

초등교사와 예비초등교사의 과학평가 실태와 지향

이혜민 · 강석진[†]

Elementary School Teachers' and Preservice Elementary School Teachers' Actual and Preferred Types of Science Assessment

Lee, Haemin · Kang, Sukjin[†]

ABSTRACT

In this study, we investigated elementary and preservice elementary school teachers' actual and preferred types of science assessment focusing on the purpose and the method of assessment. Participants were 75 elementary school teachers and 93 preservice elementary school teachers. The questionnaire adopted from previous relevant studies was administered in order to examine the types of assessment which the teachers have actually implemented and the preferred types of assessment for hypothetical situations known as the prototypical examples of constructivist assessment. The results revealed that most teachers and preservice teachers tend to possess the summative-oriented purpose of assessment. An affective assessment purpose focused on evaluating students' interest, attitude, and curiosity toward science was newly identified in this study. In analyzing teachers' responses from the viewpoint of the method of science assessment, responses of an informal mode such as observation and portfolio were noticeable. However, most of them were found to correspond to the traditional summative-oriented purpose of assessment. For prototypical constructivist assessment situations, teachers and preservice teachers were found to prefer the performance and the informal modes than the measurement mode of the assessment methods, whereas their purposes of the assessment were not found to be constructivist.

Key words: constructivist assessment, actual type of science assessment, preferred type of science assessment, elementary school teacher, preservice elementary school teacher

I. 서 론

평가는 학생의 학습 과정이나 학습 결과를 개선하는 데 필요한 교사의 의사결정을 지원하기 위한 정보의 수집, 해석, 활용 등에 관련된 모든 활동으로 정의할 수 있는데(McMillan, 2015), 평가는 수업의 모든 측면과 연결되어 있으므로 학습의 질을 결정하는 중요한 요인이다(김성훈 등, 2018). 평가는 학습 목표를 구체화하고 학습 과정과 결과를 점검하여 학생의 학습을 이해하며 학습의 개선책을 제안할 수 있도록 이루어져야 한다. 과거에는 평가의 주요 기능이 '평가를 위한 평가(assessment of evaluation)'였지만, 최근에는 '학습으로서의 평가(assessment as

learning)' 그리고 한 발 더 나아가 '학습을 위한 평가(assessment for learning)'로 평가의 기능이 변화했다. 즉, 평가는 최종적인 목표 도달 수준을 확인하는 총괄평가보다는 학습 과정에서 지속적으로 이루어지는 과정 중심의 형성평가가 되어야 한다(성태제, 2013; Butler & McMunn, 2006). 형성평가는 학생의 이해도를 점검하는 단순한 도구의 수준을 넘어 교수와 학습을 연결하는 매개가 되어야 하고(김진규, 2013), 학습 활동으로 평가를 구성해야 한다는 주장까지 있다(Earl, 2003). 이와 같이, 교수학습과 평가를 연계하는 평가의 형성적 기능 강조는 평가의 새로운 흐름이라고 할 수 있다(김난옥과 손원숙, 2019; Birenbaum *et al.*, 2015). 우리나라의 현

2019년 전주교육대학교 국립대학육성사업의 연구비 지원으로 수행하였음.

2019.12.15(접수), 2019.12.26(1심통과), 2019.12.26(2심통과), 2019.12.29(최종통과)

E-mail: kangsj@jnu.ac.kr(강석진)

행 2015 개정 교육과정에서도 핵심 역량을 구현할 수 있는 교수-학습-평가의 연계성과 학생의 학습과 성장을 지원하는 과정 중심 평가를 강조하고 있다.

그러나 실제 교육 현장에서는 형성평가나 교수 학습에서 평가 결과의 활용이 효과적으로 이루어지지 못하고 있다. 72개국에 참여한 PISA 2015의 분석 결과, 대다수 참여국에서 평가를 교수학습 개선을 위해 사용하려는 의도가 두드러졌지만, 우리나라에서는 총괄적인 의도로 평가가 시행되고 평가 결과의 활용도가 낮으며 평가의 형성적 활용도 활발하지 않은 것으로 나타났다(김난옥과 손원숙, 2019). 즉, 학교 현장에서는 수업의 마지막에 선다형 문항의 쪽지시험으로 정답 여부만 확인하는 형성평가가 이루어지고 있으며, 학생들에게 후속 학습을 위한 피드백을 제공하거나, 이후 교수학습에 반영하는 등 진정한 의미에서의 형성평가는 이루어지지 않는 실정이다(김희경 등, 2014). 이러한 현상은 일선 학교에서 형성평가 도구를 제작하고 평가 결과를 분석하여 맞춤형 피드백을 제공하는 시스템을 마련하기 쉽지 않고, 나이스를 기반으로 하는 현재의 학교 평가 체제가 형성평가에 적합하지 않다는 점 등이 원인인 것으로 분석되었다(김인숙 등, 2018).

그런데 한편으로는 교육 현장에서 형성평가가 적절히 사용된다는 실태 보고도 있다. 박지현 등(2017)의 연구에서는 거의 모든 교사들이 수업에서 형성평가를 사용한다고 응답했고, 장기적인 학생 발달을 관찰하기 위해 형성평가 전략을 사용하는 교사도 95%에 이르렀다. 또한 대부분의 교사들이 학생들에게 피드백을 제공한다고 응답했고, 피드백의 질에 대해서도 학생들의 향상을 도울 수 있을 만큼 상세하다고 응답한 교사가 76%에 이르렀다(박지현 등, 2017). 정은영과 최원호(2014)도 초등교사들이 과학 수업에서 사용하는 수행평가의 비율이 전체의 43%에 이르고, 실시 목적에 부합한다고 생각하는 수행평가의 비율도 70%에 이르는 것으로 보고했다. 그러나 교사들의 인식과 달리, 학생들은 피드백이 좀 더 구체적이기를 바라는 것으로 나타났다(박지현 등, 2017). 이러한 결과는 실제 교육 현장에서 이루어지는 평가의 실태와 평가를 담당하는 교사들의 인식 사이에 괴리가 존재할 가능성을 시사한다. 수행평가에서 수행의 과정을 중시하기보다는 여전히 결과물 중심의 평가가 이루어진다는

연구(신정윤과 양일호, 2017)나 형성평가의 목적으로 학습 목표 성취 여부를 확인하는 것이라고 응답한 교사들이 많다는 연구(강대중과 염시창, 2014; 노태희 등, 2015)는 이러한 가능성을 뒷받침한다.

교사의 평가 전문성 함양은 성공적인 교수학습 및 교사 전문성 발달을 위한 핵심적인 요소이다(민희정, 2012; Falk, 2012; Siegel & Wissehr, 2011). 다양한 평가 유형을 활용해서 학생들의 수준을 파악하고, 피드백과 적절한 학습 경험을 제공하기 위해서 교사의 구체적인 계획과 전문성 신장이 필요하다. 그러나 PISA 2015 결과에서는 우리나라 교사들의 평가효능감이 72개 국가 중 두 번째로 낮았으므로(김난옥과 손원숙, 2019), 교사들의 평가 전문성을 함양하기 위한 방안을 마련할 필요성이 있다. 특히, 초등교사들은 여러 과목을 동시에 가르쳐야 하므로 중등교사에 비해 과학 평가 전문성이 상대적으로 부족할 것으로 예상할 수 있다. 그러나 선행 연구에서는 중등교사에 비해 초등교사들이 평가 실행에 대한 자신감이 더 높았고(박지현 등, 2017), 평가 실시, 채점, 성적 부여와 평가 방법의 선정과 관련해서 자신의 전문성 수준을 높게 평가하는 것으로 나타났다(강훈식과 강석진, 2015). 강대중과 염시창(2014)도 우리나라의 초등교사 중 평가 자신감이 낮은 교사는 7%로 적었고, 자신의 평가 전문성이 낮다고 인식하는 교사도 10%에 불과하여 평가에 대해 큰 문제의식이 없다고 보고했다.

초등학교 현장에서의 평가에 대해 이와 같이 상반된 결과가 보고된다는 것은 국가적 차원에서 추진하는 과정 중심의 형성평가와 초등교사들이 생각하는 형성평가 사이에 적지 않은 불일치가 존재함을 시사한다. 따라서 과정 중심의 형성평가가 초등 과학교육 현장에 정착하기 위해서는 실제 과학 수업 상황에서 이루어지는 평가의 목적과 방법이 적절한지, 그리고 교사들이 지향하는 평가와 실제로 이루어지는 평가가 일치하는지 등에 대한 총체적인 정보를 수집할 필요가 있다. 이를 위해 이 연구에서는 우선 초등학교 과학 수업에서 진행되는 평가의 실태를 파악하고자 교사들이 실시했다고 응답한 과학평가 사례들을 조사하고, 이를 평가 목적과 방법 측면에서 분석했다. 또한 초등학교 교사들의 구성주의적 과학 평가에 대한 지향을 조사하기 위해 구성주의적 평가의 실시 가능성이 높은 전형적인 과학 평가 상황에 대한 교사들의 평가 계획

을 평가 목적과 방법 측면에서 분석했다. 한편 과정 중심의 형성평가를 과학교육 현장에 확산시키기 위해서는 효과적인 교사 양성 프로그램이 필수적인데, 예비초등교사들의 생각을 조사하고 초등교사와 비교, 분석하는 연구는 이를 위한 기초 자료를 제공할 수 있을 것이다. 따라서 이 연구에서는 예비초등교사들을 대상으로 장차 과학교육 현장에서 실시하려고 계획하는 과학 평가 사례와 구성주의적 과학 평가에 대한 지향을 조사한 후, 이를 평가 목적과 방법 측면에서 분석했다.

II. 연구 방법

1. 연구 대상

이 연구의 참여자는 전라북도 지역의 초등학교 교사 75명과 예비초등교사 93명이다. 연구 대상의 몇 가지 배경 정보는 Table 1과 같다. 초등학교 교사는 전라북도 지역의 7개 초등학교에서 각 학교의 규모를 고려하여 5~9명의 교사를 무선 표집하여 설문 조사를 실시하였다. 초등학교 교사 중 교육 경력 5년 이하의 초보 교사는 34.7%였고, 교육 경력 5년 초과와 경력 교사는 65.3%였다. 석사 이상의 학위를 가진 교사의 비율은 37.3%였다. 그러나 초등교사 중 과학 평가 관련 연수를 들은 경험이 있는 교사는 13.3%로 적었다. 예비초등교사의 경

우, 교육 평가 강의를 들은 학생은 88.2%였다. 그러나 과학 평가 관련 강의는 개설되어 있지 않았고, 과학교과교육론 관련 과목에서 과학 평가 강의를 들은 경험이 있는 학생도 6.5%에 불과했다.

2. 검사 도구

초등교사의 과학 평가 실태와 예비초등교사가 교사가 된 후 실시하려는 과학 평가를 조사하기 위해 선행 연구(노태희 등, 2015)의 검사 도구를 초등학교 맥락에 맞춰 일부 수정하여 사용했다. 초등학교 사용 검사지에는 ‘과학 과목에서 실시했던 대표적인 평가 사례’를 제시하도록 요구했고, 초등예비교사 사용 검사지에는 ‘초등학교에서 과학 과목을 가르친다면 어떤 평가를 실시할지’를 제시하도록 요구했다. 이 검사지는 선택형 문항이 응답에 미치는 영향을 배제하기 위해 교사들이 실시했던 과학 평가 사례를 목적과 방법 측면에서 자유롭게 기술하는 방식으로 구성되어 있다. 다음으로, 초등교사와 예비초등교사가 지닌 구성주의적 과학 평가에 대한 지향을 조사하기 위해 선행 연구(노태희 등, 2015)의 검사 도구를 사용했다. 이 검사지는 구성주의적 평가의 시행을 기대할 수 있는 전형적인 과학 평가 관련 상황(과학실에서의 실험 수업, 사회적 이슈에 대한 토의 수업, 자유탐구 지도)을 제시한 후, 그 상황에서 자신의 평가 계획을 목적과 방법 측면에서 자유롭게 기술하는 방식으로 구성되어 있다.

Table 1. The characteristics of the participants

특성	빈도(%)	
	교사	예비교사
성		
여성	42(56.0)	63(67.7)
남성	33(44.0)	30(32.3)
경력		
5년 이하	26(34.7)	-
5년 초과	49(65.3)	-
최종 학위		
학사	47(62.7)	-
석사 이상	28(37.3)	-
과학 평가 연수(강의) 경험		
유	10(13.3)	6(6.5)
무	65(86.7)	87(93.5)

3. 결과 분석

이 연구에서 수집한 교사들의 응답 중 무응답과 불성실한 응답을 제외한 결과, 과학 평가의 실태 분석에서는 초등교사 197개(1인당 평균 2.6개)와 예비초등교사 279개(1인당 평균 3.0개)의 응답이 분석 대상이었다. 구성주의적 과학 평가 지향 분석에서는 초등교사의 경우 실험 상황 96개, 토의 상황 94개, 자유탐구 상황 99개의 응답을 분석 대상으로 했고, 예비초등교사의 경우 실험 상황 91개, 토의 상황 92개, 자유탐구 상황 85개의 응답을 분석 대상으로 했다.

초등교사와 예비초등교사의 응답 사례에 드러난 과학 평가의 목적은 크게 총괄평가와 형성평가로 구분한 후(김난옥과 손인숙, 2019), 선행 연구(노태희 등, 2015; Davis & Neitzel, 2011)에서 사용한 분석틀을 기준으로 세분화했다. 응답 사례에 드러난

과학 평가 방법은 선행 연구(노태희 등, 2015; 백순근, 2000; Wang *et al.*, 2010)를 바탕으로 측정(measurement), 수행(performance), 비형식(informal)의 세 가지 유형으로 분류한 뒤, 다시 하위 유형으로 세분했다. 전통적인 지필평가 방법인 측정은 선택형과 서답형 평가로 분류했고, 학생들의 행동이나 사고 과정을 측정하는 평가 방법인 수행은 실험, 보고서, 논술형, 발표, 포트폴리오, 토의 평가로 분류했으며, 비형식적인 방법으로 학생을 진단하는 비형식 평가는 관찰, 면담, 동료, 자기 평가로 분류했다.

초기 분석틀을 바탕으로 초등교사와 예비초등교사의 응답을 분석하여 분석틀을 점검하고 일부 수정했다. 두 사람의 분석자가 응답 중 일부를 선택하여 각자 분석한 뒤 일치도를 확인했다. 분류가 일치하지 않은 응답에 대해서는 분석자 간의 논의를 거친 뒤 분류 기준을 수정했다. 이 과정을 여러 번 반복하여 분석자 간 일치도가 .97에 도달한 뒤, 분석자 중 한 사람이 모든 응답을 다시 분석하여 분류했다.

III. 연구 결과 및 논의

1. 과학 평가 실태

평가 목적은 평가의 성격, 평가의 시행 및 채점 방법, 결과 활용 등을 결정하는 중요한 요인이다 (Istance & Dumont, 2010). 과학 평가의 실태에 관한 초등교사와 예비초등교사의 응답을 평가의 목적 관점에서 분석한 결과는 Table 2와 같다. 초등교사와 예비초등교사가 제시한 평가 사례의 대부분은 총괄평가를 목적으로 하는 것으로 나타났다(초등교사: 98.0%, 예비초등교사: 85.3%). 이는 과학 평가의 목적이 대부분 총괄평가라는 선행 연구(강대중과 염시창, 2014; 노태희 등, 2015; Wang *et al.*, 2010)와 동일한 맥락에 있는 것으로 볼 수 있다.

총괄평가 목적 중 가장 많았던 유형은 초등교사의 89.4%와 예비초등교사의 72.8%를 차지한 학생 수준 평가였다. 초등교사들은 ‘학업 성취도를 평가하기 위하여 객관식 및 서술형 문항을 출제하여 평가했다.’ 혹은 ‘단원의 기초적인 내용을 이해했는지 알아보기 위하여 기본 수준의 객관식 문항과 서술형 문항을 출제하여 평가했다.’와 같이 인지적 측면에서의 학습 목표 도달 여부를 평가한 사례를 많이 제시했다. 예비초등교사들도 주로 인지적 측면에서

Table 2. The purposes of assessment shown in participants' responses

목적	빈도(%)	
	교사	예비교사
총괄평가	193(98.0)	238(85.3)
학생 수준 평가	176(89.4)	203(72.8)
학생 선발	14(7.1)	18(6.4)
학습 부진아 확인	0(0.0)	3(1.1)
수업 참여도 측정	3(1.5)	14(5.0)
형성평가	4(2.0)	25(9.0)
학습 기회 제공	4(2.0)	22(7.9)
기술적 피드백 제공	0(0.0)	3(1.1)
정의적 평가	0(0.0)	16(5.7)
계	197(100.0)	279(100.0)

학생 수준을 평가하겠다는 응답이 많았지만, ‘실험 절차/결과를 잘 알고 있는지 파악할 목적으로 지필 평가를 실시한다.’와 같이 다소 원론적인 응답을 하는 경향이 강했다. 선행 연구에서 초등교사의 경우 30%가 학생 선발을 목적으로 하는 평가 사례였지만(노태희 등, 2015), 이 연구에서 학생 선발 목적의 평가 사례는 초등교사 7.1%와 예비초등교사 6.4% 정도로 매우 적었다. 중등학교와 달리 초등학교에서는 과학 교과에서 선발을 목적으로 학생을 평가해야 하는 경우가 드물기 때문에 실제 평가 사례가 적은 것으로 생각된다.

일반적으로 형성평가에서는 학생들도 적극적으로 참여하는 평가, 학생들에게 평가 결과를 피드백 하는 평가, 교수학습 방법을 개선하는 평가와 같이 학습을 위한 평가로서의 기능이 강조되고 있다(성태제과 임현정, 2014). 평가 결과를 형성적으로 활용할 경우, 학생이 지각하는 평가 실재와 교사가 보고한 피드백 사이에 일관성이 높다는 연구 결과(김난옥 등, 2018)도 교수학습에서 형성평가의 중요성을 보여준다. 비록 소수이기는 하지만, 아래와 같이 이 연구에서 형성평가로 분류된 과학 평가 사례에서는 학습 기회를 제공하려는 목적이 분명히 드러나고 있었다. 그러나 교사와 학생 사이의 상호작용을 강조하는 형성평가는 초등교사의 경우 2.0%, 예비초등교사의 경우 9.0% 정도로 매우 적어서 초등학교 과학 평가의 현실을 개선할 필요성이 있는

것으로 파악된다.

수업 목표 도달 여부를 측정할 때 이산화탄소(산소) 발생 원리를 알 수 있도록 이산화탄소(산소) 발생 실험을 통해 얻은 결과를 발표함으로써 원리를 이해하는 평가를 했다.

(초등교사 A)

생물의 모습에 대해 배우고 난 뒤 그 생물을 찾을 수 있는지 확인하기 위해 야외로 나가 직접 생물을 채집해 보도록 했다.

(초등교사 B)

변인 통제, 가설 설정 등의 과학적 능력을 기르기 위한 목적으로 1년 동안 실시하는 실험 보고서법을 실시한다.

(예비초등교사 A)

일상생활에서의 과학적 사고를 기르기 위해 학생들의 사고 과정을 파악하는 구술법, 면담법을 실시한다.

(예비초등교사 B)

다른 과목과의 창의 융합을 위한 목적으로 과학 상상력을 측정하는 수행평가(글짓기 및 상상화)를 실시한다.

(예비초등교사 C)

일부 예비교사는 ‘과학에 대한 관심이 얼마나 있는지 알아보기 위한 목적으로 정의적 측면을 평가하는 질문지법을 실시한다.’, ‘학생들이 평소 과학에 대한 흥미가 있는지 판별하기 위한 목적으로 과학적 태도를 측정하는 관찰법을 실시한다.’, ‘학생의 과학에 대한 전반적인 생각을 알기 위한 목적으로 과학적 태도를 측정하는 면담법을 실시한다.’, ‘학생 각자의 과학에 대한 흥미나 태도를 파악하기 위해 자기 평가를 실시한다.’ 등과 같이 선행 연구에서 보고되지 않았던 새로운 유형의 평가 목적에 해당하는 사례를 제시했다. 이러한 평가 목적은 기존의 총괄평가나 형성평가와 달리 정의적 측면에서 학생들의 상태를 파악하는 것이 핵심이므로 정의적 평가라는 새로운 유형으로 분류했다.

과학 평가의 실태에 관한 초등교사와 예비초등교사의 응답을 평가의 방법 측면에서 분석한 결과는 Table 3과 같다. 초등교사는 측정 방식이 54.4%로 가장 많았고, 수행과 비형식 방식이 각각 22.3%와 23.3%였다. 선행 연구에서도 초등교사들은 시간 부족, 평가의 용이성과 객관성, 평가 과정이나 방법

Table 3. The methods of assessment shown in participants' responses

방법	빈도(%)	
	교사	예비교사
측정	107(54.4)	107(38.3)
선택형	60(30.5)	94(33.7)
서답형	47(23.9)	13(4.6)
수행	44(22.3)	131(47.0)
실험	11(5.6)	38(13.6)
보고서	20(10.2)	34(12.2)
논술형	2(1.0)	6(2.2)
발표	4(2.0)	5(1.8)
포트폴리오	6(3.0)	45(16.1)
토의	1(0.5)	3(1.1)
비형식	46(23.3)	41(14.7)
관찰	37(18.8)	23(8.2)
면담	4(2.0)	9(3.2)
동료	3(2.5)	6(2.2)
자기	2(1.0)	3(1.1)
계	197(100.0)	279(100.0)

에 대한 이해 부족 등으로 수행평가를 제대로 실시하지 못하는 경우가 많은 것으로 보고되었다(고민석 등, 2013). 중등교사를 대상으로 한 선행 연구(노태희 등, 2015)와 비교할 때, 이 연구에서는 수행 방식의 비율은 감소하고 비형식 방식의 비율이 증가했는데, 이는 수행 방식의 평가가 초등학생에게 부담으로 작용할 수 있으므로 초등교사들이 수행 방식의 평가 대신 비형식 방식의 평가를 선호하기 때문일 가능성이 있다.

비형식 평가 방식의 대부분을 차지한 것은 관찰 평가였는데, 관찰은 학습 과정에서 일어나는 학생들의 활동을 기록하고, 적절한 피드백을 제공함으로써 활발한 학습을 조장하고, 궁극적으로 학생의 성장을 도모할 수 있는 구성주의적 평가 방법이다. 그런데 관찰이 효과적인 평가 도구가 되기 위해서는 관찰의 의도를 분명히 하고, 구체적인 관찰 계획을 수립하며, 평가 목적에 근거한 관찰 결과 활용 방안을 수립해야 한다(남승인과 강영란, 1999). 이 연구에서 초등교사의 관찰 평가 비율이 중등교

사에 비해 높았지만, 이러한 결과가 곧 초등교사의 구성주의적 평가 전문성을 의미한다고 보기에는 어려움이 있다. 왜냐하면 체크리스트나 평가 기준 및 항목을 준비하여 실시하는 계획적인 관찰 평가 보다는 단순히 학생들을 관찰하여 평가한다는 사례가 많았기 때문이다. 초등교사들이 평가를 할 때 기존의 평가 자료를 그대로 사용하거나 부분적으로 재구성하는 경우가 대부분이라는 점(신정윤 등, 2016)을 고려한다면, 초등교사들이 사전에 평가 기준이나 항목을 꼼꼼하게 준비했을 가능성은 높지 않다고 보인다.

예비초등교사의 경우에는 수행 방식이 47.0%로 가장 많았고, 측정 방식 38.3%, 비형식 방식 14.7% 순이었다. 예비초등교사는 초등교사에 비해 수행 방식의 비율이 높았는데, 이러한 결과는 주로 포트폴리오 평가의 비율이 높아진 것에 기인한다. 중등과 달리 예비초등교사는 과학 이외에 미술이나 실과 등 포트폴리오가 강조되는 다른 교과교육 과목도 공부하기 때문에 포트폴리오 평가 방법에 대해 더 친숙할 가능성이 있다. 학생 활동의 모든 과정을 체계적으로 기록하고 누적하여 평가하는 포트폴리오 평가는 학생의 변화 과정을 종합적으로 파악할 수 있는 대표적인 구성주의적 평가 방법의 하나이므로(김찬중, 2012), 과학 수업에서 포트폴리오 평가를 활용하겠다는 응답이 많은 결과는 앞으로 과학 평가의 질이 개선될 가능성을 내포한다는 점에서 긍정적으로 볼 수 있을 것이다. 그러나 포트폴리오 평가로 분류된 응답을 평가의 목적 관점에서 분석한 결과, 45개의 사례 중 34개가 총괄평가

목적으로 분류되었다. 즉, 아래의 사례에서 볼 수 있듯이, 예비초등교사들은 보고서와 같이 일정기간 동안 진행한 학생 활동을 총괄적으로 평가하는 목적으로 포트폴리오 평가를 사용하는 경우가 많았다.

한 학기 동안 과학 수업을 통해 무엇을 배웠고 그에 따른 변화 과정을 살펴보기 위한 목적으로 총괄적으로 측정하는 포트폴리오를 제출한다.
(예비초등교사 D)

학생들이 그동안 학습 또는 실험해 온 과정과 결과물을 평가하기 위해 포트폴리오 평가를 실시한다.
(예비초등교사 E)

과학적 탐구 능력이 얼마나 발달해 왔는지 파악하기 위해 발달 과정을 측정하는 포트폴리오법을 실시한다.
(예비초등교사 F)

한 학기 동안 얼마나 열심히 수업에 임했는가 확인하는 목적으로 포트폴리오를 실시한다.
(예비초등교사 G)

2. 과학 평가 지향

구성주의적 평가의 실시를 기대할 수 있는 전형적인 과학 평가 상황인 실험, 토의, 자유탐구에 대한 초등교사와 예비초등교사의 응답에서 평가의 목적을 분석한 결과는 Table 4와 같다. 초등교사의 경우, 총괄평가 목적의 비율은 실험 상황에서 98.0%, 토의 상황에서 91.5%, 자유탐구 상황에서 90.0%였다. 예비초등교사의 응답은 실험 상황에서 96.7%,

Table 4. The frequencies of the purposes of assessment shown in participants' responses to the prototypical situations (%)

목적	실험		토의		자유탐구	
	교사	예비교사	교사	예비교사	교사	예비교사
총괄평가	94(98.0)	88(96.7)	86(91.5)	73(79.3)	90(90.0)	71(83.5)
학생 수준 평가	83(86.6)	83(91.2)	70(74.5)	59(64.1)	87(87.0)	65(76.5)
학생 선발	1(1.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
학습 부진아 확인	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
수업 참여도 측정	10(10.4)	5(5.5)	16(17.0)	14(15.2)	3(3.0)	6(7.0)
형성평가	2(2.0)	3(3.3)	8(8.5)	19(20.7)	9(9.0)	14(16.5)
학습 기회 제공	1(1.0)	1(1.1)	6(6.4)	12(13.0)	5(5.0)	9(10.6)
기술적 피드백 제공	1(1.0)	2(2.2)	2(2.1)	7(7.7)	4(4.0)	5(5.9)
계	96(100.0)	91(100.0)	94(100.0)	92(100.0)	99(100.0)	85(100.0)

토의 상황에서 79.3%, 자유탐구 상황에서 83.5%가 총괄평가 목적으로 나타났다. 구성주의적 평가를 실시할 수 있는 전형적인 평가 상황을 제시했음에도 대부분의 응답이 전통적인 총괄평가 목적으로 나타난 결과는 초등교사와 예비초등교사들의 과학 평가에 대한 인식이 구성주의적 평가를 지향하고 있지 못함을 의미한다. 한편 토의 상황에서는 총괄평가 중 수업 참여도 측정 목적의 비율이 상대적으로 높게 나타났는데, 이는 교사나 예비교사들이 토의 상황의 중요한 평가 목적으로 토의 참여 정도를 생각하고 있음을 보여준다.

초등교사와 예비초등교사가 지향하는 평가의 목적은 평가 상황에 따라 달라졌다. 토의나 자유탐구 상황에 비해 실험 상황에서 형성평가 목적의 평가가 매우 적었는데, 이는 학교에서 실험이 사용되는 방식과 관련된 것으로 보인다. 즉, 현재의 과학 수업에서는 학습한 지식을 확인하는 방식으로 과학 실험을 사용하는 경우가 많으므로, 실험 상황에서 학생들이 정확하게 이해하고 실험을 실행하는지에 대한 지향이 높게 나타날 수 있다.

중등교사를 대상으로 한 선행 연구(노태희 등, 2015)에 비해(실험: 13.9%, 토의: 33.6%, 자유탐구: 41.5%) 초등교사와 예비초등교사 모두 형성평가를 목적으로 하는 비율이 상당히 낮았다. 물론 구성주의적 평가를 알고 있는 교사들도 현실적인 제약으로 인해 실제 수업에서는 전통적 평가를 시행할 가능성이 있다(노태희 등, 2015; Thomas *et al.*, 2011). 그러나 전형적인 구성주의적 평가 상황에 대해서도 형성평가를 목적으로 하는 응답이 적은 것은 초등교사와 예비초등교사들의 구성주의적 평가 전문성이 부족할 가능성을 시사한다. 즉, 여러 과목을 가르쳐야 하는 초등교사는 중등교사에 비해 과학 교과에 대해 훈련받은 경험이 적으므로 평가 전문성도 부족할 가능성이 있다. 선행 연구에서도 초등교사들은 기준 참조 평가와 준거 참조 평가의 특징과 장단점에 대한 이해나 교수학습 과정에서 학생이 겪는 어려움을 진단하는 것과 같은 구성주의적 평가 전문성이 부족할 가능성을 지적한 바 있다(강훈식과 강석진, 2015). 학습에 대한 관점과 평가의 가치 그리고 평가의 목적에 대한 지식은 교사들이 갖추어야 할 평가 소양의 핵심 지식이므로(Abell & Siegel, 2011), 초등교사와 예비초등교사의 구성주의적 평가에 대한 이해를 높이기 위한 노력이 이루어져야

할 것이다.

실험, 토의, 자유탐구 상황에 대한 초등교사와 예비초등교사의 응답에서 평가의 방법을 분석한 결과는 Table 5와 같다. 초등교사의 경우, 측정 방식은 실험 상황에서 15.6%, 토의 상황에서 9.9%, 자유탐구 상황에서 8.1%로 과학 평가의 실태에 비해 그 비율이 매우 낮았다. 평가의 실태 분석에서 50% 이상이었던 측정 방식의 평가 방법이 전형적인 구성주의적 평가 상황에서는 큰 폭으로 감소했고, 대신 수행이나 비형식 방식의 평가 방법에 대한 지향이 크게 증가했다. 즉, 평가의 목적과 달리 평가의 방법 측면에서는 전형적인 구성주의적 평가 상황에서 다양한 방법의 평가를 시도하려는 경향이 나타났다.

실험과 토의 상황에서는 비형식 방식이 가장 많았고(실험: 57.3%, 토의: 76.9%), 하위 범주 중에서는 관찰 방법의 비율이 압도적이었다. 그러나 아래의 사례에서 드러나듯이, 관찰 방법으로 분류된 교사들의 응답은 실험 내용에 대한 이해 여부, 능숙하게 실험 기구를 다루는 기능, 적극적인 실험 활동 참여 등과 같이 대부분 총괄평가 목적이어서 구성주의적 평가라고 볼 수 없었다.

학생이 혼합물의 성질에 따라 여러 가지 실험 기구를 사용하여 실제로 분리할 수 있는지 평가하기 위해 직접 수행하게 하고 관찰 평가를 한다.

(초등교사 C)

여러 가지 실험 기구를 사용하여 혼합물을 잘 분리하는지 평가한다. 모듬별로 모듬 친구들과 협력하며 실험하는 태도도 평가한다. 여러 가지 실험 기구를 올바른 방법으로 사용하면서 혼합물의 분리를 잘하는지 수시로 확인하면서 평가한다.

(초등교사 D)

토의와 자유탐구 상황에서는 비형식 방식 중 동료 평가와 자기 평가가 일부 있었다. 동료들 사이의 상호작용을 바탕으로 제공되는 피드백이 학습을 지원하는 동료 평가나 자신의 이해 수준이나 개선점에 대해 스스로 반성하는 자기 평가는 학생에 관한 정보를 제공하므로 성적의 산출보다 학습의 지원을 목적으로 하는 구성주의적 평가 방법에 해당한다(김송자와 최창우, 2006). 그러나 노태희 등(2015)은 교사들이 수업 참여도를 평가하기 위한

Table 5. The frequencies of the methods of assessment shown in participants' responses to the prototypical situations (%)

방법	실험		토의		자유탐구	
	교사	예비교사	교사	예비교사	교사	예비교사
측정	15(15.6)	15(16.7)	9(9.9)	1(1.1)	8(8.1)	0(0.0)
선택형	5(5.2)	14(15.6)	2(2.2)	1(1.1)	2(2.0)	0(0.0)
서답형	10(10.4)	1(1.1)	7(7.7)	0(0.0)	6(6.1)	0(0.0)
수행	26(27.1)	46(51.0)	12(13.2)	30(34.5)	54(54.6)	72(84.7)
실험	5(5.2)	21(23.3)	0(0.0)	0(0.0)	0(0.0)	1(1.2)
보고서	16(16.7)	22(24.4)	6(6.6)	4(4.6)	27(27.3)	15(17.6)
논술형	0(0.0)	0(0.0)	2(2.2)	2(2.3)	0(0.0)	0(0.0)
발표	5(5.2)	2(2.2)	3(3.3)	1(1.1)	5(5.1)	5(5.9)
포트폴리오	0(0.0)	1(1.1)	0(0.0)	5(5.8)	21(21.2)	51(60.0)
토의	0(0.0)	0(0.0)	1(1.1)	18(20.7)	1(1.0)	0(0.0)
비형식	55(57.3)	29(32.3)	70(76.9)	56(64.4)	37(37.3)	13(15.3)
관찰	51(53.1)	27(30.1)	52(57.1)	41(47.2)	23(23.2)	8(9.4)
면담	0(0.0)	0(0.0)	0(0.0)	1(1.1)	4(4.0)	0(0.0)
동료	2(2.1)	0(0.0)	13(14.3)	11(12.7)	8(8.1)	3(3.5)
자기	2(2.1)	2(2.2)	5(5.5)	3(3.4)	2(2.0)	2(2.4)
계	96(100.0)	90(100.0)	91(100.0)	87(100.0)	99(100.0)	85(100.0)

목적으로 동료 평가와 자기 평가를 활용하는 경우가 있음을 보고한바 있다. 이 연구에서도 동료 평가나 자기 평가에 해당하는 대부분의 응답이 해결책의 타당성, 학습 태도, 활동에의 기여도 등에 대한 총괄평가를 지향하고 있었다. 이러한 경향을 보여주는 대표적인 사례는 다음과 같다.

토의 수업에서 지구 온난화 문제 해결 방안을 잘 마련하는지 알아보기 위해 논리적이고 타당한 주장을 펼치는지 상호 평가하게 한다.

(초등교사 E)

지구 온난화에 대한 사전 준비 자료의 준비, 합리적이고 논리적인 해결 방안을 제시하는지, 토론 후에 상호 평가를 실시한다.

(초등교사 F)

자유탐구를 소집단별로 진행하는 과정에서 활동에 충실하게 참여한 학생들을 파악하기 위해서 상호 평가를 진행한다. 평가 방법은 학습 태도, 활동에 대한 기여도, 각자 맡은 역할 등에 대한 질문지를 작성하여 평가한다.

(초등교사 G)

소집단 활동에서 자기 역할을 성실히 수행하고 꾸준히 탐구 활동을 했는지에 대해 학생 상호 평가와 자기 평가를 실시한다.

(초등교사 H)

실험 상황에서 수행 방식의 비율은 27.1%였고, 그 중 보고서 평가(16.7%)의 비율이 높았다. 토의 상황에서 수행 방식의 비율은 13.2%였고, 그 중 보고서 평가가 6.6%로 가장 많았다. 자유탐구 상황에서는 수행 방식의 비율이 54.6%로 가장 높았는데, 그중에서도 보고서(27.3%)와 포트폴리오(21.2%) 평가가 많았다. 보고서 평가는 과학 교사들이 평가의 타당도보다 평가 결과의 신뢰도를 중요하게 생각하기 때문에 많이 사용하는 수행평가 방법으로 알려져 있다(신정윤과 양일호, 2017). 비형식 방식과 마찬가지로 수행 방식에 해당하는 응답도 목적 측면에서는 대부분 전통적인 총괄평가를 지향하고 있어, 평가 목적과 방법의 불일치가 크게 나타났다. 즉, 이 연구에서는 구성주의적 평가 방법으로 알려진 평가 방법을 전통적인 목적으로 활용하는 경우가 많았는데, 이는 교사들이 구성주의적 평가 방법

에 대한 지식은 있으나, 구성주의적 평가가 무엇인지에 대한 본질적인 이해는 부족하기 때문에 해석할 수 있다.

구성주의적 평가를 활용할 수 있는 전형적인 상황을 제시한 경우, 예비초등교사들도 측정 방식의 비율이 많이 낮아지고(실험: 16.7%, 토의: 1.1%, 자유탐구: 0.0%), 수행이나 비형식 방식의 비율이 큰 폭으로 증가했다. 특히, 실험 상황(51.0%)과 자유탐구 상황(84.7%)에서 초등교사에 비해 수행 방식의 평가를 사용하겠다는 응답이 높았다. 그러나 평가의 목적 측면에서는 대부분의 응답이 전통적인 목적의 평가인 것으로 나타나, 구성주의적 평가에 대한 지향 측면에서는 초등교사들의 응답과 유의미한 차이를 발견할 수 없었다.

IV. 결론 및 제언

이 연구에서는 초등학교에서 이루어지는 과학 평가의 실태를 파악하기 위해 초등교사와 예비초등교사의 평가 사례를 조사하여 평가의 목적과 방법 측면에서 분석했다. 또한 초등교사와 예비초등교사들의 구성주의적 과학 평가에 대한 지향을 조사하기 위해 실험, 토의, 자유탐구 등 전형적인 구성주의적 평가 상황에 대한 평가 계획을 평가의 목적과 방법 측면에서 분석했다.

연구 결과, 초등교사와 예비초등교사가 제시한 평가 사례는 대부분 총괄평가 목적인 것으로 나타났다. 이러한 결과는 교육 현장에서 진정한 의미에서의 형성평가가 부족하다는 선행 연구(김난옥과 손원숙, 2019; 김희경 등, 2014)와 유사하다. 그러나 많은 교사들이 수준 높은 피드백을 제공하는 형성평가를 실시한다고 인식한다는 선행 연구(박지현 등, 2017; 정은영과 최원호, 2014)의 결과를 고려하면, 교사들이 생각하는 형성평가는 진정한 의미의 형성평가와 상당한 차이가 있는 것으로 보인다. 교사들의 생각과 달리 학생들은 피드백이 좀 더 구체적이기를 바라고 있고(박지현 등, 2017) 많은 교사들이 형성평가와 총괄평가를 단순히 개념적으로 구분하는 정도의 수준에 불과하다는 주장(강대중과 염시창, 2014)을 고려한다면, 형성평가를 후속 교수 학습을 위한 적절한 피드백과 학습 경험을 제공하는 평가로 이해하고 실천하는 교사는 많지 않은 것으로 보인다. 즉, 교사들이 형성평가라는 용어를 자

주 사용하지만 학생들의 이해도를 점검하는 도구의 의미 정도로 이해하는 것이 우리의 현실인 것이다. 수업과 평가를 연계한 평가의 형성적 기능을 강조하는 학습을 위한 평가, 그리고 2015 개정 교육과정에서 강조하는 교수-학습-평가의 연계성과 학생의 학습과 성장을 지원하는 과정 중심 평가를 구현하기 위해서는 전통적인 총괄평가에서 벗어나 형성평가의 활용이 확대되어야 할 것이다.

초등교사의 과학 평가 실태를 평가 방법 측면에서 분석한 결과, 중등과학교사를 대상으로 한 선행 연구(노태희 등, 2015)에 비해 비형식 방식, 특히 관찰 평가 방법의 비율이 크게 증가했다. 관찰 평가는 학생의 활동을 기록하고, 이를 바탕으로 피드백을 제공하는 대표적인 구성주의적 평가 방법으로 알려져 있다(백순근, 2000). 그러나 초등교사들이 제시한 평가 사례가 거의 모두 총괄평가 목적이었던 점을 고려한다면, 초등교사들이 구성주의적 평가 방법의 하나로서 관찰 평가를 사용하기보다는 단순히 학생의 성취 여부를 암묵적으로 점검한 경험을 관찰 평가로 인식하고 있을 가능성이 높다. 초등학교 과학 수업에서는 교사들이 추가적인 시간을 투입하지 않아도 되는 비형식적인 관찰 평가 방법을 선호할 가능성이 높다. 그러나 관찰이 효과적인 구성주의적 평가 도구가 되기 위해서는 명확한 의도에 근거한 체계적인 관찰 계획과 관찰 결과의 활용 방안 수립이 선행되어야 하므로(남승인과 강영란, 1999), 관찰 평가의 목적과 방법에 대한 초등교사들의 전문성을 개발할 수 있는 방안을 모색해야 한다. 한편 예비초등교사는 선행 연구(노태희 등, 2015)의 중등 과학교사나 이 연구에서의 초등교사에 비해 구성주의적 평가 방법인 포트폴리오 평가의 비율이 상대적으로 높았다. 그러나 포트폴리오 평가를 활용하겠다는 응답 중 형성평가 목적을 지닌 사례는 소수인 것으로 나타나, 과학교육 현장에서 포트폴리오 평가가 효과적으로 정착되기 위해서는 과학 평가의 목적과 기능에 대한 예비교사들의 전문성 함양에 노력을 기울여야 할 것이다.

실험, 토의, 자유탐구와 같이 구성주의적 평가를 사용할 수 있는 전형적인 평가 상황에 대해서도 초등교사와 예비초등교사들의 응답은 대부분 총괄평가 목적으로 나타나, 토의나 자유탐구에서 형성평가 목적의 평가 비율이 상당히 높았던 중등 과학교사(노태희 등, 2015)와 다른 경향을 보였다. 또한 평

가 방법 측면에서도 구성주의적 평가로 사용할 수 있는 평가 방법들이 전통적인 목적으로 사용되는 경우가 많았다. 이러한 결과는 이론적으로는 다양한 평가 방법을 알고 있지만, 실제로 교실에서 구성주의적으로 평가를 구현하기에는 교사들의 이해가 부족하고 익숙하지 않음(노태희 등, 2015; Abell & Siegel, 2011)을 시사한다. 여러 과목을 가르쳐야 하는 초등교사는 중등교사에 비해 과학 평가에 대해 체계적인 교육을 받은 경험이 적어 전문성이 부족할 가능성이 높으므로(강훈식과 강석진, 2015; 김난옥, 손원숙, 2019), 교육 현장에서 구성주의적 평가를 구현하기 위해서는 교사 연수나 교사 양성 프로그램에서 과학 평가 교육을 강화할 수 있는 방안을 모색해야 할 것이다.

우리나라의 경우, 다른 나라에 비해 평가 관련 프로그램이 예비교사나 현직교사에게 평균 이하로 제공되는 실정이고(김난옥과 손원숙, 2019), 평가 연수의 기간과 대상이 점차 축소될 뿐 아니라, 연수의 내용과 질도 평가 관련 자료를 보급하는 수준이며(김진규, 2009), 2015 개정 교육과정에서 강조되는 수행평가의 연수 경험이 없는 교사도 62%에 이르는 것으로 보고되었다(신정운 등, 2016). 예비초등교사의 경우에도 평가 능력을 함양하기 위한 교육은 소홀히 다루어져 교육평가 과목을 대학에서 기껏해야 한 과목 정도 이수하는 정도이다(박정, 2013). 평가 목적에 적절한 평가 방법을 선정, 실시하고 타당한 의사결정을 내리기 위해서는 무엇보다 교사의 평가 역량이 중요하다(김성훈 등, 2018). 평가 연수 여부에 따라 교사들의 형성평가에 대한 인식 차이가 있고(강대중과 염시창, 2014), 평가 강좌가 예비교사의 평가 인식 개선에 도움이 되며(박정, 2013), 많은 교사들이 수준 높은 평가 전문성 개발을 요구한다(김희경 등, 2014; 박지현 등, 2017)는 선행 연구를 고려할 때, 최신 평가 동향을 학습하고 부족한 평가 역량을 함양할 기회를 제공할 수 있는 교사 연수 프로그램이나 예비교사교육 프로그램을 확대할 필요성이 있다.

한편 김진규(2009)는 현행 평가 연수의 문제점으로 교사의 개별적인 역량 강화에 치중하고 현장 적용 문제를 교사의 책임으로 전가하며, 교사 간 협력과 정보 공유의 문화 형성에 둔감하다는 점을 지적했다. 강훈식과 강석진(2015)도 초등교사들의 평가 전문성을 함양하기 위해 이론적인 학습 기회 제

공에 덧붙여 교사들이 평가 도구나 문항을 직접 개발하고, 다른 교사나 전문가의 피드백을 받아 개선하는 경험을 가질 수 있는 방안이 필요함을 제안했다. 특히 우리나라 교사들은 동료 교사와의 교류를 통해 자신의 평가 능력에 자신감을 가지는 것으로 나타났으므로(손원숙 등, 2015), 단순히 지식을 전달하는 이론적인 강의보다는 교사공동체 내에서 동료 교사 간의 협력을 통해 구체적인 평가 상황에서 어떤 평가 방법을 사용할 것인가에 대한 실천적인 논의를 할 수 있는 기회를 제공하는 시도가 필요하다.

참고문헌

- 강대중, 염시창(2014). 초등학교 교사들의 형성평가에 대한 인식 및 실태 분석. *학습자중심교과교육연구*, 14(2), 27-43.
- 강훈식, 강석진(2015). 초등학교 교사의 과학 평가 전문성에 대한 자가진단 내용 분석. *초등과학교육*, 34(2), 153-163.
- 고민석, 김은애, 허진미, 양일호(2013). 초등학교 교사의 탐구에 대한 신념과 과학과 수행평가의 실제. *대한지구과학교육학회지*, 6(2), 124-135.
- 김난옥, 손원숙(2019). 우리나라 학생평가 실태 연구: PISA 2015 참여국과의 국제비교를 중심으로. *교육과정평가연구*, 22(3), 173-198.
- 김난옥, 박민애, 이빛나, 손원숙(2018). 교사의 특성과 피드백이 초등학생의 정의적, 인지적 성취에 미치는 영향. *교육과정평가연구*, 21(2), 129-151.
- 김성훈, 김신영, 김재철, 반재천, 백순근, 서민원(2018). 예비교사를 위한 교육평가. 서울: 학지사.
- 김송자, 최창우(2006). 초등학교 수학과 수행평가 실천에 관한 연구: 자기평가, 동료평가, 관찰평가를 중심으로. *한국초등수학교육학회지*, 10(1), 67-87.
- 김인숙, 박지현, 임은영(2018). 형성평가시스템 활용을 통한 단위학교의 형성평가 지원 방안. *교육과정평가연구*, 21(3), 181-199.
- 김진규(2009). 교사 학습공동체 활용 평가연수의 실천 전략. *교육평가연구*, 22(4), 939-959.
- 김진규(2013). 형성평가 101가지 기법. 서울: 교육과학사.
- 김찬중(2012). 포트폴리오 교수-학습 및 평가. 서울: 교육과학사.
- 김희경, 박종임, 정연준, 박상욱, 김창환, 이채희, 최재화(2014). 맞춤형 교육 지원을 위한 형성평가 체제 도입 (I): 온·오프라인 형성평가 시스템 설계. 연구보고 RRE 2014-9. 한국교육과정평가원.
- 남승인, 강영란(1999). 관찰을 통한 수행능력 평가 방안.

- 수학교육논문집, 8, 65-76.
- 노태희, 이재원, 강석진, 강훈식(2015). 중등 과학교사의 과학 평가 실태와 지향. *한국과학교육학회지*, 35(4), 725-733.
- 민회정(2012). 과학교사의 학생평가 실태분석 및 교사연수를 통한 평가전문성 모델 개발. *한국교원대학교 박사학위논문*.
- 박정(2013). 초등 예비교사의 학생평가 전문성 분석. *교육평가연구*, 26(1), 15-33.
- 박지현, 조지민, 홍미영, 정혜경(2017). 학교수준 학생평가 운영 실태 및 인식 조사. *교육평가연구*, 30(1), 29-52.
- 백순근(2000). 수행평가의 원리. 서울: 교육과학사.
- 손원숙, 신이나, 배주현, 박정(2015). 교사의 내·외적 요인과 교실평가 간의 구조적 관계: 한국, 싱가포르 및 핀란드 비교. *교육평가연구*, 28(2), 601-622.
- 성태제(2013). 2020 한국 초·중등교육의 방향과 과제. 서울: 학지사.
- 성태제, 임현정(2014). 형성평가의 재인식에 따른 교사와 학교교육의 변화를 위한 제언. *교육평가연구*, 27(3), 597-615.
- 신정윤, 양일호(2017). 과학과 수행평가의 문제점과 개선 방향. *예술인문사회융합멀티미디어논문지*, 7(1), 873-883.
- 신정윤, 유선아, 양일호(2016). 초등학교 교사들의 과학과 수행평가 실시 현황 및 요구 분석. *청람과학교육연구논총*, 22(2), 18-29.
- 정은영, 최원호(2014). 우리나라 초·중·고등학교 과학과의 평가 실태. *과학교육연구지*, 38(1), 168-181.
- Abell, S. K. & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In Corrigan, D., Dillon, J. & Gunstone, R. (Eds.), *The professional knowledge base of science teaching* (pp. 205-221). London, UK: Springer.
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L. & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117-140.
- Butler, S. M. & McMunn, N. D. (2006). *A teacher's guide to classroom assessment: Understanding and using assessment to improve student learning*. San Francisco: Jossey-Bass.
- Davis, D. S. & Neitzel, C. (2011). A self-regulated learning perspective on middle grades classroom assessment. *The Journal of Educational Research*, 104(3), 202-215.
- Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousands Oaks: Corwin Press.
- Falk, A. (2012). Teachers learning from professional development in elementary science: Reciprocal relations between formative assessment and pedagogical content knowledge. *Science Education*, 96(2), 265-290.
- Istance, D. & Dumont, H. (2010). Future directions for learning environments in the 21st century (pp. 317-388). In Dumont, H., Istance, D. & Benavides, F. (Eds.), *The nature of learning: Using research to inspire practice*. Paris, France: OECD Publishing.
- McMillan, J. H. (2015). *교실평가의 원리와 실제: 기준참조 수업과의 연계*. 손원숙, 박정, 강성우, 박찬호, 김경희 역, 서울: 교육과학사.
- Siegel, M. & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, 22(4), 371-391.
- Thomas, L., Deaudelin, C., Desjardins, J. & Dezutter, O. (2011). Elementary teachers' formative evaluation practices in an era of curricular reform in Quebec, Canada. *Assessment in Education: Principles, Policy & Practice*, 18(4), 381-398.
- Wang, J.-R., Kao, H.-L. & Lin, S.-W. (2010). Preservice teachers' initial conceptions about assessment of science learning: The coherence with their views of learning science. *Teaching and Teacher Education*, 26(3), 522-529.

이혜민, 전주교육대학교 대학원 학생(Lee, Haemin; Graduate student, Jeonju National University of Education)

† 강석진, 전주교육대학교 교수(Kang, Sukjin; Professor, Jeonju National University of Education)