

교육과정평가연구

The Journal of Curriculum and Evaluation

2021, Vol. 24, No. 1, pp. 231~251

DOI: <https://doi.org/10.29221/jce.2021.24.1.231>

과학교사의 수행평가에 대한 평가 전문성 탐색¹⁾

김성기 (한국교육과정평가원 부연구위원)*

민희정 (연남중학교 교사)

하민수 (강원대학교 교수)

백성혜 (한국교원대학교 교수)**

요약

본 연구는 학생의 성장을 지원하는 형성평가와 관련하여 수행평가에 대한 과학교사의 평가 전문성을 탐색하는데 목적이 있다. 이를 위해 서울의 중학교 과학교사 4명을 대상으로 한 학기 동안 평가 장면을 포함한 수업을 관찰하고, 면담을 실시하여 과학교사의 수행평가와 관련한 평가 전문성을 탐색하였다. 선행연구를 통해 평가 전문성과 관련한 분석을 ‘교육 목표’, ‘전략’, ‘평가 목적과 방법’, ‘평정’, ‘피드백’의 범주로 나누어 진행하였다. 각 범주의 특성이 충분히 관찰되면 높은 수준으로, 특성이 나타나지 않는 경우 낮은 수준으로 판단하였다. 연구결과 5가지 범주에서 과학교사들은 대부분 낮은 수준으로 판정되었다. 또한, ‘교육 목표’, ‘전략’, ‘평가 목적과 방법’과 같이 주로 인지적 측면에 대한 평가 전문성이 ‘평정’이나 ‘피드백’과 같이 주로 실천적 측면의 평가 전문성에 비해 상대적으로 높은 수준을 보였다. 이는 과학교사들이 실천적 역량으로 전문성이 확대될 때 그 수준이 낮아짐을 보이는 것으로 교사가 알고 있는 지식을 실천까지 확대할 수 있는 실천적 역량의 평가 전문성 함양과 관련된 연구가 요구됨을 알 수 있다. 특히, ‘피드백’ 측면에서 모든 과학교사의 평가 전문성이 낮게 판정되었다. 이는 평가를 통해 학생의 학습을 지원하기 보다는 학생의 성적을 평정하는 측면으로 수행평가가 진행되고 있음을 보여준 것으로 이와 관련한 교사 연수가 절실함을 알 수 있다. 이 연구에서 탐색한 결과를 종합하여 교사의 평가 전문성을 향상하기 위한 방안을 제안하였다.

주제어 : 평가 전문성, 수행평가, 형성평가

1) 본 연구는 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019S1A5C2A04081191/NRF-2019R1A2B5B01069840)

* 제1저자, mcarey2000@kice.re.kr

** 교신저자, shpaik@knue.ac.kr

I. 서 론

평가는 목적에 따라 형성평가와 총괄평가로 나누어질 수 있다(Dixson & Worrell, 2016). 총괄평가란 교수·학습이 끝난 다음 교수 목표 달성, 성취 여부를 종합적으로 판정하는 평가 형태로 평가 결과가 교육과정이나 교육방법 등 장기적인 교육의 질 관리에 이용될 수 있으며, 집단 간의 성취도를 상호 비교할 수 있는 정보를 제공한다는 점에서 유용하다. 하지만 학습이 모두 완료된 시점에서 평가가 진행됨에 따라 학생의 성장을 도모하기 보다는 학습이 완료된 후 학습의 결과적 측면에 초점을 두고 있다. 이와 달리, 형성평가는 수업이 진행되는 상태에서 현재 진행되는 교수활동이 계획대로 진행되고 있는지를 확인하는 평가로, 실시간 학생의 학습을 점검한다는 의미가 있다. 이러한 형성평가는 수업 중 다양한 학생의 활동을 포함하며, 학습에서 그 역할이 점점 더 중요해지고 있다(Heritage, 2010). 형성평가를 능숙하게 사용함으로써 교사는 학생들이 학습하고 성취하도록 동기를 부여 할 수 있다(Darling-Hammond et al., 2013; Lyon, 2013a). 그렇기 때문에 학생의 학습 지원과 성장을 위한 평가로 형성평가가 강조되고 있다.

Black & Wiliam(1998)은 학생의 학습에 대한 효과를 증진할 수 있는 최고의 방법이 형성평가라고 주장하였다. 이러한 의미에서 교육부는 지난 20년 동안 형성평가를 강조해왔다(교육부, 1998, 2017). 김현재, 송민영(1997)은 수행평가란 학생의 실제 수행의 과정과 결과를 평가한다고 하였다. 신정윤, 양일호(2017)는 수행평가는 학생의 성장을 위해 학습의 현재 상태를 진단하고 이에 맞는 피드백을 제공하는 과정을 포함한다고 하였다. 특히, 수행평가는 수업시간에 진행되며, 모둠이나 개인별로 진행됨에 따라 수행에 대한 구체적인 피드백을 학생 개인에게 제공하기 때문에 형성평가의 형태로 운영될 수 있다. 따라서 수행평가는 과정중심평가에 적합한 실천 방법이기도하며(노은희, 서민원, 2016), 따라서 학생의 성장을 위한 평가로써 수행평가의 가치는 크다. 교육부는 평가에 대한 패러다임을 학습을 지원하는 평가로 변화하기 위해 다양한 교사 연수 프로그램을 운영하고 있지만, 여전히 많은 교사들이 수행평가를 형성평가의 목적에 부합하게 실시하기 보다는 총괄평가의 목적으로 활용하고 있는 실정이다(노태희 외, 2015). 이러한 현상은 비단 우리나라만의 일이 아니며, 여러 나라에서 많은 교사들이 교실에서 형성평가를 실행하는데 어려움을 겪고 있다(Panadero et al., 2014; Puad & Ashton, 2020). 우리나라를 포함하여 많은 나라의 교사들은 객관성에 대한 논쟁을 피하기 위해 단순한 지식을 평가하는 것을 선호하는 것으로 나타났다(김성기 외, 2015; Heritage, 2010; Stiggins & Duke, 2008; Volante & Beckett, 2011). 이 현상에 대한 주요한 이유는 바로 교사의 부족한 평가 전문성 때문이다(Bayat & Rezaei, 2015; DeLuca et al., 2016).

기존의 평가 전문성과 관련된 연구는 지필평가 중심의 평가 전문성을 분석이거나(김성기 외, 2015; 임천택, 2010; Kim & Paik, 2016), 성취기준에 근거한 평가 전문성 분석(김석우 외, 2017), 학교에 필요한 교사 전문성 탐색(곽영순, 2015; 박휴용, 2019) 및 평가에 대한 교사의 신념(김성기, 민희정, 백성혜, 2020) 측면에서 수행되었으나, 학생의 학습 지원을 위한 수행평가

중심의 연구는 활발하지 못한 실정이다. 특히, 노태희 외(2015)의 연구에서 알 수 있듯이 교사들은 학생의 학습을 지원하고 성장을 도모하는 평가에 어려움을 갖고 있다. 과학교사의 평가에 어려움을 해소하고, 교사들이 평가에 전문성을 갖춘 전문가가 되는데 도움을 주기 위해서는 먼저 이와 관련한 연구가 선행되어야 한다. 그러한 면에서 수행평가에 대한 과학교사의 평가 전문성을 탐색하는 것은 평가 전문성과 관련된 구체적 범주별로 그 수준에 대한 실태를 보여준다는 점에서 의미가 있다. 이러한 탐색은 구체적인 교사의 평가 전문성 향상을 위한 전략으로 활용될 수 있다는 점에서도 가치가 있다.

본 연구에서는 수행평가와 관련하여 평가 전문성을 탐색하기 위해 중학교 과학교사의 평가 전문성을 5가지 측면에서 심층적으로 분석하고자 한다. 수행평가에 대한 과학교사의 평가 전문성 탐색을 통해 보다 구체적으로 학생의 성장을 지원하는 평가를 위한 평가 전문성 신장 방안을 제안하고자 한다.

II. 연구 방법

1. 연구 대상

이 연구의 목적은 중학교 과학교사들의 평가 전문성을 탐구하는 것이었다. 이러한 탐구는 대규모 양적 연구로 진행하기에는 심도 있는 탐색이 어렵기 때문에(김성기, 민희정, 백성혜, 2020) 질적 연구방법을 선택하였다.

연구 목적을 충족하기 위해 중학교가 적절하다고 판단하였다. 그 이유는 고등학교에서 시행되는 수행평가의 경우 대학입시와 직접적인 관련이 있어 교사의 평가 소양과 관련하여 자신의 전문성을 드러내기가 어렵고, 중학교에 비해 다소 틀에 맞춰진 평가가 이루어지기 때문에 교사의 평가 전문성을 탐색하기에 적절하지 않다고 판단되었기 때문이다. 본 연구에서 표집된 학교는 서울에 위치한 중학교로 총 28개 학급의 규모를 가진다. 또한, 이 학교의 경우 본 연구의 연구자 중 한명이 재직하고 있는 학교이기 때문에 교사들의 언어나 행동을 이해하는데 필요한 학교의 문화적 요소를 잘 알고 있어 그들의 언어와 행동을 이해하는데 도움을 줄 수 있으며, 이 학교의 포괄적 데이터를 얻는데도 용이하였다.

이 학교는 6 명의 과학교사가 있었고, 두 명의 교사가 각각 같은 학년을 담당하고 있었다. 이 중 이 학교에 재직한 연구자에 해당하지 않은 학년을 담당하는 4명의 과학교사(1 학년, 3학년)가 연구에 참여하였다. 이 학교의 교장은 연구 목적을 이해하고 연구자들이 교실 활동을 관찰하고 교사와의 면담을 할 수 있도록 허락하였으며, 연구에 참여한 4명의 교사 모두 연구 참여에 동의하였으며 평가와 관련된 장면에 연구자가 관찰하는 것에 동의하였다. 이렇게 표집된 2개의 공동체에 대한 교사의 특성은 <표 1>과 같다.

〈표 1〉 연구 대상

특성	공동체1 (중학교 1학년)		공동체2 (중학교 3학년)	
	교사A	교사B	교사C	교사D
성별	남자	여자	남자	여자
연령	30대 후반	40대 초반	50대 초반	40대 초반
전공	물리	물리	물리	지구과학
교육 경력	7년	9년	25년	14년

2. 자료 수집

과학교사의 평가 전문성을 포괄적으로 탐색하기 위해 한 학기 동안 다양한 유형의 자료(교실 수업 녹화, 교실 활동을 기록한 노트, 교사의 수업 준비 자료, 교사와 학생 인터뷰 등)를 수집하였다. 이와 같은 다양한 유형의 자료는 교사의 심도있는 평가 전문성에 대한 탐색을 도와주기도 하였으며, 연구자의 질적 해석을 타당화하는데 활용되기도 하였다. 자료 수집을 수행하기 위해 학교 관리자 및 교사의 동의를 얻었으며, 연구 윤리 지침을 준수하였다.

관찰 자료는 크게 2가지 유형으로 나누어 수집되었다. 하나는 일반 수업에 대한 관찰이며, 다른 하나는 수행평가와 관련된 수업 관찰이다. 모든 수업 관찰은 녹화되었다. 일반 수업 관찰 녹화는 교사 당 총 10~15개의 수업이 녹화되었다(교사 A와 B는 10개, C는 11개, D는 15개). 수행 평가와 관련한 수업 녹화는 수행평가 계획대로 4명의 교사 모두 1회만 녹화되었다.

교사에 대한 심층 면담은 교사별로 3회씩 실시되었으며, 면담의 길이는 1회당 50~90분정도 소요되었다. 모든 면담은 교사 개별로 수행되었으며, 면담 내용은 휴대 전화를 이용하여 녹음하였으며, 면담 후 모든 녹음 자료는 전사하였다. 교사의 수행평가 장면을 관찰하기 이전에 먼저 1차 면담을 실시하였다. 면담 질문으로 교사의 교육 철학, 과학교육의 목표, 교육과정 등 교육에 대한 일반적인 내용을 포함하였다. 2차 면담은 수행평가가 진행된 이후 실시되었다. 면담의 초점은 교사가 실시한 수행평가이었다. 면담에서 평가활동의 목적과 방법, 평정 방법, 피드백 등에 대한 질문을 하였다. 1, 2차 면담 분석 및 관찰 평가를 토대로 해석한 내용에 기반하여 3차 면담을 진행하였다.

학생 면담은 총 16명의 학생을 대상으로 반구조화된 면담을 진행하였으며, 20~30분정도가 소요되었다. 4명의 학생이 하나의 조가 되어 면담을 하였으며, 점심시간 또는 방과후 시간을 활용하여 면담을 진행하였다. 수업시간에 연구자에 의해 관찰된 교수의 수업이나 평가에 대해 학생들의 인식을 보다 구체화되도록 질문하였다.

3. 자료 분석

면담을 통해 수집된 평가와 관련된 정보, 연구자 중 한 명이 기록한 교실 관찰 노트, 수업 녹화 자료를 삼각 측량을 통해 반복적으로 비교하였다. 이 데이터는 수행평가에 대한 교사의 평가 전문성을 판단하는데 활용되었다.

교사의 평가 전문성과 관련된 요소는 연구자마다 다양하게 보고되었다. 김성기 외(2015)와 Lyon(2013b)은 적절한 평가 방법의 선정 능력, 평가 도구 개발 능력, 성적 부여 능력, 결과 활용 및 의사소통 능력, 평가의 윤리성 인식 능력을 평가 전문성에 요구된 요소로 보고하였다. JCESS(2015)는 평가 실천을 강조하면서 평가의 시행과 피드백과 관련된 부분이 평가 전문성에 강조되어야 한다고 하였다. Brookhart(2011)은 평가 전문성에 대해 평가에 대한 지식 기반에 평가가 시행되는 것이라 하였다. Abell & Siegel(2011)과 Gearhart 외(2006)는 교육 목표를 이해하고 이를 위한 다양한 교수 전략 활용의 중요성을 강조하였으며, Xu & Brown(2016)은 실천의 영역으로 평가 전문성이 확대되기 위해서는 평가와 관련된 지식 기반에 평정과 피드백 등과 같은 실천이 결합되어야 한다고 하였다. Hattie & Timperley(2007)은 피드백의 효능을 강조하면서 평가 전문성에 피드백과 관련된 요소를 강조하였다. 종합하면 평가 전문성은 평가가 진행되는 특정한 시점에 요구되는 역량이라기보다는 평가를 수행하기 위해 교과와 목표를 이해하고 이를 위한 전략을 수립할 수 있어야 하며, 여러 평가의 목적을 이해하고 이에 따른 방법을 이해할 수 있어야 한다. 또한, 실제 평가 시점에서 학생을 평정하며, 평정 중 또는 이후에 상황에 맞는 피드백을 제공해야 한다. 이를 연구자는 <표 2>와 같이 5가지 범주로 재구성하였다.

각 범주에서 평가 전문성의 정도를 문헌에서 제안된 특성에 따라 높은 수준과 낮은 수준으로 구분하였다. 물론, 각 범주에 대해 중간 수준 설정도 가능하지만, 본 연구에서 명확하게 합의된 중간 수준의 판정 기준에 대한 진술이 다소 어려운 측면이 있었다. 또한, 범주별로 평가 전문성 경향을 비교하기에는 이분법적 수준 판정이 용이하다고 판단되어 2개 수준으로만 구분하였다. 각 범주의 특성이 충분히 관찰되면 높은 수준으로, 특성이 나타나지 않는 경우 낮은 수준으로 판단했다. 따라서 본 연구에서 평가한 과학교사의 평가 전문성의 정도는 정량적인 절대 판정이라기보다는 정성적 상대적 정도를 나타낸다. 각 범주에 대한 수준은 <표 2>와 같다.

〈표 2〉 범주별 수준 판정 기준

범주	수준	판정 기준
교육 목표	높은 수준(High)	단순한 개념 학습을 뛰어 넘은 학습을 추구한 경우
	낮은 수준(Low)	단순한 개념 학습을 목표로 하는 경우
전략	높은 수준(High)	학생 중심의 수업을 활용하기 위해 다양한 전략을 활용한 경우
	낮은 수준(Low)	교사 중심의 수업 전략을 활용한 경우
평가 목적과 방법	높은 수준(High)	교사가 수행평가의 형성적 측면을 인식하며, 이에 따라 다양한 방법을 고려한 경우
	낮은 수준(Low)	수행평가를 총괄적 특성으로 인식하여, 특정한 하나의 방법만을 활용한 경우
평정	높은 수준(High)	평정 과정에 학생과의 상호작용이나 학생들의 의견을 고려한 경우
	낮은 수준(Low)	교수 중심으로 일방적인 평정을 실시한 경우
피드백	높은 수준(High)	평가가 학습을 지원하기 위해 과정과 결과에 대한 피드백이 적절하게 제공된 경우
	낮은 수준(Low)	학생에게 평가에 대한 피드백을 제공하지 않는 경우

5가지 범주에 따라 과학 교육 박사 3인이 수집한 자료를 독립적으로 해석하였다. 이때 해석은 각 범주별 과학교사의 평가전문성에 대한 수준과 대표적인 예시를 중심으로 진행되었다. 이후 각각의 해석에 대해 일치도를 점검하였다. 3인의 판정에 불일치가 보인 부분에 대해서는 3인이 같이 자료를 반복적으로 분석하고 논의하는 과정을 통해 어떠한 판정이 자료의 본질을 가장 적절하게 포착했는지에 대한 합의에 도달했다(Braun & Clarke, 2006). 토론을 통해 합의에 도달하는 데 어려움을 겪는 경우 추가 인터뷰를 실시하기도 하였다.

III. 연구 결과

4명의 과학교사의 5가지 범주에서 탐색된 평가 전문성을 요약하면 〈표 3〉과 같다. ‘교육 목표’, ‘전략’, ‘평가 목적과 방법’에 대한 측면에서는 상대적으로 높은 전문성을 보였으나, ‘평정’과 ‘피드백’과 관련된 측면에서는 대부분의 교사들이 낮은 평가 전문성을 보였다. 구체적인 범주별 평가 전문성은 다음과 같다.

〈표 3〉 범주별 수준 판정

범주	교사별 수준			
	교사 A	교사 B	교사 C	교사 D
교육 목표	높음	낮음	낮음	높음
전략	높음	높음	낮음	낮음
평가 목적과 방법	높음	높음	높음	낮음
평정	낮음	높음	낮음	낮음
피드백	낮음	낮음	낮음	낮음

1. 교육 목표 측면

4명의 과학교사는 과학교육의 목표에 대해 다양한 수준을 보였다. 교사 A는 과학교육 목표를 충분히 이해하였다. 면담에서 교사 A는 자신의 교육 철학을 다음과 같이 설명하였다.

교사 A: 과학교육의 목표는 과학적 사고를 향상시키는 것이라 생각합니다. 그래서 저는 학생들이 스스로 과학을 공부하게 했습니다. 학생들이 선생님과 함께 과학 수업을 정말로 즐긴다면 과학적 목표를 얻기에 충분하다고 생각합니다. 또한, 이것은 학생들이 과학교육의 목표를 달성하고 있다는 것을 의미한다고 생각합니다.

반면, 교사 B의 교육 목표는 과학적 개념에 대한 명확한 이해였다. 따라서 단순한 개념의 이해를 목표로 갖는 교사 B는 교육 목표에 대한 수준은 높지 않음을 유추할 수 있었다.

연구자: 과학 수업에서 가장 중요한 것이 무엇이라고 생각하십니까?

교사 B: 학생들에게 가장 중요한 것은 과학의 개념을 명확히 이해하는 것입니다. 과학 개념을 정확하게 이해하면 과학 학습이 발생할 수 있습니다.

교사 A와 B의 이러한 다른 목표는 중간고사 주관식 문제를 채점할 때 두 교사가 나눈 대화에서도 확인할 수 있었다. 다음은 30kg인 물체가 지구에서 받는 중력의 크기를 기록하는 주관식 문제에 대한 교사의 채점에 대한 대화 내용이다.

교사 B: 30kgf를 어떻게 채점해야 할까요?

교사 A: 굳이 9.8을 곱해서 N으로 변환하지 않아도, kgf도 힘이니까 맞다고 해야 하지 않을까요?

교사 B: 문제가 지구에서 30kg의 중력을 구하라고 했으니, 학생이 쓴 f가 지구에 해당하는 중력 가속도 값인 9.8인지 달에 해당하는 중력 가속도 값인지 알 수 없으니 틀렸다고 해야 할 것 같아요.

교사 A: 저희가 의도한 것이 질량으로부터 중력을 환산하는 것이 목적이었다면, 학생이 쓴 f가 지구에서의 중력

가속도인지 확신 없다고 해도 이미 안거라고 할 수 있지 않아요?

교사 B: 그래도 지구에서의 중력 값을 물어보았으니, 정확하게 9.8 곱하여서 N으로 표현한 것만 맞다고 해야 한다고 생각해요.

교사 B는 지구라는 것이 명시가 되었기 때문에 정확히 9.8을 곱한 N값을 제시해야 한다고 생각하였다. 반면에 교사 A는 학생들이 질량(kg)에 각 행성의 중력 가속도(f)를 곱한 값이 그 행성에서의 중력(kgf)이라고 배우기 때문에, f에 굳이 9.8을 곱하지 않아도 학생이 이러한 과정을 사고하였다고 판단하여 유사 답안으로 인정하여야 한다고 생각했다. 이는 두 교사가 갖는 다른 교육목표에 기반한다. 교사 A는 과학적 사고가 교육목적이기 때문에 이를 정답으로 인정할 수 있었으나, 교사 B는 그렇지 않았다. 이와 유사하게도 지구의 중력가속도를 9.8이 아닌 10으로 곱한 학생에 대해서도 교사 A는 유사 답안으로 인정해야함을 주장하였으나, 교사 B는 교과서에 제시된 중력 가속도 값인 9.8만을 고수하였다. 사실 실제로 지구에서 중력 가속도는 위도마다 다르기 때문에 교과서에 제시된 중력 가속도 9.8은 평균값임에도 불구하고, 교사 B는 교과서의 제시된 하나의 중력 가속도 값만을 강요하였다.

교사 C의 교육 목표에 대한 지식은 과학적 개념을 체계적으로 이해하는 것이었다. 그의 목표는 교사 B의 목표와 비슷했다. 반면, 교사 D의 교육 목표는 과학적 문제 해결력을 증진하는 것이었다. 따라서 교사 D의 교육 목표에 대한 수준은 높았다. 이 둘의 다른 교육목표 역시 정체 전선을 그리는 주관식 문제 채점에서 확인할 수 있었다.

교사 D: 정체전선을 온난전선과 한랭전선이 겹치게 그리기만 하면 맞다고 해야 하지 않을까요?

교사 C: 아니죠, 우리나라를 기준으로 해서 온난전선과 한랭전선의 위치까지 맞아야 맞은거죠?

교사 D: 과학적으로 그게 맞긴 하지만, 학생들이 정체전선이 어떻게 형성되었는지에 대한 과정을 묻는 것이 의도였다면 위치는 중요하지 않다고 생각해요.

교사 C: 이렇게 그린 학생들도 정체전선이 만들어지는 원리는 알고 있다고 생각하지만, 문제에 우리나라 지도가 있으니, 전 맞다고 할 수 없다고 생각해요.

교사 D는 정체전선이 만들어지는 과정을 옳게 이해한 학생들은 교육 목표를 달성하였다고 판단하였지만, 교사 C는 교사 B와 유사하게 교과서에 제시된 위치가 정확한 정체전선만이 옳은 정답이라고 판단하였다. 김성기 외(2020)는 채점과정에서 교사의 교육 목표를 엿보았으며, 이를 통해 교사가 갖는 교육 목표에 대한 신념을 도출하였다. 본 연구에서도 인터뷰를 통해 판정된 교육 목표의 수준과 채점 과정에서 판정된 수준을 비교하였으며, 이러한 수준이 선행연구와 동일하게 일치하였다.

2. 전략 측면

교사 A의 과학교육 목표는 학생들의 관심과 호기심을 기르고 사고력을 높여 사려 깊은 아이디어를 만들어내는 것이었다. 따라서 그는 수업에서 교사와 학생 간의 상호 작용하며 질문과 대답 기술을 활용하여 기본 과학 개념을 소개했다. 그는 수업시간에 항상 학생이 토론에 참여할 수 있도록 하였다. 교사 A는 다음과 같이 부력의 크기를 측정하는 방법을 도입하는 수업에서도 일방적으로 부력의 크기를 측정하는 방법을 도입하기 보다는 학생들의 사고를 촉진하는 질문을 던지면서 학생들이 사고할 수 있는 전략을 구사하였다.

교사 A: 우리가 부력의 크기가 얼마인지 확인하려면 어떻게 하면 좋을까요?

학생 1: 물속에서 가볍게 느껴지게 하는 게 부력이니까, 공기 중에서 물체의 무게를 측정하고 물속에서 물체의 무게를 측정하면, 이 차이가 부력이 될 거예요.

학생 2: 물이 가득 비커에 추를 넣어요. 그러면 물이 넘치는데 넘친 물의 양을 저울로 측정하면 부력을 알 수 있어요.

교사 A: 그럼 두 학생이 말한 방법이 모두 부력의 크기를 측정한다고 했으니, 이 둘이 동일한 값으로 측정되는지 실험해 볼까요?

물론 과학 개념을 도입할 때는 교사 A의 설명이 있긴 하였으나, 학생들의 사고를 촉진하기 위해서 다양한 발문을 활용하였으며, 필요에 따라서는 시범실험을 보여주고 학생들이 생각하며 그것을 검증하도록 하는 등 다양한 전략을 활용하였다. 따라서 교사 A는 전략에서 높은 수준으로 판단되었다.

교사 B의 수업에서 학생들은 각 모듈별로 학습지를 바탕으로 스스로 탐구하였다. 교사 B는 과학적 개념을 학생들에게 도입하기 이전에 학생들이 학습지에 제공된 문제 상황을 인식하고 이를 스스로 해결해 보는 전략을 활용하였다. 실험을 이용하여 이를 탐구하기도 하고, 모듈원들간의 토론을 통해 이를 해결하도록 하는 등 다양한 방법을 활용하여 교사가 과학 개념을 도입하기 이전에 학생들이 스스로 탐구하도록 하였다. 교사 B는 학생들을 적극적인 참여자로 간주하였다. 따라서 교사 B는 교수 전략에 대한 수준이 높다고 판정하였다.

교사 C의 수업 전략은 수행평가에서 수행한 실험을 제외하면, 일방적 분필을 활용한 강의식 수업이었다. 교사 C의 설명은 체계적이었고 분필을 효과적으로 사용하였다. 그러나 교사 C는 학생들에게 필기 할 충분한 시간을 주지는 않았다. 그는 학생들이 수동적이라고 생각하고 교사 중심 방식으로 수업을 제공했다. 예를 들어, 그는 세 번 정도 시연을 한 후, 보여준 시연의 원리를 설명했다. 학생들이 보여준 시연에 대해서 왜 그러한지를 스스로 탐구하도록 질문하지 않았다. 수업시간에 학생들은 개념을 이해했는지에 대한 반응을 표현하지 않았다. 그는 학생들이 대답하기를 기다리지 않고 계속 설명했다.

연구자: 학생들에게 질문을 한 후 기다리지 않고 직접 답변했습니다. 왜 그랬나요?

교사 C: 틀렸다는 것을 압니다. 이것은 젊은 선생님들과는 다르지만 제가 학생이었을 때 이런 종류의 가르침을 배웠습니다. 이런 방법을 사용하지 않으려고 했지만 습관을 깨기 힘들네요. 학생 중심의 수업을 하는데 너무 많은 시간을 낭비한다고 생각합니다. 학생들에게 질문을 하지만 수업 시간이 짧기 때문에 기다릴 수 없습니다.

교사 C는 자신의 교수 전략에 문제가 있다는 것을 인식했지만 그의 방법을 바꾸지는 않았다. 이러한 관찰과 면담을 통해 교사 C의 교수 전략의 수준이 낮음을 알 수 있다. 교사 D는 수업에서 학생들이 교과서를 읽고 문제를 해결하도록 했으며, 학생들이 교과서 개념을 설명할 때 의미있게 이해하도록 비유를 사용했다. 교사 D가 보여준 전략은 비유를 활용한 강의법이였다. 학생들도 교사 D의 수업을 회상하면 비유가 가장 생각난다고 하였다.

연구자: 선생님 D의 수업에서 기억나는 게 뭐가 있어?

학생: 비유하여 설명하는 것이 기억나요. 앙금생성 반응 수업 때 앙금이 되는 것을 누구는 만날 수 없는 인연이다 하면서 설명했던 것이 기억나요.

학생 인터뷰처럼 교사 D의 수업에 대한 전략은 비유를 이용한 강의식 수업이였다. 물론, 비유를 활용한 전략이 옳지 않다는 것은 아니지만 비유만을 활용한 수업 전략은 그 수준이 높다고 판단할 수 없었다. 또한, 실험과 관련된 수업을 할 때도, 학생들이 스스로 실험을 하여 탐구하기 보다는 시범으로 교실을 돌아다니면서 보여주는 방식의 전략만을 활용하였다.

3. 평가의 목적과 방법 측면

교사 A는 수행평가의 목적이 학생이 아는 것을 확인하는 것이 아니라 학생의 학습 과정을 모니터링하고 학습을 돕는 것으로 인식하였다. 그래서 그는 수행평가를 위해 실험이 끝날 때 학생들이 제출한 실험보고서를 평가에 활용하는 것에 부정적인 견해를 가졌다.

연구자: 그럼 실험보고서를 수행평가에 활용한 적이 없어요?

교사 A: 아니요. 한두 번 활용했어요. 하지만 가능한 활용하지 않으려고 노력합니다. 수행평가는 중간고사나 기말고사와 같은 지필고사에서 평가할 수없는 것을 고려해야한다고 생각합니다. 부득한 경우에만 수행평가로 실험보고서를 사용했습니다. 가능하다면 다른 다양한 수행평가 방법과 전략을 찾고 있습니다.

교사 A는 실험보고서 역시 학생의 과정을 보지 않고 기록된 결과만을 평가하기 때문에 수행평가의 목적에 부합하지 않는다고 생각하였다. 그래서 교사 A는 지필평가와 다른 형태로 수행평가를 진행하고자 하였으며, 역할극과 같은 창의적인 방법 활용하여 학생들의 능력을 평가하려고 하였다. 교사 A와 같은 학년을 가르친 교사 B 역시 학생들의 태도와 창의적인 발표

및 UCC 등 다양한 수행평가를 시도하였다.

연구자: 수행평가의 목적이 무엇이라고 생각하세요?

교사 B: 수행평가의 목적은 학생들의 학습을 돕기 위한 것입니다. 평가 과정에서 학생들은 과학을 배울 기회도 갖게 됩니다. 수행평가를 하면 지필식 총괄평가에서 드러나지 않은 학생의 능력을 확인할 수 있습니다. 학생들은 다양한 능력을 가지고 있다고 생각하며, 수행평가를 통해 그러한 능력을 발견하고 개발할 수 있습니다.

이와 같이 교사 B는 수행평가가 총괄평가에서 관찰되지 않는 학생들의 강점을 발견할 수 있다고 생각하였다. 분명히 교사 B는 총괄평가와는 다른 수행평가의 목적을 알고 있었다.

반대로 교사 C와 D는 실험수업 후 학생들이 제출 한 실험 보고서를 수행평가로 활용하였다. 보고서 평가는 학생들의 수행과정을 볼 수 없기 때문에 결과 평가에 해당한다. 면담과정에서 교사 C는 수행평가의 목적이 학생들의 학습을 돕는 것이라고 믿었기 때문에 실험 보고서 평가가 과정을 평가하기에는 부적절한 방법임을 인정하였다.

교사 C: 작년에 저는 학생들이 수행평가에서 모르는 문제를 해결하기 위해 서로 토론을 할 수 있도록 했습니다.

하지만 올해에는 D 선생님과 협의하여 수행평가를 결정했기 때문에 이것을 할 수 없었습니다. D 선생님은 수행평가가 다른 사람의 도움 없이 개별 학생이 할 수 있는 것을 평가해야한다고 생각했습니다.

교사 D는 수행평가가 총괄평가의 한 유형이라고 믿고 있었다. 교사 D는 학생들이 교사의 도움 없이 스스로 평가를 수행하도록 요구하였다.

연구자: 선생님께서는 수업시간에 학생들의 다양한 질문에 친절하게 답변해 주셨습니다. 그런데 수행평가에서는 '이제 질문에 답하지 않겠습니다!'라고 하셨는데요. 왜 그런거죠?

교사 D: 학생들이 수업에서 배운 것을 평가하기 때문에 학생들이 스스로 문제를 해결해야한다고 생각합니다. 수업시간에 학생들에게 평가와 관련된 내용을 학습할 기회를 제공하였습니다.

이와 같이 교사 D는 수행평가를 통해 정확한 학생의 성취 수준을 판단하고자 하였다. 이러한 수행평가의 목적에서는 학생의 성장과 학습 지원은 존재하지 않았다. 이처럼 교사 A와 B, C는 수행평가의 목적을 잘 알고 있었다. 하지만 동일한 공동체에 속한 A와 B만 자신이 알고 있는 목적에 부합하는 방식으로 평가를 진행하였다. 반면에 교사 C는 같은 공동체에 속한 교사 D 때문에 자신의 알고 있는 대로 실천을 옮기지 못하고 교사 D의 방식을 따라가고 있었다.

여러 연구(Bell & Cowie, 2001; Harlen & Deakin Crick, 2003; Ruiz-Primo & Furtak, 2007)에 따르면 높은 수준의 평가 전문성을 갖춘 교사는 평가의 목적을 인식하여야 하며, Lyon(2013b)은 학생의 성장과 학습을 지원하기 위해 교사는 다양한 방법의 평가를 사용할 수 있어야 한다고 하였다. 그러한 의미에서 교사 A와 B는 높은 수준으로, 교사 D는 이 둘을 모두

충족하지 않기 때문에 낮은 수준으로 판정되었다. 교사 C의 경우 수행평가의 목적을 높은 수준으로 인식하고 있었으나, 교사 D로 인해 자신이 인지한 수행평가의 목적과 부합되지 않는 방법을 활용하고 있었다. 따라서 교사 D의 영향이 없었다면 교사 C는 자신의 인지한 수행평가 목적과 부합과 방법을 실행할 것으로 판단하여 교사 C는 평가 목적과 방법 측면에서 높은 수준으로 판정하였다. 또한, 교사 D와 같은 형태는 여러 연구(노태희 외, 2015; 남정희 외, 2005, 김성기 외, 2020)에서도 많이 보고된 것으로 아직까지 수행평가에 대한 목적을 제대로 인식하지 않는 교사가 존재함을 알 수 있다.

4. 평정에 대한 측면

교사 A와 B는 수행평가로 고무 찰흙을 이용하여 단층을 만들기 활동을 하기로 결정하였다. 이 수행평가의 가장 중요한 기준은 제한된 시간 내에 구조물을 구축하는 것이었다. 두 교사 모두 주어진 시간을 초과하면 점수가 차감될 것이라고 학생들에게 알렸다.

교사 A는 [그림 1]과 같이 지필평가와 유사한 교실 상황에서 수행평가를 진행하였다. 그는 즉시 학생들의 수행한 과제를 확인하고 점수를 매겼다. 이 평가과정에서 학생들의 의견은 고려되지 않았으며, 교사 A는 학생들이 만든 구조를 평가하고 구조가 올바른지 여부에 따라 점수를 매겼다.

연구자: 선생님께서 점수를 매길 때, 결과물을 보고 점수를 주는데 평가하는데 애매하거나 어려운 점은 없었나요?
교사 A: 애매한건 없었어요. 학생들에게 얘기했거든요. 내가 못 알아보면 틀린거다. 선생님이 알아보게 확실하게 만들라고요.

반대로 교사 B는 [그림 1]과 같이 정규 수업시간과 비슷한 과학 실험실에서 평가를 진행했다. 그녀는 교사 A와 수행 평가 기준에 동의했기 때문에 학생들에게 3분 안에 구조를 구축하도록 요구하였다. 교사 B는 학생들에게 자신이 수행한 과제를 설명하도록 요청하고 학생의 설명을 참고하여 평가하였다.

연구자: A선생님은 수행평가 할 때 바로바로 점수를 주는데, 선생님은 어떤 것이 상반인지 하반인지 학생들에게 물어보면서 점수를 매기던데요?
교사 B: 사실 그게 중요하다고 생각합니다.
연구자: 만든 모양이 잘 안보여서 질문 한건가요?
교사 B: 그러한 것도 있어요. 그리고 중간에 위, 아래를 바꾼 애들도 있어요. 정확하게 학생들이 아는지 모르는지를 체크하려고 했던 거죠.



(가) 교사 A의 평가 장면



(나) 교사 B의 평가 장면

[그림 1] 공동체 1에서 교사별 수행평가 장면

동일한 수행평가를 진행한 교사 A와 B와 달리 교사 C와 D는 다른 수행평가를 시행하였다. 다만, 두 교사 모두 태도 평가로 학습지에 대한 평가를 동일하게 하였다. 그래서 동일한 학습지를 사용하였다. 교사 C는 학생들에게 사전에 알리지 않고 도선 주변의 자기장 실험 중에 수행평가를 실시했다. 그는 학생들의 성적은 자연스럽게 평가되어야하며 미리 준비하는 것을 허용해서는 안 된다고 믿었다. 그는 자신이 제작한 평가 기준표를 활용하여 학생들의 결과를 관찰하고 점수를 매겼다. 그러나 그는 평가 과정에서 즉시 피드백을 제공하지 않았으며 수행평가 한 후 평가 기준표를 활용하기 보다는 자신의 기억에만 의존하여 피드백을 제공하였다. 태도평가는 노트평가로 진행하였다. 그는 노트평가를 미리 학생들에게 예고하지 않았다. 좋은 평가를 받기 위해 학생들이 인위적으로 준비하는 모습이 아닌, 평소 학생의 성실한 모습을 평가하기를 원했다. 그래서 교사 C의 노트평가는 불시에 이루어졌으며, 그날 노트를 가져오지 않은 학생들은 감점을 받았다. 그러나 그는 학습지에 빈칸이 있거나 하는 등의 사소한 학습내용의 누락에는 신경 쓰지 않았다.

교사 C: 수행평가는 10점 만점입니다. 10점 중 대부분은 실험 보고서입니다. 일부분이 수업에서 학생의 태도입니다. 학생들에게 보고서 점수가 완벽하더라도 태도 점수 때문에 점수를 잃을 수 있다고 말합니다. 태도 점수도 그날 노트를 갖고 오면 되요. 사실 어떤 내용을 기록했는지는 보지 않아요.

교사 D는 자신의 평가가 객관적인지 확인하기 위해 상당한 노력을 기울였다. 실험평가전 평가와 관련된 유불리를 없애기 위해서 실험대와 실험 조를 편성하기 위해 제비뽑기를 이용하여 정하였다. 또한, 보고서를 작성할 때 학생들의 질문에 대해 대답하지 않았다. 특히 그녀는 태도 평가를 위해 각별히 주의를 기울였다. 학습지의 장수와 빈칸 누락에 대해 각각 0.1점의 감점을 주었다. 그리고 이렇게 산출된 점수는 다시 학급별 차이를 고려한 환산점수를 부여하였다.

연구자: 매 시간 그 많은 학습지를 평가하는 것이 힘들지 않으세요?

교사 D: 힘들긴 해요. 하지만 학생들을 철저히 평가하기 위해서입니다. 수업에 참여하는 모든 학생들은 공정하게 대우 받아야합니다. 그러나 여전히 문제가 남아 있습니다. C 선생님의 점수는 저와는 다른 기준으로 평가되었습니다. 우리 반의 일부 학생이 낮은 점수를 받고 C 선생님에 속한 다른 반의 학생이 높은 점수를 받는 것은 불공평합니다. 그래서 상대 평가를 통해 점수를 환산합니다. 이렇게 하는 데 많은 시간과 노력을 들였기 때문에 정말 힘듭니다.

평가 결과에 대해 학부모와 학생들의 불만을 경험 한 교사 D는 주관적 평가를 시행하지 않기로 결정했다. 결과적으로 그녀는 평가 후 증거로 활용할 수 있는 자료만 고려하여 평가를 수행하려고 노력했다.

교사와 학생의 상호 작용을 기반으로 평가 한 교사 B를 제외하고 다른 교사는 교사 중심의 기준을 사용하여 학생을 평가했다. 교사 중심의 기준을 사용한다는 것은 낮은 수준의 전문성을 보여준다. 박종윤 외(2005)는 상호 작용을 강화하는 평가가 학생들의 학습에 긍정적인 영향을 미친다고 보고하였으며, 이러한 상호작용을 반영한 수행평가를 강조하고 있다. 하지만 여전히 평정과정에서 교사는 전통적인 평가 방식을 고수하였으며, 전반적으로 낮은 수준의 전문성을 보였다.

5. 피드백 측면

교사 A의 경우 평가가 시행될 때마다 즉각적으로 평가결과를 학생에게 고지하였다. 반면 교사 B의 경우 평가결과를 꼼꼼히 기록하고 이를 다시 검토한 뒤 수행평가 점수로 최종 고지하였다. 이처럼 교사 A는 수행결과에 대해 즉각적 판단을, 교사 B의 경우 데이터를 누적한 뒤 이를 바탕으로 판단을 하고 있었다.

교사 A는 판단시기 뿐만 아니라 판단의 주체도 교사 자신의 판단이 학생의 성적을 결정하는 데 충분하다고 생각하였다. 그러나 교사 B의 경우 자신의 판단뿐만 아니라, 학생 본인 및 동료의 판단 또한 고려하였다. 즉 평가에 참여하고 있는 다른 대상자의 판단도 자료로 이용하고 있었다.

교사 A는 수행의 가시적 정확성에 초점을 맞추었고, 교사 B는 수행하는 학생의 이해도 고려하였다. 또한, 교사 B는 평정과정에서 학생들에게 질문을 하여 과정 피드백을 제공하기도 하였다. 두 교사는 시기의 차이는 있지만 학생에게 평가의 결과를 고지하였다. 교사 B는 평가를 위해 수집한 학생의 여러 정보를 평정을 위해 사용하였지 이를 학생을 위한 피드백 자료로 제공하지 않다. 교사 A, B 모두 최종적으로 학생들에게 최종 점수만을 제공하였다.

교사 C는 평가결과를 즉각적으로 산출하지만 등급 및 그 등급에 할당된 이유를 학생들에게 구체적으로 고지하지 않는다. 그러므로 교사 C의 학생들은 자신이 왜 그 등급에 할당되었는지 알지 못했다. 교사 C는 즉각적 판단을 하지만 그 결과를 학생에게 고지하지 않았다. 그렇기 때문에 학생들은 교사의 피드백이 반영된 다음의 수행을 진행할 수 없었다.

연구자: 수행평가에서 A를 받은 이유가 무엇이라고 생각하니??

학생: 모르겠어요.

연구자: 왜?

학생: 선생님이 말씀 안 해주시고, 그냥 수행평가 점수만 알려주셨어요. 사실 어떤 애들이 A를 받는지 잘 모르겠어요.

이처럼 교사 C는 단순히 학생들을 판정한 결과만 통보하였지, 어떠한 점을 보완해야하는지 아니면 어떠한 점에서 우수한 평가를 받았는지와 같은 피드백을 제공하지 않았다. 반면, 교사 D의 학생들은 자신의 받은 등급과 평가기준 그리고 노트평가 점수가 상대평가로 등급화된다는 것에 이르기까지 구체적으로 알고 있었다. 그러므로 그들은 자신이 왜 그 등급에 할당되었는지 알고 있었다. 교사 D는 데이터에 근거한 판단을 하며 그 결과를 학생에게 고지하였으므로 학생들은 다음의 수행을 준비할 수 있었다. 하지만 교사 D는 평가 과정에서는 학생이 스스로 문제를 해결해야하기 때문에 과정 중에 피드백이 이루어지면 안 된다고 생각하였다.

학생: 선생님, 동서남북이 없어요?

교사 D: 학생 마음이에요. 네가 그린 걸 기준으로 채점할 거예요.

학생: 선생님. 전류의 방향은 어디예요?

교사 D: 조용히 하고, 학생이 알아서 하세요. 더 이상 질문하지 마세요.

이처럼 교사 D는 과정에 대한 평가와 이와 관련된 피드백을 전혀 고려하지 않았다. 학생들이 동서남북의 위치나 전류 방향의 설정에 어려움이 있다면, 이에 대한 힌트를 제공하여 다음 과제를 수행하도록 한 후 이를 과정 평가에 반영하면 된다. 이러한 과정 중의 평가 사항이 학생들에게 충분히 의미 있는 피드백이 될 수 있음에도 결과 중심의 평가와 피드백만을 수행하고 있었다.

평가에는 효과적인 피드백이 매우 중요하다(Hattie & Timperley, 2007). 평가를 통한 학생에 대한 이해는 학습의 다음 단계를 안내하기도 한다(Brookhart, 2011). 또한, 평가의 피드백은 학생들의 학습을 향상시킬 수 있는 방법을 알려주는 강력한 방법이다(Moeed, 2015). 따라서 교사는 평가 정보와 피드백을 학생에게 제공할 의무가 있다 (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Sadler, 1989). 하지만 본 연구에서 유일하게 교사 B가 평가 중 일부 과정 피드백을 제공하긴 하였지만, 결과에 대한 피드백을 제공하지 않았다. 즉, 이 연구에 참여한 4명의 과학교사 모두 적절한 피드백을 학생들에게 제공하지 못하고 있었다. 단지 그들은 판정 또는 점수 부여를 위한 평가를 하고 있었으며, 피드백과 관련된 측면에서 낮은 전문성을 보였다.

IV. 결론 및 제언

본 연구는 학생의 성장을 지원하는 형성평가와 관련하여 수행평가에 대한 과학교사의 평가 전문성을 탐색하는데 목적이 있다. 이를 위해 서울의 중학교 과학교사 4명을 대상으로 한 학기 동안 평가 장면을 포함한 수업을 관찰하고, 면담을 실시하여 과학교사의 수행평가와 관련한 평가 전문성을 탐색하였다. 선행연구를 통해 평가 전문성과 관련한 탐색을 ‘교육 목표’, ‘전략’, ‘평가 목적과 방법’, ‘평정’, ‘피드백’의 범주로 나누어 분석하였고 그 수준과 특성을 분석하였다.

연구결과, ‘교육 목표’에서 교사 A와 D는 상대적으로 높은 수준을 보였으며, ‘전략’에서는 교사 A와 B가 높은 수준을 보였다. 각각에 대해 절반정도가 높은 수준과 낮은 수준을 보였다. ‘평가의 목적과 방법’ 측면에서 교사 A, B, C가 높은 수준을, 교사 D가 낮은 수준을 보였다. 상대적으로 다른 범주에 비해 이 부분에 대해서 높은 수준을 보였다. 흥미로운 것은 높은 수준을 보인 교사 C는 같은 공동체에 속한 교사 D에 의해 자신의 전문성을 실천하지 못하고 낮은 수준의 교사 D에 따른 모습을 보였다. 이는 평가 전문성을 실천할 때, 개인적 차원의 전문성뿐만 아니라 공동체에 속한 교사의 평가 전문성이 영향을 줄 수 있음을 보여준다. ‘평정’에 대한 측면에서 교사 B를 제외하고 나머지 교사들은 모두 낮은 수준을 보였으며, ‘피드백’측면에서는 참여한 모든 교사가 낮은 수준을 보였다.

연구 결과를 통해 다음과 같은 결론을 도출할 수 있었다.

첫째, 교사들의 수행평가에 대한 평가 전문성은 전반적으로 낮으며, 특히 실천적 측면에서 보다 더 어려움을 갖고 있다. 본 연구에서 5개의 평가 전문성 판정에 교사 A와 B가 3개부분에서 높은 수준을 판정 받았으며, 교사 C와 D는 1개 부분에서만 높은 수준으로 판정받았다. 이처럼 과학교사들은 대부분 낮은 수준으로 판정되었다. 5개의 범주는 엄격하게 인지적 측면과 실천적 측면으로 구분되기 어려우나 그 강조점에 따라 2가지 측면으로 구분될 수 있다. 특히, 평가 방법의 경우 실천적인 성격이 강할 수 있으나, 선행연구에서 평가 방법이라는 것이 교사가 인지한 평가 목적에 크게 영향을 받는다고 하였다. 이러한 점을 고려할 때, ‘교육 목표’, ‘전략’, ‘평가 목적과 방법’은 인지적 측면이 강한 평가 전문성 범주로, ‘평정’이나 ‘피드백’은 실천적 측면이 강한 평가 전문성 범주로 분류할 수 있다. 이 분류로 교사의 수준을 비교해 보면 인지적 측면에 대한 평가 전문성 수준이 실천적 측면에 대한 평가 전문성 수준이 비해 상대적으로 높았다. 이는 과학교사들이 실천적 역량으로 전문성이 확대될 때 그 수준이 낮아지는 것을 보이는 것으로 교사가 알고 있는 지식을 실천까지 확대할 수 있는 실천적 역량의 평가 전문성 함양과 관련된 연구가 요구됨을 알 수 있다. 특히, ‘피드백’ 측면에서 모든 과학교사의 평가 전문성이 낮게 판정되었다. 이는 평가를 통해 학생의 학습을 지원하기 보다는 학생의 성적을 평정하는 측면으로 수행평가가 진행되고 있음을 보여준 것으로 이와 관련한 교사 연수가 절실함을 알 수 있다.

둘째, 교사의 평가 전문성 수준이 높다고 하여 반드시 높은 실천의 수준을 보이지는 않을 수 있으며, 교사가 속한 공동체에 영향을 받는다. 본 연구에서 교사 C는 수행평가 목적에 대해 높은

수준의 인식을 갖고 있었다. 하지만 자신이 속한 공동체의 동료 교사 D에 의해 이러한 자신의 높은 수준의 인식을 실천하는데 어려움을 겪었으며, 결국 낮은 수준의 실천을 하는데 합의하는 모습을 보였다. 따라서 궁극적으로 학교 현장에 높은 수준의 평가 전문성이 발현되기 위해서는 교사 개인적 차원의 노력뿐만 아니라 공동체 차원의 노력이 필요함을 알 수 있다. 이러한 점은 평가 전문성과 관련하여 예비교사를 포함한 교사 교육에서 반영되어야 할 것이다.

이 연구의 결론을 토대로 과학교사들의 평가 전문성 향상을 위한 제언은 다음과 같다.

첫째, 실천적 역량 중심의 교사 연수가 필요하다. 본 연구에서 얻은 결과와 같이 교사들은 전반적으로 인지적 측면의 평가 전문성은 상대적으로 높은 수준을 보였으나, ‘평정’이나 ‘피드백’과 같은 실천적 측면에서 낮은 수준을 보였다. 이론중심의 교사 연수에서 벗어나 실제적인 평가 장면을 토대로 실습의 기회를 제공하여 교사들이 교실 현장에서 학습과 성장을 지원할 수 있는 역량을 기르도록 해야 할 것이다.

둘째, 교사 연수시 개인별 연수보다 교사가 속한 공동체가 함께 참여할 필요가 있다. 이 연구에서 동일 공동체에서 다른 수준의 평가 전문성을 갖는 교사가 속해있을 때, 높은 평가 전문성을 가진 교사의 의견을 따르기 보다는 그 공동체에서 상대적인 지위나 서열에 의해 의사결정이 이루어져서 교사가 높은 평가 전문성이 갖고 있더라도 같은 공동체에 속한 상대 교사의 영향을 받아 낮은 평가 전문성을 실천할 수 있음을 알 수 있었다. 따라서 높은 수준의 평가 전문성 실천은 공동체 차원의 영향을 받기 때문에 공동체에 속한 교사들이 같이 연수를 받으면서 서로의 평가에 대한 의견을 공유하면서 보다 높은 차원으로 실천할 수 있도록 지원하는 것이 필요하다.

셋째로, 본 연구는 한학기의 수업을 관찰하고 분석하기에는 많은 교사를 연구대상으로 하기 어려웠다. 그래서 4명의 교사를 대상으로 질적 연구를 수행하였다. 그렇기 때문에 이 연구 결과를 과학교사의 전반적으로 확대 해석하는데 주의가 필요하다. 따라서 보다 전반적인 과학교사의 수행평가에 대한 평가 전문성을 탐색하기 위해서 본 연구 결과를 기반한 대규모의 양적연구가 필요하다. 이러한 연구는 보다 일반화된 과학교사의 수행평가에 대한 평가 전문성을 보여줄 것이며, 이는 교사교육에 보다 깊은 시사점을 줄 것으로 기대된다.

넷째로, 본 연구에서 활용한 평가 전문성 범주별 수준 판정은 중간 수준의 설정의 어려움과 범주별 구별되는 경향을 비교하기 위해 2개 수준으로 제한하여 판정하였다. 하지만 연구 방법에서 밝혔듯이 2개 수준이외에도 보다 다양한 수준으로 이를 설정할 수 있으며, 이를 통해 보다 다양한 범주별 경향 탐색이 가능할 것이다. 본 연구는 수행평가와 관련하여 기초적인 수준의 연구를 진행하여 2개 수준으로 제한하였으나, 2개 수준이외의 다양한 수준에 따른 교사의 평가 전문성 탐색은 보다 실제적인 교사의 도움을 줄 것으로 기대된다.

참고문헌

- 곽영순(2015). 미래 학교교육 변화 및 교육과정 재구성에 필요한 교사 전문성 탐색. **교과교육학연구**, 19(1), 93-111.
- 교육부(1998). 수행평가의 이해.
- 교육부(2017). 과정을 중시하는 수행평가 어떻게 할까요?
- 김석우, 김성숙, 장재혁, 이승배. (2017). 이원목적분류와 성취기준에 근거한 중등교사의 평가전문성 분석. **교육평가연구**, 30(2), 173-197.
- 김성기, 민희정, 백성혜(2020). 학습공동체를 통한 과학교사의 평가에 대한 신념에 관한 연구. **교원교육**, 36(1), 1-21.
- 김성기, 최은주, 백성혜(2015). 중학교 과학과 서술형 평가의 문항 특성 및 요인 분석. **대한화학회지**, 59(5), 445 - 453.
- 김현재, 송민영 (1997). 과학 교육에서의 수행 평가-포트폴리오 평가를 중심으로. **열린교육연구**, 5(2), 77-94.
- 남정희, 최준환, 고문숙, 김재홍, 강순민, 임재향, 공영태(2005). 형성평가를 이용한 교수-학습 전략이 중학교 학생들의 과학개념 이해, 학습동기, 메타인지 능력에 미치는 영향. **대한화학회지**, 49(3), 311-320.
- 노은희, 서민원(2016). 창의·융합형 인재 양성을 위한 수행 평가 시행의 과제와 현장 안착 방안, 창의·융합형 인재 양성을 위한 학생 평가 및 기록 개선 방안 토론회. 3-16. 한국교육과정평가원. ORM 2016-85.
- 노태희, 이재원, 강석진, 강훈식(2015). 중등 과학교사의 과학 평가 실태와 지향. **한국과학교육학회지**, 35(4), 725-733.
- 박종윤, 남정희, 유희선(2000). 상호작용을 강화한 형성평가 수업전략이 중학교 과학학습에 미치는 영향. **한국과학교육학회지**, 20(3), 468-478.
- 박휴용(2019). 포스트휴먼 시대의 학교교육의 변화와 교사전문성의 방향. **교육철학연구**, 41(2), 47-80.
- 신정윤, 양일호(2017). 과학과 수행평가의 문제점과 개선 방향. **예술인문사회융합멀티미디어논문지**, 7(1), 878-883
- 임천택(2010). 초등교사의 국어과 평가 전문성 실태 분석-평가 문항 개발을 중심으로. **학습자중심교과교육연구**, 10(2), 347-370.

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.). *The professional*

- knowledge base of science teaching* (pp. 205-221). Netherlands: Springer.
- Bayat, K., & Rezaei, A. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education*, 3(1), 139-146.
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85(5), 536-553.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12.2011
- Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2013). Developing and accessing beginning teacher effectiveness: The potential of performance assessments. *Educational Assessment, Evaluation and Accountability*, 25(3), 179-204.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into Practice*, 55(2), 153-159.
- Gearhart, M., Nagashima, S., Pfothner, J., Clark, S., Schwab, C., Vendlin, T., ... & Bernbaum, D. J. (2006). Developing expertise with classroom assessment in K-12 science: Learning to interpret student work. Interim findings from a 2-year study. *Educational Assessment*, 11(3-4), 237-263.
- Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, 10(2), 169-207.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Heritage, M. (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?* Los Angeles: Council of Chief State School Officers.
- JCSEE. (2015). Program evaluation standards statements. The Joint Committee on Standards for Educational Evaluation. Retrieved from <http://www.jcsee.org/program-evaluation-standards-statements>.

- Kim, S., & Paik, S. (2016). An analysis of science teachers' stages of concern and levels of use on descriptive assessment. *Journal of the Korean Chemical Society*, 60(5), 353-361.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Lyon, E. G. (2013a). Learning to assess science in linguistically diverse classrooms: Tracking growth in secondary science preservice teachers' assessment expertise. *Science Education*, 97(3), 442-467.
- Lyon, E. G. (2013b). Conceptualizing and exemplifying science teachers' assessment expertise. *International Journal of Science Education*, 35(7), 1208-1229.
- Moeed, A. (2015). *Science investigation: Student views about learning, motivation and assessment*. NY: Springer.
- Panadero, E., Brown, G., & Courtney, M. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice*, 21(4), 365-383.
- Puad, L. M. A. Z., & Ashton, K. (2020). Teachers' views on classroom-based assessment: An exploratory study at an Islamic boarding school in Indonesia. *Asia Pacific Journal of Education*, 1-13.
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57-84.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.
- Stiggins, R., & Duke, D. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan*, 90(4), 285-291.
- Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education*, 34(2), 239.
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.

· 논문접수 : 2021.01.05. / 수정본접수 : 2021.01.29. / 게재승인 : 2021.02.16.

ABSTRACT

Exploring the science teachers' assessment literacy for performance assessment

Sung-ki Kim

Associate Research Fellow, Korea Institute for Curriculum and Evaluation

Hee-jung Min

Teacher, Eonnam Middle School

Min-su Ha

Professor, Kangwon National University

Seoung-hey Paik

Professor, Korea National University of Education

The purpose of this study is to explore the assessment literacy of science teachers for performance assessment which is related to formative assessment to support students' growth. To this end, four middle school science teachers in Seoul were observed during the semester, including evaluation scenes, and interviews were conducted to explore the assessment literacy related to the performance assessment of science teachers. Through prior research, analysis related to assessment literacy was divided into the categories of 'Educational goal' and 'Strategy', 'Assessment objective and method', 'Rating', and 'Feedback', and the level and characteristics were analyzed. If the characteristics of each category were sufficiently observed, it was judged as a high level, and if the characteristics did not appear, it was judged as a low level. As a result of the study, most science teachers were judged at a low level in the five categories. In addition, assessment literacy mainly in cognitive aspects, such as 'Education goal' and 'Strategy' and 'Assessment objective and method', showed a relatively high level compared to the practical aspects, such as 'Evaluation' and 'Feedback'. This shows that when science teachers expand their expertise to practical competency, the level of assessment literacy decreases. Therefore, research on assessment literacy extended to the practical area is required. In particular, in terms of 'feedback', the assessment literacy of all science teachers was judged to be low. This shows that teachers use performance assessment to evaluate student grades rather than to use it as an evaluation that supports learning, and teacher training in this regard is necessary. By synthesizing the results of this study, several plan can be drawn to improve teachers' assessment literacy.

Key Words: *Assessment literacy, Performance assessment, Formative assessment*