습을 점검하도록 제시할 수 있다. 이러한 맥락에서 자기 평가에 참고할 수 있는 평가 문항이나 학습 내용, 동료의 피드백 등 레퍼런스가 명시적으로 제시되어 있는 경우 '명시'로, 명시적으로 제시되지 않는 경우 '비명시'로 분류하였다.

6) 평가 준거의 제작

학생들이 스스로 혹은 교사와 함께 평가 준거를 제작하는 것은 수행하는 과제를 더욱 세밀하게 파악할 수 있도록 한다(Andrade & Valtcheva, 2009). 학생들은 평가 준거를 제작하는 활동을 통해 학습 과정에 더욱 적극적으로 참여하고 학습을 통해 기대되는 성취나 학습 목표 등을 명확히 이해할 수도 있다(이선우와 남정희, 2016; Ainsworth & Christinson, 1998; Panadero, 2017). 또한 자기 평가의 준거를 제작하는 활동에 참여한 학생들은 자기 평가를 긍정적으로 인식하기도 하였다(Ross *et al.*, 1998). 따라서 교과서에서 자기 평가를 위한 준거를 제시한 경우 '제시'로, 학생이 준거를 제작하도록 하는 경우 '제작'으로 분석하였다. 단, 응답 형태가 '기술'인 자기 평가는 평가 준거의 제작 여부를 판단할 수 없으므로, 이 범주의 분석에서 제외하였다.

7) 피드백

학생들의 수행을 질적으로 향상하고 학습을 촉진하기 위해 자기 평가를 실시하고 난 후 피드백을 제공하는 것이 중요하다(Andrade, 2010). 예를 들어 자기 평가의 결과를 바탕으로 피드백을 위한 요소를 도출하고 이에 근거하여 추후 학습 활동을 제안하는 것은 자기 평가를 효과적으로 활용하는 방법으로 제안되고 있다(Boud, 2003; Harris & Brown, 2013). 예를 들어 "기체가 공간을 차지하고 있는지 어떻게 알 수 있을까요?"라는 준거를 제시하고 관련된 내용이 제시된 교과서의 쪽수나 추가 학습 자료를 안내할 수 있다. 이에 따라 자기 평가 이후 준거에 대한 해설, 평가 결과에 따른 추가 학습 범위나 학습 활동 등의 피드백을 제시하는 경우 '제시'로, 피드백을 제시하지 않는 경우 '없음'으로 구분하였다.

3. 분석 방법

확정한 분석 기준을 바탕으로 7개 출판사에서 발행한 초등 과학과 교과서 중 무작위로 한 출판사의 교과서를 선택하여 2인의 연구자가 각각 예비 분석을 실시하였다. 분석 결과가 다르게 나타난 부분에 대해서는 예비 분석을 실시한 2인의 연구자를 포함해 총 3인의 연구자가 함께 논의하여 연구자 간 일치도를 95% 이상 확보하였고 이후 한 명의 연구자가 나머지 교과서를 모두 분석하였다. 분석한 결과는 분석 기준의 각 항목에 해당하는 자기 평가의 빈도와 백분율(%)로 나타내었다. 하나의 자기 평가가 여러 평가 준거를 포함하여 다양한 평가 대상이나 내용을 평가할 수 있으며 복수의 응답 형태를 가질 수도 있다. 이에 '평가 대상 및 내용'과 '응답 형태'는 중복 코딩을 허용하였고, 따라서 두 항목에서는 백분율의 계가 100%를 초과하는 경우도 있었다. 또한 본 연구의 분석 대상인 초등학교 3학년 과학과 교과서에는 평가 준거를 학생들이 직접 제작하는 자기 평가가 제시되지 않는 등 분석 기준의 각 범주에 해당하는 사례가 나타나지 않는 경우가 있었다. 이 경우에는 초등학교 과학과 교과서의 교사용 지도서나 국내의 중·고등학교 과학 교과서, 국외의 과학 교과서 등의 사례를 참고하여 연구 결과를 논의하였다.

Ⅲ. 연구 결과 및 논의

7개 교과서에 나타난 자기 평가 빈도는 총 106개

Table 2. The frequencies of self-assessment by publisher

단위:개

빈도	A	B	C	D	E	F	G	계
과학	12	18	2	8	5	16	18	79
실험관찰	1	17(17)	13	15(4)	3(3)	1(1)	4(2)	54(27)
계	13	18	15	19	5	16	20	106

*괄호 안의 숫자는 실험관찰 교과서에 제시된 자기 평가 중 과학 교과서와 중복되는 자기 평가의 수를 의미하며, 계는 중복되는 자기 평가를 제외하고 계산함.

로, 한 교과서당 평균적으로 15.1개의 자기 평가가 나타났다(Table 2). 교과서별로 자기 평가의 제시 빈도는 큰 편차를 보였다. G 교과서는 과학 교과서 18개, 실험관찰 교과서 4개, 중복 2개를 포함한 총 20개로 자기 평가를 가장 많이 제시하였다. 반면에 E 교과서는 과학 교과서 5개, 실험관찰 교과서 3개, 중복 3개를 포함한 총 5개로 가장 적은 빈도를 보였다. 이외에 나머지 교과서는 각각 D 19개, B 18개, F 16개, C 15개, A 13개 순으로 자기 평가를 제시하였다.

또한 과학 교과서와 실험관찰 교과서에 자기 평가를 제시하는 방식도 차이를 보였다. 먼저 A와 C 교과서는 실험관찰 교과서에서 과학 교과서에 제시하지 않았던 새로운 자기 평가만을 제시하여 중복된 자기 평가가 없었다. 그러나 다른 교과서들은 과학 교과서에서 제시한 자기 평가의 일부를 실험관찰 교과서에서 중복하여 제시하였으며 그 방식은 조금씩 달랐다. B 교과서는 총 18개의 자기 평가 중 17개를 중복하여 제시하였는데, 과학 교과서에서는 “만드는 재료에 따라 물체를 분류할 수 있어요.”와 같이 평가 준거만을 제시하고 실험관찰 교과서에서 동일한 준거를 제시한 후 3단계로 평정하는 척도를 추가로 제시하는 형태였다. 다음으로 D, E, F, G 교과서는 ‘창의융합’에서 각각 2개, 2개, 1개, 2개의 자기 평가를 중복하여 제시하였다. 4개의 교과서 모두 과학 교과서에서는 별도의 평가 준거 없이 자기 평가를 위한 활동만을 제시하고, 실험관찰 교과서에서 구체적인 평가 준거나 응답 형태 등을 제시하는 방식이었다. 예를 들어 D 교과서의 경우 과학 교과서에서는 “설계한 물체의 장점과 단점을 물질의 성질과 관련하여 토의해 봅시다.”라고만 제시하였으며, 실험관찰 교과서에서는 “물체를 사용 목적에 맞게 창의적으로 설계했나요?”, “물체의 기능에 알맞은 물질의 성질을 이용했나요?”, “물체를 사용하기 편리한가요?”와 같이 자기 평가를 위한 준거와 학생들이 자기 평가 결과를 기록할 수 있는 빈칸을 제시하였다. 또한 D와 E 교과서는 ‘창의융합’ 이외의 부분에서도 각각 2개, 1개의 자기 평가를 과학 교과서와 실험관찰 교과서에 동일하게 제시하였다. “우리 모둠이 만든 새집의 잘된 점과 개선할 점을 물질의 성질과 관련지어 이야기해 볼까요?”와 같이 탐구 활동 중 알게 된 점이나 느낀 점을 기술하는 자기 평가가 이에 해당하였다.

1. 위치

교과서에서 제시된 자기 평가의 위치를 분석한 결과는 Table 3과 같다. ‘창의융합’을 제외한 세 가지 영역 중 ‘마무리’에 제시된 자기 평가가 56.6%(60개)로 가장 높은 비율을 차지하였으며, ‘본문’에 27.4%(29개), ‘도입’에 1.9%(2개)의 자기 평가가 제시되었다. 이처럼 ‘마무리’에 제시된 평가가 가장 많

Table 3. The frequencies of the positions of self-assessment by publisher

단위:개(%)

위치	A (13개)	B (18개)	C (15개)	D (19개)	E (5개)	F (16개)	G (20개)	계 (106개)
도입	0 (0.0)	0 (0.0)	0 (0.0)	2 (10.5)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.9)
본문	1 (7.7)	14 (77.8)	0 (0.0)	13 (68.4)	1 (20.0)	0 (0.0)	0 (0.0)	29 (27.4)
마무리	12 (92.3)	2 (11.1)	13 (86.7)	2 (10.5)	2 (40.0)	13 (81.3)	16 (80.0)	60 (56.6)
창의융합	0 (0.0)	2 (11.1)	2 (13.3)	2 (10.5)	2 (40.0)	3 (18.8)	4 (20.0)	15 (14.2)

았던 것은 모든 교과서가 분석 대상인 두 대단원을 마무리할 때 자기 평가를 제시하였기 때문이다. 또한 B와 D 교과서를 제외한 나머지 6개 교과서가 많은 수의 자기 평가를 '마무리'에서 제시했다. 특히 A 교과서는 '본문'에서의 1개(7.7%)를 제외한 나머지 자기 평가를 모두 소단원과 대단원을 마무리하는 '마무리'(12개, 92.3%)에서 제시하였다. C, F, G 교과서도 '마무리'에 제시된 자기 평가의 비율이 각각 86.7%(13개), 81.3%(13개), 80.0%(16개)로 모두 80%를 넘었다. '마무리'에 제시된 자기 평가의 전형적인 사례로는 대단원 마무리에 제시된 것을 들 수 있다. 예를 들어 F 교과서는 "물질의 상태" 학습을 모두 마친 대단원 마무리에서 '스스로 확인하기'로 "우리 주변의 물체를 이루는 물질의 상태에 관심과 호기심을 가져요.", "고체, 액체, 기체의 성질을 설명할 수 있어요.", "우리 주변에 기체가 있음을 예를 들어 설명할 수 있어요." 등의 준거에 따라 3단계로 평정하는 자기 평가를 제시하였다.

이처럼 자기 평가가 '마무리'에서 가장 많이 제시된 결과는 전통적인 평가 관점인 학습에 관한 평가(assessment of learning)에서 강조하는 결과 중심의 평가에 가까운 모습이라고 볼 수 있다. 그러나 자기 평가는 학습을 위한 평가(assessment for learning)나 학습으로서의 평가(assessment as learning)의 관점에서 학습 과정을 더욱 중요시하는 평가 방법이라고 할 수 있다(Black & Wiliam, 1998; Earl, 2003). 나아가 2015 개정 교육과정에서도 과정 중심 평가를 강조하고 있으므로 '마무리'뿐 아니라 '본문'이나 '도입'에서도 균형 있게 자기 평가를 제시할 필요가 있다.

한편, B와 D 교과서에서는 '본문'에 제시된 자기 평가의 비율이 각각 77.8%(14개), 68.4%(13개)로, 교과서에 제시된 자기 평가의 대부분을 차지하였다. '본문'에 제시된 자기 평가의 사례로는 물, 주스의 모양과 부피가 담는 그릇에 따라 어떻게 변하는지 관찰하는 탐구 활동 중 '스스로 평가하기'에서 "담는 그릇에 따라 물과 주스가 어떻게 변하는지 관찰할 수 있어요."라는 준거가 제시된 경우(교과서 B)를 들 수 있다. D 교과서는 '도입'에서 자기 평가를 제시한 유일한 교과서로(2개, 10.5%), 2개 모두 대단원을 배우기 전에 앞으로 배우고자 하는 내용을 기술하여 학습 계획을 세우는 형태였다. 예를 들어 물질의 성질 단원에서는 학습할 내용과 관련하여 무엇을 알고 싶은지 궁금한 점을 서술하도록 하였다.

'창의융합'은 대단원별로 한 번씩 제시되어 모든 교과서에서 총 두 번 나타났다. 이때 A 교과서를 제외한 모든 교과서가 두 번의 '창의융합'에서 각각 하나 이상의 자기 평가를 제시하였다. 먼저 B, C, D, E 교과서는 산출물에 대한 장단점, 보완할 점 등을 서술하는 자기 평가를 한 번씩 제시하여 총 2개의 자기 평가를 제시하였다. 그리고 F와 G 교과서는 한 '창의융합'에서 하나 이상의 자기 평가를 포함하여 각각 3개, 4개의 자기 평가를 제시하였다. 예를 들어 F 교과서는 물질의 성질을 이용하여 책가방을 설계하는 '창의융합'에서 탐구 활동에 포함된 형태로 자신이 설계한 책가방의 좋은 점과 보완할 점을 물질의 성질과 관련지어 기술하는 자기 평가를 제시하였고, 탐구 활동을 마친 후 제시된 '스스로 확인하기'에서 책가방을 사용하기 편리하게 설계했는지, 친구들에게 잘 소개했는지를 3단계로 평정하는 자기 평가를 포함하였다. A 교과서는 '창의융합'에서 자기 평가를 제시하지 않았지만, 다른 교과서에서 제시한 자기 평가와 유사하게 산출물의 기능에 대한 동료 평가를 제시하였다. 이처럼 대부분의 교과서가 '창의융합'에서 하나 이상의 자기 평가를 제시한 이유는 '창의융합'의 활동이 과학 개념의 직접적인 학습을 목적으로 하기보다는 대단원에서 학습한 내용을 바탕으로 실생활 문제를 과학적으로 탐구하는 적용 활동이기 때문에 자기 평가를 비교적 수월하게 적용할 수 있었던 것으로 보인다.

2. 평가 대상 및 내용

자기 평가의 대상과 내용을 분석하여 Table 4에 정리하였다. 평가 대상을 분석한 결과 '학생'에 대한 자기 평가가 155.7%(165개)로, 학생의 '지식'이나 '역량', '태도'를 평가하는 자기 평가가 많았으며 하나의 자기 평가에서 여러 내용을 평가하는 경우가 있어 그 비율이 100%를 넘었다. '산출물'과 '기타'는 각각 9.4%(10개)와 7.5%(8개)로 적은 비율을 차지하였다.

'학생' 대상의 자기 평가를 평가 내용에 따라 분석한 결과 '지식'이 56.6%(60개), '역량'이 58.5%(62개)로 유사한 비율로 나타났다. '태도'에 대한 평가도 40.6%(43개)로 적지 않은 비율이었다. 이처럼 평

Table 4. The frequencies of the targets and contents of self-assessment by publisher

단위:개(%)

평가 대상 및 내용		A (13개)	B (18개)	C (15개)	D (19개)	E (5개)	F (16개)	G (20개)	계 (106개)
학생	지식	10 (76.9)	7 (38.9)	11 (73.3)	8 (42.1)	2 (40.0)	12 (75.0)	10 (50.0)	60 (56.6)
	역량	12 (92.3)	15 (83.3)	12 (80.0)	6 (31.6)	2 (40.0)	6 (37.5)	9 (45.0)	62 (58.5)
	태도	1 (7.7)	4 (22.2)	13 (86.7)	13 (68.4)	2 (40.0)	2 (12.5)	8 (40.0)	43 (40.6)
	소계	23 (176.9)	26 (144.4)	36 (240.0)	27 (142.1)	6 (120.0)	20 (125.0)	27 (135.0)	165 (155.7)
산출물		1 (7.7)	1 (5.6)	0 (0.0)	2 (10.5)	3 (60.0)	1 (6.3)	2 (10.0)	10 (9.4)
기타		0 (0.0)	1 (5.6)	0 (0.0)	4 (21.1)	1 (20.0)	0 (0.0)	2 (10.0)	8 (7.5)

가 내용의 전체 결과는 세 가지 평가 내용이 상대적으로 고른 비율을 차지하고 있었다. 우리나라의 과학과 교육과정(교육부, 2015)은 지식뿐 아니라 탐구 능력이나 과학에 대한 태도나 흥미 등 또한 중요한 목표로 제시하고 있으므로 세 가지 내용에 대한 평가가 고른 비율을 차지한 결과는 긍정적이라고 할 수 있다. 그러나 출판사별로 결과를 살펴보면 각 교과서의 특징이 두드러졌다.

먼저 C, E, G 교과서는 세 가지 내용을 고르게 평가하고 있어 전체 결과와 유사한 모습을 보였다. 예를 들어 C 교과서는 물질을 섞었을 때 나타나는 성질 변화를 알아보는 탐구 활동에서 “서로 다른 물질을 섞었을 때 나타나는 변화를 설명할 수 있나요?”, “서로 다른 물질을 섞었을 때 나타나는 물질의 성질 변화를 관찰할 수 있나요?”, “탐구 활동에 적극적으로 참여하였나요?”와 같은 평가 준거를 제시함으로써 ‘지식’, ‘역량’, ‘태도’를 고르게 평가하였다.

반면, 전체 결과와 다른 양상을 보이는 교과서도 있었다. F 교과서는 ‘역량’(6개, 37.5%)이나 ‘태도’(2개, 12.5%)에 비해 ‘지식’(75%, 12개)을 평가하는 경우가 많았다. 이는 액체가 모양과 부피가 담는 그릇에 따라 어떻게 변하는지 관찰하는 탐구 활동 후 “용기의 모양이 달라지면 고체의 모양과 부피가 어떻게 되는지 설명할 수 있어요.”라는 평가 준거만 제시한 경우처럼 ‘지식’만을 평가하는 경우가 많았기 때문이다. 그러나 B 교과서는 같은 탐구 활동에서 ‘지식’뿐 아니라 “담는 그릇에 따라 물과 주스가 어떻게 변하는지 관찰할 수 있어요.”라는 준거로 ‘역량’도 함께 평가하였다. 또한 A 교과서는 ‘태도’(1개, 7.7%)에 대한 평가를 거의 제시하지 않았다. 그러나 D 교과서는 ‘태도’에 대한 평가가 13개(68.4%)로 세 가지 내용에 대한 평가 중 가장 큰 비율을 차지하였다. D 교과서는 “물과 주스를 관찰할 때 적극적으로 참여했어요.”와 같이 태도에 대한 자기 평가를 자주 제시하였다.

이처럼 교과서에 따라 일부 내용에 편중된 평가로 세 가지 내용을 고르게 평가하지 못하는 경우가 있었다. 하나의 자기 평가에서 여러 내용을 동시에 평가할 수 있으므로, 본 연구에서도 평가 내용의 범주를 중복 코딩하였다. 그러므로 향후 교과서 개발 과정에서는 세 가지 내용을 고르게 평가하도록 노력할 필요가 있다. 또한 자기 평가는 상대적으로 객관적 측정이 어려운 흥미, 호기심과 같은 정의적 영역을 학생들이 스스로 점검하도록 한다는 점에서 다른 평가 방법보다 정의적 영역을 더욱 효과적으로 반영할 수 있다(Karaman, 2021; Panadero *et al.*, 2017). 그러므로 ‘태도’를 강조한 자기 평가를 적극적으로 추가하여 학생들의 ‘태도’ 발전을 지원할 필요도 있다.

산출물에 대한 자기 평가 대신 동료 평가를 제시했던 C 교과서를 제외한 나머지 6개 교과서는 모두 ‘산출물’에 대한 자기 평가를 하나 이상 제시하였다. 예를 들어 A 교과서는 물질의 성질을 이용하여 다양한 기능이 있는 물체를 설계하는 탐구 활동에서 칫솔꽂이를 만들고 “내가 설계한 칫솔꽂이의 좋은 점과 보완할 점”을 평가하도록 하였다. 또한 D 교과서는 ‘창의융합’의 장난감을 설계하는 활동에서 “만든 장난감을 이용하여 놀이를 할 때 작동이 잘 되었나요?”와 같은 준거를 제시하여 산출물의 기능, 편리성 등을 다양한 측면에서 평가하도록 하였다.

한편, E 교과서는 ‘산출물’에 대한 평가가 60%(3개)로, 다른 출판사에 비해 높은 비율을 차지하였다. 이는 E 교과서의 자기 평가 개수가 다른 교과서에 비해 상대적으로 적은 5개에 그쳤고, 이 중 ‘산출물’

에 대한 평가가 두 번의 '창의융합'에서 1개씩 총 2개와 '본문'에서의 1개를 합하여 총 3개나 제시하였기 때문이다. 반면 '학생'에 대한 평가는 두 번의 대단원의 '마무리'에서 '학생'의 '지식', '역량', '태도'를 한 번에 평가하는 데 그쳤다. 자기 평가는 '산출물'을 평가할 때만이 아니라 '학생'을 평가할 때에도 유용한 방법이다. 그러므로 E 교과서와 같이 '창의융합'에서 '산출물'을 평가할 때만이 아니라 '도입', '본문' 등에서 '학생'의 '지식'이나 '역량', '태도'를 평가할 때에도 자기 평가를 적극적으로 활용할 필요가 있다.

'기타'에 해당하는 자기 평가는 D 교과서가 4개(21.1%)로 가장 많았다. 나머지 교과서는 '기타'에 해당하는 자기 평가를 제시하지 않거나(교과서 A, C, F), 1-2개의 자기 평가(교과서 B, E, G)만이 이에 해당하였다. 구체적인 사례로는 '본문'의 탐구 활동에서 "이 활동으로 알게 된 점을 이야기해 봅시다."와 같이 전반적인 평가 내용을 포괄하는 경우(교과서 D)나 "'물질의 상태'를 되돌아보고, 무엇을 알게 되었는지 스스로 생각하면서 써 봅시다."와 같이 배운 점에 대해 작성해 보는 경우(교과서 G)가 있었다.

3. 응답 형태

자기 평가에 대한 학생의 응답 형태를 분석하여 Table 5에 나타내었다. 응답 형태 중 '평정'이 71.7%(76개)로 가장 높은 빈도를 보였으며 모두 3단계 척도로 제시되었다. 특히 '평정'은 '체크리스트'(15개, 14.2%)와 '기술'(17개, 16.0%)에 비해 4배 이상의 빈도를 차지하였다. '평정' 형태의 자기 평가는 학생의 이해도를 빠르게 확인할 수 있으며 활용 방법이 간단하여 자기 평가에 익숙하지 않은 학생들에게 효과적이라는 특징이 있다(홍소영, 2018; Brown & Harris, 2013). 이에 초등학교 교과서에서는 학생들이 자기 평가에 익숙하지 않을 수 있다는 점을 고려하여 다른 응답 형태에 비해 '평정'을 많이 사용한 것으로 보인다.

분석 결과를 교과서별로 살펴보면 A, B, D, F, G의 5개 교과서에서 '평정'에 해당하는 자기 평가가 각각 92.3%(12개), 88.9%(16개), 68.4%(13개), 93.8%(15개), 80.0%(16개)로 전체 경향과 유사하게 가장 높은 비율을 차지하였다. 그리고 이중 A, B, F, G의 4개 교과서는 '체크리스트'를 전혀 활용하지 않았다. 반면 C 교과서는 '체크리스트'가 86.7%(13개)로, '평정'(2개, 13.3%)에 비해 아주 높은 비율이었다. 즉, 자기 평가를 5개밖에 제시하지 않았던 E 교과서를 제외하면, 나머지 교과서가 '체크리스트'나 '평정' 중 하나의 형태로 대부분의 자기 평가를 제시하여 응답 형태가 다양하지 못하고 단조로운 것으로 나타났다. 이러한 현상은 일관된 형식으로 평가 활동을 반복적으로 구성함에 따라 나타난 결과로 보인다.

그런데 이처럼 자기 평가의 응답 형태를 단일화하려는 시도가 평가 준거와 응답 형태 사이에 어색한 호응을 가져오는 경우가 있었다. 예를 들어 B 교과서는 "플라스틱 막대와 나무 막대를 관찰할 수 있어요."라는 평가 준거를 제시하고, '좋음', '보통', '부족함'을 나타내는 것으로 보이는 표정이 그려진 그림을 활용해 3단계의 평정 방식으로 응답하게 하였다. 즉, 부사나 형용사 없이 제시된 '관찰할 수 있다'는 서술어는 '관찰할 수 있다' 혹은 '관찰할 수 없다'의 2단계로 평가하는 것이 적절해 보이나, 구체적인 채점 기준도 주어지지 않은 상태에서 3단계의 평정을 요구하고 있었다. 그러므로 형식의 일관성을 유지하려는 노력과 함께 주어진 평가 준거나 평가 내용 등을 종합적으로 고려하여 자기 평가의 응답 형태를 유연하게 선택하려는 노력도 필요해 보인다.

응답 형태가 '기술'인 자기 평가는 전체 17개 중 '항목 비제시'가 10개(9.4%)로, '항목 제시'(7개,

Table 5. The frequencies of the response types of self-assessment by publisher

단위:개(%)

응답 형태		A (13개)	B (18개)	C (15개)	D (19개)	E (5개)	F (16개)	G (20개)	계 (106개)
체크리스트		0 (0.0)	0 (0.0)	13 (86.7)	2 (10.5)	0 (0.0)	0 (0.0)	0 (0.0)	15 (14.2)
평정		12 (92.3)	16 (88.9)	2 (13.3)	13 (68.4)	2 (40.0)	15 (93.8)	16 (80.0)	76 (71.7)
기술	항목 제시	0 (0.0)	1 (5.6)	0 (0.0)	0 (0.0)	3 (60.0)	1 (6.3)	2 (10.0)	7 (6.6)
	항목 비제시	1 (7.7)	1 (5.6)	0 (0.0)	6 (31.6)	0 (0.0)	0 (0.0)	2 (10.0)	10 (9.4)
									17 (16.0)

6.6%)에 비해 더 많았다. 기술해야 하는 항목이 모호하거나 잘한 점, 보완할 점, 알게 된 점 등을 복합적으로 작성하도록 하는 자기 평가는 응답자인 학생들에게 혼란을 가져와 평가가 타당하게 이루어지지 않을 가능성이 있다(Sadler, 2005). 또한 초등학생은 인지 수준이 높지 않고 정신 용량(mental capacity) 또한 충분하지 않으므로 서술해야 하는 항목을 더욱 명확히 제시하는 것이 필요할 수 있다. 그러므로 '기술' 형태의 자기 평가는 학생들이 작성해야 할 내용을 명확하게 제시하고 항목을 세분하여 나타낼 필요가 있다.

4. 채점 기준

자기 평가의 채점 기준 제시 여부를 분석한 결과, 모든 교과서에서 자기 평가의 채점 기준을 제시하지 않았다. 이와 달리 국외의 초등학교 과학 교과서와 국내 중・고등학교 과학 교과서에서는 채점 기준을 제시한 사례가 있었다. 먼저 우리나라 초등학교 3학년과 유사한 나이인 영국의 Year 4 학생이 사용하는 과학 교과서 중 하나는 "고체와 액체의 특성(How solids and liquids behave?)" 단원에서 실험을 한 후 "입자 모델을 사용하여 고체와 액체의 특성을 얼마나 잘 설명할 수 있나요?(How well can you use the particle model to explain the properties of solids and liquids?)"라는 준거에 3단계로 평정하는 자기 평가를 제시하였다. 그리고 3단계 척도에 대한 채점 기준을 "도움 없이 다른 사람에게 설명할 수 있다.(I get it! I can even explain to others.)", "도움이 조금 필요하다.(I need a little more help.)", "많은 도움이 필요하다.(I don't get it. I need a lot of help.)"로 세분하여 제시하였다. 또한 국내 중학교 과학 교과서 중 하나는 대단원을 마무리할 때 각 소단원에서 학습한 내용을 3단계 척도로 평정하는 자기 평가를 제시하고, 각 단계의 채점 기준을 '자신 있게 설명할 수 있다.', '알지만 설명하기 힘들다.', '이해하기 어려웠다.'로 명시하였다. 이처럼 채점 기준이 명확히 제시될 때 효과적인 자기 평가가 이루어질 수 있다. 특히 초등학생의 경우 자기 평가가 학생들의 내적 특성에 영향을 받아 관대하게 평가될 가능성이 크다(김민주와 임채성, 2019). 그러므로 향후 초등학교 과학 교과서에서 자기 평가를 제시할 때는 구체적이고 명확한 채점 기준을 제시하려는 노력이 필요하다.

분석 대상으로 삼았던 과학 교과서나 실험관찰 교과서에서는 채점 기준을 제시하지 않았지만, 교사용 교과서나 교사용 지도서에서는 채점 기준을 제시한 교과서가 있었다. B 교과서는 교사용 지도서에 추가로 제시한 수행평가에서 자기 평가를 포함하고 있었는데 여기서 구체적인 채점 기준을 제시하고 있었다. 예를 들어 "공기의 성질을 생각하며 압축 물휴지 변화와 수조의 물 높이 변화를 관찰하고, 관찰한 내용을 바탕으로 친구들과 의사소통하기"라는 준거에 대한 성취 정도를 "잘함", "보통", "노력 요함"의 3단계로 평정하는 자기 평가가 있었다. 그리고 "잘함"은 "공기의 성질을 생각하며 실험에 참여하고, 관찰한 내용을 바탕으로 친구들과 적극적으로 의사소통하여 공기의 성질을 설명할 수 있다.", "보통"은 "공기의 성질을 생각하며 실험에 참여하고, 친구들과 의사소통할 때 적극적으로 참여할 수 있다.", "노력 요함"은 "공기의 성질을 생각하며 실험에 참여할 수 있다."로 3단계의 성취 수준에 대한 채점 기준을 구체적으로 제시하였다. 이처럼 교사용 교과서나 교사용 지도서에 채점 기준이 제시될 경우 교사는 학생들에게 이를 구체적으로 안내하여 더욱 효과적인 자기 평가를 유도할 필요가 있다. 이뿐만 아니라 교과서나 교사용 지도서 등 교육과정 자료에 채점 기준이 제시되지 않더라도 교사가 직접 채점 기준을 마련하여 학생들에게 제공할 수도 있다. 이와 같이 교과서에 제시된 채점 기준을 그대로 활용하는 것이 아니라 교사가 자신의 수업 상황에 맞는 채점 기준을 직접 마련할 때 각 교실 상황에 특이적인 더욱 구체적인 채점 기준이 마련될 수 있다. 또한 이러한 교사의 노력과 별개로 자기 평가의 채점 기준을 교사용 지도서뿐 아니라 학생들이 사용하는 과학 교과서나 실험관찰 교과서에 직접 제시할 필요도 있을 것이다.

한편 하나의 준거에서 두 가지 이상의 평가 내용을 포함하여 채점 기준을 더욱 모호하게 하는 경우도 있었다. 예를 들어, A 교과서는 물질의 성질을 활용하여 칫솔꽂이를 설계하는 활동에서 "토의에 적극적으로 참여하고, 나만의 칫솔꽂이를 소개했어요."와 같이 하나의 준거에서 여러 내용을 평가하는 자기 평가를 제시하였다. 이처럼 하나의 준거에서 여러 가지 평가 요소를 혼합하여 제시하는 것은 학

생들에게 혼란을 야기하고 평가 의도를 파악하기 어렵게 만들 수 있다. 그러므로 구체적인 채점 기준 제시에 앞서 평가 준거를 명확히 할 필요도 있을 것이다.

5. 레퍼런스

자기 평가를 위한 레퍼런스의 제시 여부를 분석한 결과, 모든 교과서에서 레퍼런스를 제시하고 있지 않았다. 그러나 국내 중학교 과학 교과서 중 하나에서 레퍼런스를 명시한 사례가 있었다. 이 교과서는 대단원 마무리를 크게 마인드맵, 평가 문항, 자기 평가의 세 가지로 구성하였는데, 자기 평가는 교육과정에서 제시한 성취기준을 그대로 평가 준거로 삼고 이에 대한 성취를 3단계 척도로 평정하도록 하는 형태였다. 이때 각 평가 준거에 해당하는 평가 문항의 번호를 평가 준거와 응답 형태 바로 옆에 함께 배치하여 학생들이 해당 문항을 풀고 그 결과를 바탕으로 자기 평가를 수행하도록 하였다.

이처럼 레퍼런스를 명시하는 자기 평가는 학생들이 자신의 성취 수준을 정확하게 평가하도록 돕는다(Boud, 1999; Panadero *et al.*, 2016). 따라서 학생들이 자기 평가를 할 때 교사나 동료의 피드백, 학습 과정에서의 과제 점수 등 외부의 레퍼런스를 활용하여 평가할 수 있도록 도움을 제공해야 한다(Joughin *et al.*, 2019; Yan & Brown 2017; Yan & Carless, 2022). 특히, 초등학생은 중·고등학생보다 인지발달 수준이 비교적 낮으므로 자신의 학습을 평가하기 위한 근거가 부족한 채로 감각에만 의존하여 자기 평가를 할 가능성이 있다. 그러므로 초등학교 과학 교과서에서는 레퍼런스를 더욱 적극적으로 제시할 필요가 있다. 이와 동시에 교과서의 모든 자기 평가에 레퍼런스를 제시하는 것도 현실적으로 어려우므로 수업 시간에 이루어진 교사나 동료의 피드백 등을 참고하여 학생들이 자기 평가를 할 수 있도록 교사가 안내할 수도 있다.

6. 평가 준거의 제작

평가 준거의 제작 여부를 분석한 결과, 모든 교과서에서 자기 평가의 준거를 제시하고 있었으며 학생 스스로 혹은 교사와 함께 준거를 제작하도록 한 사례는 없었다. 2015 개정 초등학교 교육과정에서도 '학생들이 자신의 수행 과정과 결과물을 평가하는 준거를 교사와 학생이 함께 만들어 자기 평가'하도록 권고하고 있음에도 교과서에서는 이러한 사례가 나타나지 않았다. 국내 한 출판사의 고등학교 화학 교과서에서 동료 평가에 해당하긴 하였으나 평가 준거를 제작하도록 하는 사례를 찾아볼 수 있었다. 이 교과서에서는 화학이 실생활의 문제 해결에 기여한 사례를 조사하는 탐구 활동에서 모둠별로 조사한 사례를 발표하고 이를 모둠끼리 평가하도록 하는 동료 평가를 제시하였다. 이때 각 모둠의 발표를 평가할 수 있는 준거를 직접 제작하고 5단계 척도로 평정하도록 하였다.

학생들이 자기 평가의 준거를 제작하도록 하는 것은 과제를 세밀하게 파악하게 함으로써 여러 이점을 가질 뿐만 아니라 교육과정에서도 권고하고 있으므로 초등학교 교과서에서도 평가 준거를 제작하는 자기 평가의 비율을 높여갈 필요가 있다. 이때 초등학생의 인지발달 수준이 높지 않아 '지식'이나 '역량', '태도'를 평가하는 준거를 제작하기 어렵다면 '창의융합'에서 산출물을 제작할 때 산출물을 평가하기 위한 평가 준거를 제작해보도록 할 수 있다. 예를 들어 물질의 성질을 이용하여 다양한 기능을 갖는 물체를 설계하는 활동에서 물체의 심미성이나 편리성 등을 평가하기 위한 준거를 학생들이 직접 제안하고 학급 전체의 합의를 거쳐 준거를 완성하는 방식을 생각할 수 있을 것이다.

7. 피드백

전체 7개 교과서 중 C 교과서만 자기 평가에 대한 피드백을 제시하고 있었다. C 교과서는 대단원 마무리에 '체크리스트' 형태의 자기 평가를 제시한 이후 각 평가 준거와 관련된 내용을 설명하는 교과서의 쪽수를 제시하여 학생들이 관련 내용을 복습할 수 있도록 유도하였다. 또한 C 교과서는 교사용 지도서에서도 추가적인 피드백을 제공하였다. 예를 들어 "물체는 어떤 물질로 만들어졌을까요?"라는 평가 준거에 대해 교사용 지도서에 "물체는 금속, 플라스틱, 나무, 고무 등 다양한 물질로 만들어집니다."라는 답을 제공하였다.

B 교과서는 교과서가 아닌 교사용 지도서에서만 자기 평가에 대한 피드백을 제시하였다. 예를 들어 물질의 성질을 활용하여 놀이기구를 설계하고 토의하는 활동에서는 "토의에 적극적으로 참여하기"라

는 준거에 “잘함”, “보통”, “노력 요함”의 3단계로 평정하는 자기 평가가 있었다. 여기서 “노력 요함”에 대한 피드백으로 “미도달 학생은 물질의 성질 관련 내용을 다시 학습하도록 지도한다.”와 “미도달 학생은 토의에 적극적으로 참여하고, 다른 친구의 의견을 경청하는 태도를 갖도록 안내한다.”를 제시하여 자기 평가 결과에 따라 교사가 학생에게 피드백을 제공하도록 촉진하였다.

자기 평가 이후에 이루어지는 피드백은 학생의 학습을 촉진할 수 있다(Andrade & Valtcheva; 2009; Hattie & Timperley, 2007). 그러므로 교과서에서 자기 평가를 제시하는 것에서만 그치는 것이 아니라 이후 피드백까지 제시하는 것이 바람직하다. 그리고 현재 교과서와 지도서에 제시된 것 이외에 더욱 다양한 형태의 피드백을 활용할 필요도 있다. 예를 들어 자기 평가를 바탕으로 자신의 강점과 약점을 기록하는 것, 학습에 사용한 전략이나 산출물을 수정하는 것 등 자기 평가의 결과를 바탕으로 피드백에 필요한 요소를 도출하고 추후 학습 활동을 제안하는 방법도 유용할 수 있다(Boud, 2003; Harris & Brown, 2013; Harris & Brown, 2018). 실제로 국내의 한 고등학교 화학 교과서에서는 대단원 마무리의 자기 평가를 마친 후 “이 단원에서 더 공부해야 하는 부분을 점검하고, 학습 계획을 세워 보자.”라고 함으로써 자기 평가의 결과를 바탕으로 향후 학습 계획을 세우는 활동을 제시하고 있었다. 이처럼 다양한 피드백을 교수학습 맥락에 따라 적절히 제공할 때 학습을 위한 평가나 과정 중심 평가의 관점에서 더욱 바람직한 자기 평가가 이루어질 수 있을 것이다.

Ⅳ. 결론 및 제언

이 연구에서는 우리나라 초등학교 3학년이 사용하는 과학과 교과서의 물질 영역에 나타난 자기 평가를 분석하였다. 7개 출판사의 과학 교과서와 실험 관찰 교과서를 모두 분석한 결과, ‘마무리’에 제시된 자기 평가가 가장 많았다. 그러나 ‘본문’에 가장 많은 자기 평가를 제시한 교과서도 있었으며 ‘도입’에 자기 평가를 제시한 경우도 있었다. 한 교과서를 제외한 나머지 모든 교과서에서 ‘창의융합’에 자기 평가를 제시하였다. 대부분의 자기 평가가 ‘학생’을 평가하였으며 ‘산출물’을 평가하는 것은 적었다. ‘학생’의 평가 내용은 ‘지식’, ‘역량’, ‘태도’가 상대적으로 고른 비율을 차지하고 있었다. 그러나 일부 교과서는 ‘지식’을 많이 평가하는 등 세 가지 내용을 고르게 평가하지 못하였다. 응답 형태는 ‘평정’이 가장 많았다. 각 교과서가 ‘체크리스트’나 ‘평정’으로 응답 형태를 단일화하는 경향이 있었으며 이러한 경향이 평가 준거와 응답 형태의 어색한 호응을 가져오기도 하였다. ‘기술’ 형태의 자기 평가는 ‘항목 비제시’가 더 많았다. 모든 교과서가 자기 평가를 위한 채점 기준과 레퍼런스를 제시하지 않았는데, 채점 기준의 경우 교과서가 아닌 교사용 지도서에서 제시한 출판사가 있었다. 평가 준거를 제작하도록 한 교과서도 없었으며, 피드백의 경우 7개 출판사 중 C 교과서만 자기 평가와 관련된 학습 내용의 범위를 나타내는 피드백을 제시하였다.

그동안 구성주의 관점에서 학생 중심 평가의 다양한 방법 중 하나로 자기 평가가 널리 알려져 있음에도 불구하고 과학 교과에서 자기 평가를 효과적으로 사용하기 위한 고민과 노력은 상대적으로 부족하였다. 본 연구에서 교과서에 따라 자기 평가를 활용하는 방법이나 자기 평가의 구체적인 모습이 큰 차이를 보였던 결과는 이러한 주장을 뒷받침한다. 특히나 각 교과서의 형식상 일관성을 유지하기 위해 자기 평가의 응답 형태 또한 단일화했던 결과 등은 자기 평가에 대한 고려가 부족함을 더욱 잘 보여준다. 따라서 앞으로 과학 교과에서 자기 평가를 더욱 효과적으로 활용하기 위해서는 무엇보다 관성적으로나 형식적으로 자기 평가를 다루기보다는 교수학습 과정에서 실제 평가를 위한 목적으로 자기 평가를 의미 있게 활용하려는 자세가 필요할 것이다. 특히 교육과정에서도 자기 평가를 강조하고 있는 만큼 과학교육 연구자나 학교 현장의 과학 교사부터 자기 평가에 대한 인식 변화가 필요해 보인다.

본 연구의 결과를 바탕으로 향후 교과서를 집필할 때 자기 평가를 더욱 효과적으로 제시하기 위한 구체적인 개선점을 정리하면 다음과 같다. 먼저 과정 중심 평가를 강조하는 최근 교육과정의 사조와 학습을 위한 평가라는 평가 패러다임의 변화를 고려하여 자기 평가를 학습 과정 중에 활용할 필요가 있다. 평가 내용에 있어서는 지식과 역량, 태도를 고르게 평가하려는 노력이 필요하다. 그리고 일부 교과

서는 학생보다 산출물을 평가할 때를 중심으로 자기 평가를 활용했으므로 학생을 평가할 때에도 자기 평가를 적극적으로 활용할 필요가 있다. 교과서에 따라 응답 형태를 단일화하려는 경향이 있었으므로 평가 준거나 평가 내용, 탐구 활동의 특성 등을 종합적으로 고려하여 적절한 응답 형태를 다양하게 사용해야 할 것이다. 기술 형태의 자기 평가는 학생들이 작성해야 할 내용을 명확하게 제시하고 항목을 세분하여 나타낼 필요가 있다. 채점 기준은 모든 교과서에서 제시하고 있지 않았으므로 국외의 사례나 국내 중·고등학교의 사례를 참고하여 구체적이고 명확한 채점 기준을 제시하려는 노력이 필요하다. 레퍼런스의 경우 모든 자기 평가에서 제시하기 어렵다면 학생이 감각에만 의존하여 평가할 가능성이 높거나 평가에 어려움을 겪을 수 있는 자기 평가를 중심으로 레퍼런스를 제시해 볼 수 있다. 평가 준거를 제작하도록 하는 것 또한 전면적인 도입이 어려울 수 있으므로 조금씩 그 비율을 높여갈 필요가 있다. 즉, 결과에서 논의한 바와 같이 학생이 아닌 산출물을 평가하는 '창의융합'의 탐구 활동 등을 중심으로 학생들이 준거를 제작하는 것에 익숙해지도록 한 후 점점 그 적용 범위를 확대할 수 있다. 마지막으로 피드백은 일부 교과서에서만 제한된 형태로 활용하고 있었다. 그러므로 더욱 다양한 형태의 피드백을, 더욱 많은 자기 평가에서 제공할 필요가 있다.

교과서의 개선과 함께 과학 수업에서 자기 평가를 더욱 효과적으로 활용하려는 교사의 노력도 동반되어야 한다. 먼저 구체적인 채점 기준처럼 자기 평가와 관련된 일부 내용이 교사용 교과서나 교사용 지도서에만 제시되는 경우가 있었으므로 교사들은 이를 학생들에게 적극 안내할 필요가 있다. 레퍼런스의 경우 교과서에 제시되지 않더라도 수업 시간에 이루어진 교사나 동료의 피드백 등을 참고하여 학생들이 자기 평가를 할 수 있도록 교사가 안내할 수 있다. 평가 준거 또한 교과서에 모두 제시되었더라도 수업의 특성에 맞게 자율적으로 학생들이 평가 준거를 제작하도록 하고 이에 따라 자기 평가를 하도록 할 수 있다. 또한 교과서의 '마무리'에 자기 평가가 제시되더라도 교사가 수업 과정 중에 자기 평가를 활용한다면 교과서를 바람직하게 재구성하여 활용한 사례가 될 것이다.

이상에서 언급한 개선점에 따라 교과서를 개선하기 위해서는 자기 평가와 관련한 후속 연구 또한 활발히 이루어질 필요가 있다. 먼저 자기 평가는 학생이 스스로 자신의 학습을 평가하는 것이므로 학생을 대상으로 하는 기초 연구가 필요하다. 교수학습 맥락에 따라 가장 적절한 자기 평가의 응답 형태를 탐색하기 위하여 여러 응답 형태에 따라 학생의 자기 평가 결과가 어떻게 달라지는지 등을 조사하는 연구를 생각할 수 있다. 기술 형태의 자기 평가에서 항목을 명확히 제시하기 위해서는 학생들이 기술 형태의 자기 평가에 응답한 내용을 분석하는 연구도 필요할 것이다. 또한 학생들이 자기 평가의 준거를 제작하도록 한 후 준거 제작 과정에서 고려한 요소와 제작한 준거를 분석하여 학생의 평가 준거 제작을 위한 기초 자료로 활용할 수 있다. 이러한 기초 연구의 결과를 바탕으로 자기 평가를 위한 레퍼런스와 피드백의 다양한 형태를 개발하는 적용 연구도 이루어져야 한다. 다른 한편으로는 자기 평가에 대한 과학 교사의 인식과 실태를 조사함으로써 교사의 인식을 개선하고 자기 평가를 더욱 적극적으로 활용하도록 하기 위한 연구도 필요할 것이다.

참고문헌

강남화, 김민지(2020). 고등학교 통합과학 수행평가 사례를 통해 탐색한 교사의 수행평가 실천 특성과 PCK 사이의 관련성. 한국과학교육학회지, 40(3), 291–305.

교육부(2015). 2015 개정 과학과 교육과정. 세종: 교육부.

교육부(2017). 2015 개정 교육과정 총론 해설 중학교. 세종: 교육부.

교육부(2022). 2022 개정 초·중등학교 교육과정 총론. 세종: 교육부.

김민주, 임채성(2018). 초등과학영재학생의 과학창의성에 대한 자기 평가, 교사 평가, 객관적 평가의 비교 분석. 초등과학교육, 37(4), 440-454.

김민주, 임채성(2019). 초등과학영재학생의 과학창의성에 대한 자기 평가, 동료 평가의 비교 분석. 초등과학교육, 38(4), 439-452.

김민환, 박다해, 노태희(2022). 과학탐구실험의 '역사 속의 과학 탐구'에서 과학교사의 평가 실태와 평가 지향 조사. 한국과학교육학회지, 42(6), 597-610.

김유라, 최애란(2019a). 자유학기제 과학과 평가에 대한 교사의 인식과 실제. 한국과학교육학회지, 39(1), 143–160.

김유라, 최애란(2019b). 자유학기제 과학과 평가에서 과

학 교사가 겪는 어려움과 해결방안. 대한화학회지, 63(3), 166–182.
김현정(2020). 2015 개정 과학과 교육과정 초등학교 3~4 학년 운영 실태 분석. 현장과학교육, 14(3), 307-320.
남정희, 김지연, 한인식(2002). 자기평가가 중학교 학생들의 과학 개념 이해 및 과학 관련 태도에 미치는 영향. 대한화학회지, 46(3), 287-295.
남정희, 최준환, 공영태, 문성배, 이석희(2004). 자기평가에 대한 피드백 유형이 중학교 학생들의 과학 개념 이해와 과학 관련 태도에 미치는 영향. 한국과학교육학회지, 24(3), 646-658.
남형재, 류성림(2000). 수학 수업에서 학생의 자기평가 방법. 과학·수학교육연구, 23(3), 55-71.
박정, 이진주(2015). 학생 자기평가에 대한 초등교사의 인식과 실제. 학습자중심교과교육연구, 15(11), 601-620.
백종호, 변태진, 이동원, 심현표(2020). 2015 개정 교육과정 '과학탐구실험' 평가 도구 및 평가 현황 탐색. 한국과학교육학회지, 40(5), 515-529.
윤도운, 고은아, 최애란(2018). 2015 개정 과학과 교육과정의 과학과 핵심역량 하위요소 추출 및 적용. 학습자중심교과교육연구, 18(24), 1301-1319.
이선우, 남정희(2016). 고등학교 논의기반 탐구 과학 수업에서 학생 평가 활동이 반성적 사고에 미치는 영향. 한국과학교육학회지, 36(2), 347-360.
조은비, 정도준, 남정희(2021). 협상에 기반한 동료평가 및 자기평가 활동 과학 영재 고등학생들의 모델링에 미치는 영향. 대한화학회지, 65(6), 455-467.
홍소영(2018). 학생 자기평가의 학습효과에 관한 메타분석. 교육평가연구, 31(1), 309-331.
황성원, 김희경, 유준희, 박승재(2001). 중학교 3학년 학생들의 개방적 탐구에서 과학적 탐구과정기능에 대한 자기평가 수행 분석. 한국과학교육학회지, 21(3), 506-515.
Ainsworth, L., & Christinson, J. (1998). Student Generated Rubrics: An Assessment Model To Help All Students Succeed. Assessment Bookshelf Series. White Plains, NY: Dale Seymour Publications.
Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and self-regulation of learning. In H. L. Andrade, & G. J. Ciizek (Eds.), Handbook of formative assessment (pp. 90-105). New York, NY: Routledge.
Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. Theory Into Practice, 48(1), 12-19.
Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7-74.
Boud, D. (1999). Avoiding the traps: Seeking good practice in the use of self assessment and reflection in professional courses. Social Work Education, 18(2), 121-132.
Boud, D. (2003). Enhancing learning through self assessment. London, UK: Routledge.
Boud, D., & Falchikov, N. (2007). Rethinking assessment in higher education (pp. 191-207). London, UK: Routledge.
Brown, G. T., & Harris, L. R. (2013). Student self-assessment. In J. McMillan (Ed.), The Sage handbook of research on classroom assessment (pp. 367-393). Thousand Oaks, CA: SAGE.
Brown, G., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. Frontline Learning Research, 2(1), 22-30.
Earl, L. M. (2003). Assessment as Learning: Using Classroom Assessment to Maximize Student Learning (pp. 95-97). Thousand Oaks, CA: Corwin Press.
Fiske, S. T., & Taylor, S. E. (1991). Social cognition. New York, NY: McGraw-Hill.
Grob, R., Holmeier, M., & Labudde, P. (2021). Analysing formal formative assessment activities in the context of inquiry at primary and upper secondary school in Switzerland. International Journal of Science Education, 43(3), 407-427.
Harris, L. R., & Brown, G. T. (2018). Using self-assessment to improve student learning. New York, NY: Routledge.
Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. Teaching and Teacher Education, 36, 101-111.
Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81-112.
Joughin, G., Boud, D., & Dawson, P. (2019). Threats to student evaluative judgement and their management. Higher Education Research & Development, 38(3), 537-549.
Karaman, P. (2021). The impact of self-assessment on academic performance: A meta-analysis study. International Journal of Research in Education and Science, 7(4), 1151-1166.
Kearney, S. (2013). Improving engagement: the use of

'Authentic self-and peer-assessment for learning to enhance the student learning experience. Assessment & Evaluation in Higher Education, 38(7), 875-891.

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. Learning and Instruction, 22(2), 121-132.

Lindblom-Ylanne, S., Pihlajamaki, H., & Kotkas, T. (2006). Self-, peer-and teacher-assessment of student essays. Active Learning in Higher Education, 7(1), 51-62.

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools. Practical Assessment, Research, and Evaluation, 10(1), 17.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. Frontiers in psychology, 8, 422.

Panadero, E., Brown, G. T. L., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. Educational Psychology Review, 28(4), 803-830.

Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. Learning and Individual Differences, 22(6), 806-813.

Ramdass, D., & Zimmerman, B. J. (2008). Effect of self correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. Journal of Advanced Academics, 20(1), 18-41.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1998). Skills training versus action research in-service: Impact on student attitudes to self-evaluation. Teaching and Teacher Education, 14(5), 463-477.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. Assessment & Evaluation in Higher Education, 30(2), 175-194.

Taras, M. (2010). Student self-assessment: Processes and consequences. Teaching in Higher Education, 15(2), 199-209.

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. Learning and instruction, 20(4), 270-279.

Wanner, T., & Palmer, E. (2018). Formative self-and peer assessment for improved student learning: the crucial factors of design, teacher participation and feedback. Assessment & Evaluation in Higher Education, 43(7), 1032-1047.

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. Assessment & Evaluation in Higher Education, 42(8), 1247-1262.

Yan, Z., & Carless, D. (2022). Self-assessment is about more than self: the enabling role of feedback literacy. Assessment & Evaluation in Higher Education, 47(7), 1116-1128.

김민환, 전남대학교 교수(Minhwan Kim; Professor, Chonnam National University)
노준혜, 서울대학교 대학원생(Junhye Roh; Graduate student, Seoul National University)
† 노태희, 서울대학교 교수(Taehee Noh; Professor, Seoul National University)