

발성 사고법을 이용한 예비 화학 교사의 지필평가 문항 제작 과정의 특징 분석

노태희¹, 김혜리¹, 한재영², 강훈식^{3*}

¹서울대학교, ²충북대학교, ³서울교육대학교

An Analysis of the Characteristics of the Processes of Pre-Service Chemistry Teachers in Making Written Test Items Using Think-Aloud Method

Taehee Noh¹, Hyeree Kim¹, JaeYoung Han², Hunsik Kang^{3*}

¹Seoul National University, ²Chungbuk National University, ³Seoul National University of Education

ARTICLE INFO

Article history:

Received 6 February 2017

Received in revised form

6 March 2017

14 April 2017

Accepted 15 April 2017

Keywords:

processes of making written test item, think-aloud method, pre-service chemistry teacher

ABSTRACT

In this study, we investigated the characteristics of the individual processes of pre-service chemistry teachers in making written test items. For this, we analyzed the think-aloud processes of eight pre-service chemistry teachers while making six written test items about gas laws and their in-depth interview scripts. The characteristics of the processes were found to be divided into six categories; the lack of systematicity in planning stage for making written test items, the lack of awareness and consideration for the interrelationship between instruction and assessment, the diverse criteria in determining test item types, the difficulties in judging behavioral domains and difficulties of items in the table of specification, the limitation in using teacher's guides and misunderstanding of teacher's guides, and the limitation in point and/or content of review and differences depending on the time of review. These can provide some significant guidelines and implications in finding ways to improve pre-service chemistry teachers' ability to make written test items.

1. 서론

학교 현장에서 실시되는 평가는 학생들을 선발하고 분류하는데 필요한 정보 수집의 기능을 가지는 동시에 학생들의 능력을 파악하고 진보를 관리하며 교수·학습에 대한 개선 정보를 제공하는 등의 기능을 수행한다(McMillan, 2014). 이처럼 평가는 교수 전략이나 학습 운영 등과 함께 교사의 전문성이 요구되는 실천적 과정 변인으로서 교육과정의 충실한 운영과 수업의 질 개선을 위해 필수적인 요소이며(Kim, 2002), 이는 과학 학습 평가에도 동일하게 적용된다. 즉, 과학 학습의 평가 결과는 학생들의 학습 과정과 결과 및 학습에서의 다양한 특성과 장단점을 파악하고, 교수 목표와 수업 계획, 수업 운영, 평가 방법 등의 적절성을 평가함은 물론, 교수·학습 및 평가 방법을 개선하여 학습 효과를 높이고 학생들의 진로를 지도하는 데 유용하게 활용된다(Atjonen, 2014; Kwon *et al.*, 2012; McMillan, 2014). 따라서 교사에게는 교수 목표에 부합하는 성취 기준을 설정하고, 성취 기준 도달 정도를 측정할 수 있는 적합한 평가 방법이나 도구를 선정 및 개발하여 평가와 채점을 실시한 후 성적을 부여하고, 평가 결과를 분석 및 해석하여 교수·학습 개선, 학생과 학부모 및 유관 기관와의 의사소통에 효과적으로 활용할 수 있는 능력이 요구된다(Abell & Siegel, 2011; Brookhart, 2011; Kim *et al.*, 2012; Nam *et al.*, 2006). 예를 들어, 교사는 평가의 실시 시기 및 유형에 따른 목적과 특징을 이해하고, 학생들의 특성과 수준에 대한 이해를 토대로 학습 목표에

부합하는 평가도구를 선택 및 개발할 수 있어야 한다. 특히 지필평가의 경우 문항의 타당도, 신뢰도, 난이도, 변별도 등에 대한 이해와 문항 유형별 제작 원리를 충실히 반영하여 평가도구를 개발할 수 있어야 한다. 또한 명확한 채점 기준을 수립하여 채점하되, 학생의 특성을 반영하여 채점 기준을 수정 및 보완할 수 있음을 이해하고, 평가의 목적 및 내용에 기초하여 평가 결과를 해석하고 이를 학생이나 교수 및 학습에 대한 의사결정 자료로 활용할 수 있어야 한다. 그리고 평가 방법이 교수 및 학습 활동에 미치는 긍정적인 영향과 부정적인 영향에 대해서도 파악할 수 있어야 한다(Kim *et al.*, 2012; Nam *et al.*, 2006). 이와 같은 교사의 평가 전문성은 교사 전문성의 핵심 요소로서 다른 전문성 요소들과 밀접한 관련이 있으므로(Black *et al.*, 2004; Cho & Ko, 2008; Hashweh, 2005; Magnusson *et al.*, 1999; Park & Oliver, 2008), 교사의 전문성을 발달시켜 성공적인 교수·학습을 유도하는 데 중요한 역할을 담당한다.

그럼에도 불구하고 교사들의 평가 전문성은 만족할 만한 수준에 이르지 못하는 것으로 보고되고 있다. 예를 들어 많은 교사들이 평가 방법과 도구들을 적절하게 선택 및 개발하거나, 평가 결과를 타당하게 분석 및 해석하거나, 평가 결과에 기초하여 교수·학습 과정에서 학생들이 겪는 어려움을 진단하고 교수·학습 및 평가 전략을 개선하는 데 있어서 다양한 어려움을 겪는 것으로 보고되고 있다(Kang & Kang, 2015; Lee *et al.*, 2004; Min, 2012). 또한 과학 교사들의 평가 목적에 대한 인식이 부족할 뿐만 아니라 평가 목적과 방법이 일치하

* 교신저자 : 강훈식 (kanghs@snu.ac.kr)

** 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2014R1A1A4A01008263).

http://dx.doi.org/10.14697/jkase.2017.37.2.0225

지 않는 경우도 있었다(Noh *et al.*, 2015). 따라서 교사의 평가 전문성을 제고하기 위한 전략이나 방안을 모색하는 일은 현 시점에서 매우 중요하고 필요하다. 특히 2012년부터 시행되고 있는 성취평가제의 성공적인 정착을 고려할 때, 그 의미와 파급 효과는 크다고 할 수 있다.

한편, 교사 전문성 제고를 위해서는 교육 현장에서 전문성 향상을 도모할 수 있는 능력을 갖춘 교사들이 개선을 향한 의지를 가지고 이를 위한 실천적 수행 과정을 지속적으로 거치는 것이 무엇보다 중요하다. 평생교육의 관점에서 볼 때 이러한 능력과 의지는 교사교육이 시작되는 예비교사 교육과정에서부터 교직 경력이 증가함에 따라 지속적으로 발달되므로, 예비교사 교육과정에서부터 수업 전문성 제고를 위한 기본적 토대를 체계적으로 제공할 필요가 있다(Cho, 2004). 또한 현직교사 연수에 비해 상대적으로 시간과 대상의 제약이 적고 집중적인 교육이 가능한 예비교사 교육과정의 질적 변화를 우선적으로 도모하고 이를 교사 연수와 연계하는 것이 보다 효과적이고 경제적인 것이다. 따라서 교사의 평가 전문성을 신장시키기 위해서는 예비교사 교육과정에서부터 평가 전문성 함양을 위한 체계적이고 집중적인 교육 기회를 제공할 필요가 있다. 특히 교사의 전문성이 실천적 지식의 성격을 지니고 있으므로, 실제 교실 상황에서 이루어지는 교사의 평가 실행 상황에서 평가 전문성의 발달을 촉진할 수 있는 교육 기회가 필요하다(Min, 2012; Sato *et al.*, 2005). 그러나 우리나라의 경우 예비교사 교육과정이나 각종 교사 연수 과정에서 평가 전문성 함양을 위한 실제적이고 실천적인 교육 기회가 충분하게 제공되지 못하고 있을 뿐만 아니라 평가 전문성 강좌에 대한 교육적 요구도 부족한 실정이므로(Kim *et al.*, 2010; Song *et al.*, 2014), 현 상황이 그리 낙관적이지는 않다고 볼 수 있다.

이를 개선하기 위해서는 우선적으로 예비 과학 교사들의 평가 전문성 수준에 대한 정보를 확보하고 이에 기초하여 예비교사들의 평가 전문성 향상을 위한 구체적인 방안을 모색할 필요가 있다. 특히 평가 전문성 요소 중 평가 도구의 제작 능력은 평가 전문성 수준 분석을 위한 중요한 기준이 되므로(Kim *et al.*, 2012; Nam *et al.*, 2006), 예비교사들이 평가 도구의 핵심적인 구성 요소인 평가 문항을 제작하는 과정을 살펴보고 그 과정에서 그들이 보이는 특징적인 측면과 평가 전문성 측면에서 부족한 점 등을 분석하는 과정은 매우 의미있고 필요하다고 할 수 있다. 그러나 과학 평가 전문성에 관한 선행연구들은 주로 초, 중등 현직 교사들을 대상으로 설문지나 면담 등을 통해 전반적인 평가의 실태 및 평가에 대한 인식을 조사하거나(Jang & Kim, 2002; Jeong & Choi, 2014; Kim *et al.*, 2000; Ko *et al.*, 2013; Noh *et al.*, 2015; Park, 2016; Shin *et al.*, 2016), 과학과 행동 영역이나 수행평가 방법 등과 같은 특정 평가 요소나 방법에 대한 인식을 조사하거나(Kim & Hyun, 2005; Kim & Kim, 2002; Park *et al.*, 2011), 평가관(Noh *et al.*, 2009) 및 평가 전문성에 대한 인식(Kang & Kang, 2015)을 조사하거나, 수행평가 시 나타나는 채점자간 불일치 유형을 분석(Kim & Yoo, 2012)하는 것이 대부분이었다. 일부 연구자들은 교사가 제작한 서술형 평가 문항의 행동영역 내용 타당도 및 이에 영향을 미치는 요인을 분석하거나(Choi & Paik, 2016), Klopfer의 교육 목표 분류 체계에 의한 초등학교 과학과 지필평가 문항을 분석하거나(Yang *et al.*, 2008), 지필평가 또는 수행평가 문항을 제작 및 실행하는 과정을 교수내용지식(Pedagogical Content Knowledge, 이

하 PCK)와의 관련성 측면에서 조사하기도 하였다(Falk, 2012; Min, 2012).

이처럼 지금까지의 연구들은 현직 과학 교사와 수행평가에 치중되어 있으며, 예비 과학 교사들을 대상으로 진행된 연구는 국내외적으로 매우 부족한 실정이다. 국내에서 예비교사의 형성평가를 위한 선다형 물리 개념 검사 문항의 제작 수준을 조사하는 연구(Choi & Kim, 2013)가 진행된 바 있으나, 예비교사가 제작한 선다형 문항에 대하여 결과물로서 선다형 문항의 완성도 및 수준을 분석하는 것에 초점을 두고 있다. 이로 인하여, 예비 과학 교사들이 지필평가 문항을 개별적으로 제작하는 과정의 특징에 대한 체계적이고 총체적인 정보는 부족한 실정이다. 지필평가는 학교 현장에서 실시되는 평가 중 가장 보편적인 형태로서 학생들의 과학 지식 측정 뿐 아니라 탐구 활동 과정에 대해서도 교사의 재량에 따라 이용할 수 있으며, 시간이 많이 소요되지 않고, 많은 인원을 동시에 측정할 수 있으며, 평가가 용이하다는 장점을 가지고 있다(Lee *et al.*, 2004; Marzano, 2006). 이러한 장점 때문에 지필평가는 학교 현장에서 수행평가와 더불어 유용한 평가 방법으로 활용되고 있으며, 심지어는 학교급이 올라갈수록 수행평가보다 지필평가의 활용 비율이 증가하고 있다(Jeong & Choi, 2014). 따라서 지필평가와 관련된 예비 과학 교사들의 전문성을 신장시키기 위한 노력이 필요하다.

한편, 발성 사고법(think-aloud method)은 특정 과제를 해결하면서 머릿속에 떠오르는 생각들을 그대로 말로 표현하게 하는 방법으로서, 학생들의 과제 해결 과정과 이에 영향을 주는 변인, 과제 해결의 실패 원인 등에 대하여 의미 있는 정보를 제공할 수 있다는 장점이 있다(Ferguson *et al.*, 2012; Schellings & Broekkamp, 2011). 따라서 발성 사고법은 예비 과학 교사들의 지필평가 문항 제작 과정의 특징을 체계적이고 총체적으로 파악하는 데 유용할 수 있을 것이다. 이에 이 연구에서는 예비 화학 교사들이 발성 사고법을 통하여 지필평가 문항을 제작하는 과정의 특징을 바람직한 평가문항 제작 과정의 관점에서 귀납적으로 분석하였다.

II. 연구 방법

1. 연구 참여자

연구 참여에 동의한 충청북도 지역 사범대학 3학년 학생 8명이 본 연구에 참여하였으며, 구체적인 정보는 Table 1과 같다. 성별은 남학생이 3명이고 여학생이 5명이었으며, 참여자 모두 교육학 과목 중 교육평가 강좌와 과학교육론 강좌에서 과학 학습 평가의 목적과 기능, 과학 학습 평가 영역, 과학 학습 평가 유형, 평가 도구의 질 판단 기준, 평가 문항의 제작 원리 등을 이미 학습하였다. 그리고 8명 중 3명은 개인 교습 지도 경험이나 교육평가 강좌 수강 등을 통해 비과학 교과 영역에서 지필평가 문항 제작 경험을 가지고 있었고, 2명은 이원목적분류표 작성 경험이 있으며, 1명은 과학교과교육학 강좌를 수강하면서 과학 교과에서 지필평가 문항 제작 경험을 가지고 있었다.

Table 1. The characteristics of the participants

	예비교사 A	예비교사 B	예비교사 C	예비교사 D	예비교사 E	예비교사 F	예비교사 G	예비교사 H
성별	남자	여자	남자	여자	여자	남자	여자	여자
평가관련 수강 강좌	교육평가, 과학교육론	교육평가, 과학교육론	교육평가, 과학교육론, 물리과 교재 연구 및 지도법	교육평가, 과학교육론	교육평가, 과학교육론	교육평가, 과학교육론, 물리과 교재 연구 및 지도법	교육평가, 과학교육론	교육평가, 과학교육론
지필평가 문항 제작 관련 경험	교육평가 강좌에서 문항 유형별 문항 제작 과제 수행	×	×	교육평가 강좌에서 이원목적분류표 작성	개인 교습 지도 과정에서 비과학 교과 영역 문항 제작	물리교재 연구 및 지도법 강좌에서 형성평가 문항 제작, 학창 시절 내용별 문항 유형 분석 경험 풍부	×	교육평가 강좌에서 1개 대단원에 대한 문항 제작, 이원목적분류표 작성

2. 연구 절차

연구 참여자들은 본 연구의 직전 학기에 화학교육 목표, 화학 교수·학습 이론, 화학 실험 지도, 화학 학습 평가 등의 내용으로 구성된 과학교육론 강좌를 이수하였다. 그리고 화학 학습 평가의 경우 학기 말에 화학 학습 평가를 위한 이론적인 내용만을 학습한 상태로 학기가 마무리되었다. 이에 본 연구의 자료 수집은 연구 참여자들이 수강하고 있는 3학년 과학교육 관련 강좌(3학점)의 1주와 2주차에 걸쳐 이루어졌다. 즉, 직전 학기에 배운 화학 학습 평가 관련 학습 내용을 환기시키는 오리엔테이션을 간단하게 실시한 후, 개별적으로 지필평가 문항을 제작하는 흐름으로 연구가 진행되었고 이 과정에서 필요한 자료를 수집하였다. 참고로, 해당 강좌의 이후 학습은 화학 학습 평가와는 무관한 내용으로 이루어졌다. 구체적인 연구 절차는 다음과 같다.

1주차에는 오리엔테이션을 진행하였다. 즉 과학 학습 평가의 목적과 기능, 평가 시기와 목표 및 방법에 따른 과학 학습 평가의 분류, 평가 도구의 타당도와 신뢰도, 지필평가 문항의 제작 원리 등에 대하여 구체적인 사례를 통해 설명하였다. 특히 지필평가 문항의 제작 원리에서는 이원목적분류표의 특징과 과학 교과에서의 사례를 제시하였고, 문항 유형별 특징과 제작 시 유의점 등을 과학 교과에서의 예를 보여주면서 설명하였다.

2주차에는 발생 사고법을 활용한 개별 지필평가 문항 제작 활동을 진행하였다. 먼저 발생 사고법에 대한 미숙함이 연구에 미치는 영향을 최소화하기 위해 이 연구와 관련이 없는 내용으로 발생 사고법에 대한 연습을 30분 동안 실시하였다. 그 후 중학교 1~3학년군 과학 ‘분자 운동과 상태 변화’ 단원의 보일 법칙과 샤를 법칙에 대하여 평가 영역의 행동 차원 중 ‘지식’ 영역에서 각각 선택형 2문항과 서답형 1문항을 120분 동안 발생 사고법에 따라 개별적으로 제작하도록 하였다. 이 때, 천재교육(이)과 비상교육 2개 출판사의 2009 개정 과학과 교육과정에 따른 중학교 과학 1 교과서와 교사용 지도서를 참고 자료로 제공하였으며, 스마트폰을 사용한 정보 검색도 가능함을 안내하였다.

예비교사들이 자동화된 사고 과정을 거치는 경우에는 이들의 생각이 발생 사고법으로 잘 드러나지 않는 경우가 있으므로, 사후 면담도 실시하였다. 이때, 동시에 여러 명의 예비교사들이 독립적인 장소에서 발생 사고 과정에 따른 지필평가 문항 제작 활동을 실시하고 문항 제작 종료 즉시 면담을 실시하기 위해, 예비교사 2명당 1명의 면담자

가 면담을 진행하였다. 그리고 면담을 실시하기 전에 연구자는 면담 시나리오를 작성하여 면담자들에게 제공하였고, 이에 대하여 여러 차례 설명해주고 숙지 여부를 확인함으로써 연구에 미치는 면담자의 영향을 최소화하려고 노력하였다. 4명의 면담자는 담당 예비교사의 샤를 법칙에 대한 문항 제작 과정을 관찰하면서 궁금하거나 특징적인 부분들을 관찰 노트에 메모하였으며, 이를 토대로 사후 면담을 개별적으로 30분~60분 동안 실시하였다. 모든 문항 제작 과정과 면담 과정은 녹음·녹화하여 전사본을 작성하였다. 이 전사본을 주 분석 자료로 하고 예비교사들이 출제한 문항 자료와 해당 수업의 과제로 제출한 개인별 강의 수강 저널 등을 보조 자료로 활용하여 결과를 분석 및 해석하였다.

3. 결과 분석 방법

본 연구의 주 분석 자료인 개별 발생 사고 및 면담 자료 해석의 초점은 예비교사들이 지필평가 문항을 제작하는 전반적인 과정에서 나타나는 특징을 찾고 이를 평가 전문성의 관점에서 분석하여 예비교사교육에 대한 시사점을 찾는 것이었다. 이를 위하여 수집한 자료들은 자료 분석을 통해 도출한 범주를 바탕으로 자료를 재검토하는 과정을 지속적으로 반복하여 범주를 정교화시키는 지속적 비교 방법 (Strauss & Corbin, 1998)을 사용하여 분석하였다. 즉 분석자 2인이 모든 수집 자료들을 반복적으로 분석하면서 예비교사 개인별로 발생 사고 과정에서 나타나는 특징들을 발췌 및 정리하여 프로파일을 만들었다. 그 후 이 프로파일들을 분석하여 예비교사들이 개별적으로 지필평가 문항 제작을 위해 계획하고 실행하며 검토하는 과정에서 나타나는 특징에 대한 하위 범주를 추출하였다. 그러나 연구진 내부 논의 과정에서 각 하위 범주의 위계와 수준의 일관성이 부족하고 문항 제작 절차에 따른 분석의 필요성이 제기되어 이를 바탕으로 지필평가 문항 제작 과정의 특징과 그 과정에 영향을 미치는 요인에 대한 하위 범주를 각각 구분하여 추출하였다. 추출한 범주를 정교화한 후 모든 연구자 및 현직 과학 교사와 과학교육 전공 대학원생들로 구성된 집단 세미나를 통하여 그 범주의 적절성과 타당성을 논의하였다. 그 결과 문항 제작 단계별 하위 범주의 일관성과 명료성이 부족하고 개수도 다르며, 각 단계 내에서도 일부 하위 범주의 경우에는 다른 하위 범주와의 위계와 수준이 다르며, 지필평가 문항 제작 과정에 영향을 미치는 요인은 특징이라기보다는 특징이 나타난 원인으로 보아야 한

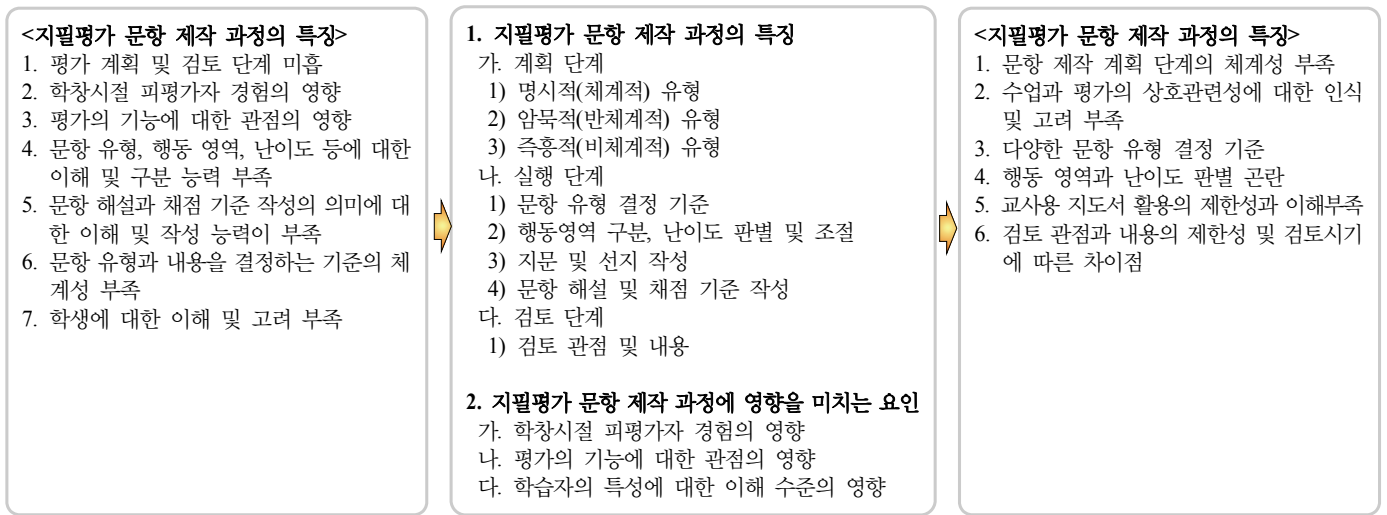


Figure 1. Process of change in analysis category

다는 지적이 있었다. 이런 지적을 토대로 지필평가 문항 제작 과정에서 나타난 특징을 중심으로 범주를 재구성하였다. 또한 분석자 2인이 모든 수집 자료들을 다시 분석하여 하위 범주의 적절성과 타당성을 점검한 후 범주를 정교화시켰다. 이러한 과정을 반복하여 최종적으로 6가지 하위 범주를 확정된 후, 각각에 대하여 구체적으로 예시하고 논의하였다. 주요 범주의 변화 과정은 Figure 1에 정리하였다.

각 범주에 대한 논의 내용은 사후 면담 자료, 예비교사들이 출제한 문항 자료와 해당 수업의 과제로 제출한 개인별 강의 수강 저널 등 수집한 모든 자료들과 지속적으로 비교하여 정당화하는 삼각측정의 과정을 거침으로써 결론의 타당성을 확보하고자 하였다. 또한, 연구의 타당성과 신뢰성을 높이기 위하여 모든 연구자, 현직 과학 교사, 과학교육 전공 대학원생들로 구성된 세미나를 통하여 연구 결과의 해석과 논의의 적절성과 타당성을 점검받아 수정하였다.

III. 연구 결과 및 논의

수집된 자료를 분석한 결과 예비교사의 지필평가 문항 제작 과정의 특징으로 문항 제작 계획 단계의 체계성 부족, 수업과 평가의 상호관련성에 대한 인식 및 고려 부족, 문항 유형 결정 기준의 다양성, 문항의 행동 영역과 난이도 판별 곤란, 교사용 지도서 활용의 제한성과 교사용 지도서에 대한 오해, 검토 관점과 내용의 제한성 및 검토시기에 따른 차이점 등 6가지가 나타났다. 각 특징별 구체적인 내용과 논의는 다음과 같다. 참고로, 이 특징들은 일부 예비교사들이 보인 특징적인 부분을 추출한 것이므로, 이를 감안하여 논문의 결과 및 해석을 이해해야 할 것이다.

1. 문항 제작 계획 단계의 체계성이 부족하다.

교사의 학생평가 전문성 기준(Kim et al., 2012; Nam et al., 2006)에 의하면, 교사는 학습 목표에 부합하는 평가 도구를 선택 및 개발할 수 있는 능력을 갖추어야 한다. 따라서 양질의 평가 도구를 개발하기 위해 교사는 평가 문항 제작 전 우선적으로 학습 목표에 근거하여 체계적으로 출제 계획을 수립하는 과정을 거쳐야 한다. 그러나 이 연구에 참가한 예비교사 8명중 6명은 단원 전체에 대해 학습 목표에

기초한 평가 계획을 체계적으로 수립하지 않고 즉흥적이고 직관적으로 개별 문항을 출제하는 경향이 있었다. 즉, 이들은 제공된 교과서와 교사용 지도서의 해당 단원 내용을 차례대로 넘겨보다가 주관적 판단으로 적당하다고 생각하는 문항을 한 개씩 제작해가며 부과된 총 문항 개수를 맞춰가는 모습을 보였다. 예를 들어, 예비교사 A는 교과서의 내용을 순서대로 읽다가 본인이 중요하다고 판단되는 내용을 발견하면 해당 내용을 묻는 문제를 출제하였다. 예비교사 B는 문항마다 내용 영역 결정에 꽤 많은 시간을 들였으나 구체적인 계획 없이 ‘포뮬러 내야 될까?’라는 말만을 반복하며 초조해하다가 즉흥적으로 내용과 문항 유형을 결정하는 모습을 보이기도 하였다.

선택형 두 문항을 만들어 볼까. 뭘 만들어 보지? (교과서를 보면서) 먼저 사를의 법칙은 어디서 나오지? 응. 사를의 법칙은 기본적으로 압력이 일정할 때 온도에 따른 부피 변화를 보는 거니깐 먼저 압력에 따른 압력이 일정할 때 온도에 따른 부피 변화를 한다는 걸 알아야겠네. 기본적인 것부터 한번 내볼까? 온도에 따른 기체의 분자 운동. 그래 풍선을 예로 한번 들어서 문제를 한번 내보자. 풍선이란 273도에서의 개념까지 아는 지 한번 물어보려면 문제를 어떻게 내야 되지?

(예비교사 A의 발성 사고 내용 중에서)

예비교사: 뭘 만들어야 되지? 어떤 게 좋은지 서답형부터. 아니다 아까 그게 선택형이었나? 그럼 이걸 서답형으로 해야 되나? 허, 어떤 문제를 만들어야 되나? 아 애들 수준을 모르겠는데? 사를의 법칙, 허, 문제를 뭘 내야 되지?

면 담 자: 지금 무슨 생각하는 거예요?

예비교사: 문제를 뭘 내야 될지 모르겠어요. 음 사를의 법칙을 또 어떤 문제를 내야 되냐. 문제집 같은 게 있었으면 도움이 될 텐데. 어떤 걸 내야 되지? . . . (생략) . . . 어떤 문제를 내야 좋은 걸까? 음, 사를의 법칙 음, 개념을 물어보는 걸 해볼까? 아, 보자. 아. 일단 문제를 그냥 선택형으로 해야겠다.

(예비교사 B의 발성 사고 내용 중에서)

한편 예비교사 C의 경우 문항 제작 전에 단원 전체 내용을 검토한 측면에서는 위의 두 예비교사에 비하여 좀 더 체계적인 접근을 시도했다고 볼 수 있지만, 학습 목표에 근거하여 출제 계획을 수립하지 않았다는 점에서는 차이가 없었다. 이는 예비교사 C의 사후 면담 자

료를 통해서도 확인할 수 있었다. 즉, 그는 문항 제작 과정을 전체적으로 설명해달라는 면담자의 요구에 대하여 제공된 교육과정 자료의 내용을 전체적으로 살펴보다가 ‘그냥 갑자기 눈에 띄는 것’을 문항 소재로 활용하였다고 대답하였다.

교과서에 있다는 거는 애네가 배웠다는 거잖아요? 대부분 배웠다는 거니까 애네가 뭘 배웠는지 한번 찾아봤어요. 뭘 배웠는지 찾아보고 어떠한 소재를 문제로 내면 좋을지를 찾아봤는데... 이거를 배웠다고 해서 이거를 바로 문제로 만들 수 없는 것도 있고. 뭔가 여러 가지를 결합해야 뭔가 문제로 만들 수도 있는 게 있는 것도 같아서 꺼리가 뭐가 있는지 봤는데, 교과서에 쓰여 있는 거 보고 그냥 갑자기 눈에 띄는 거를 주로 보고 만들어봤어요.

(예비교사 C와의 사후 면담 내용 중에서)

이에 비하여 주목할 만한 대조적 사례가 있었다. 즉, 학습 목표를 확인하면서 각 문항의 체제와 세부 내용에 대한 전체적인 출제 계획을 총괄적으로 수립하는 예비교사가 2명 있었다. 이들은 세부 문항 제작 전에 제공된 교과서 2종과 지도서 2종 자료를 모두 검토하여 공통으로 수록된 학습 목표 및 내용을 확인하고 이를 근거로 문항의 유형 및 학생 수준을 고려하여 전체 문항에 대한 세부적인 출제 계획을 수립하는 것으로 나타났다.

여기서 학생들이 알아야 할 게, (학습 목표를 확인하며) 기체의 온도와 부피의 관계니까 압력이 일정할 때, 온도가 높아지면 기체의 부피가 일정하게 증가한다 이거 하나를 알아야 되고, 아... 분자의 운동 모형으로 설명할 수 있어야 하니까... (생략)... 이 두 가지를 알면 되는데. 혹시 내가 빠뜨린 부분이 좀 있을 수 있으니까 지도서를 한번 볼까? 이거인데, (지도 목표를 확인하며) 온도에 따른 기체 부피 변화 측정하게 하고, 그래프, 그릴 수 있게 하고, 실험 결과 그래프를 해석하여 온도와 기체 분자 운동을 설명... 음. 그러면, 1번 선택형을 쉽게, 온도와 기체, 온도와 기체 부피 사이의 관계를 알면 풀 수 있는 문제를 하나 만들고, 분자 운동으로 설명하게 하는 거로 서답형을 하나 만들고, 좀 어려운 문제로 그래프를 주고 거기서 그래프를 해석해서 찾아낼 수 있는 문제를 만들어야겠다. 이 실험을 마지막 거에서 내면 좋은가?

(예비교사 D의 발성 사고 내용 중에서)

교과서를 좀 읽어 봐야겠다. (학습 목표를 확인하며) 온도와 기체의 부피 관계를 기체 분자 운동 모형으로 설명할 수 있다. 어쨌든 샤를의 법칙이 온도랑 부피 관계를 설명한 거니까, 이것을 학습 목표를 달성할 수 있게 하면 되는 거지... (생략)... 음, 다른 교과서는 마찬가지로 (학습 목표를 확인하며) 샤를의 법칙을 이해하고 분자 운동으로 이걸 설명할 수 있다. 그러니까 이 두 가지가 제일 중요한 것으로 생각해서 내야겠네. 음 지도서를 보면 실험을 하거나 생활 예시를 찾아보거나 그런 게 있으니까 실생활 문제도 하나 내고 그리고 개념, 실생활 서술형하면 되겠다.

(예비교사 E의 발성 사고 내용 중에서)

이러한 출제 계획 과정을 거치게 된 이유에 대한 사후 면담 질문에, 예비교사 D는 교과서 전체에서 문항을 고르게 출제하고 난이도를 조절하기 위해서라고 응답하였다.

면담자: 전체를 놓고 이거 한 문제, 이거 한 문제, 이거 한 문제 내야 되겠네? 이런 식으로 진행하시더라도요. 혹시 그런 아이디어를 어디서 얻으신 거예요?

예비교사: 아~. 왜냐면 중요한 걸 놓칠 수도 있고 그리고 또 그런 계획을 안세우면 너무 한 문제에 모든 개념이 다 들어가거나 나중에는 그런 개념을 넣어야 되는데 못 넣거나 뭐 이럴 수도 있을 거 같아서. 난이도 조절할 때도 그렇고, 이 개념은 어려우니까 뒤에 가서 더 집중적으로 물어봐야지 하고? 앞에 쉬운 문제 계획하고.

(예비교사 D와의 사후 면담 내용 중에서)

평가의 기준은 교육 활동을 통해 달성하고자 하는 학습 목표나 학생의 변화된 행동 목표에 두어야 하므로, 평가 도구의 질을 제고하기 위해서는 기본적으로 학습 목표에 대한 분석 작업이 선행되어야 한다(Park, 2012). 즉, 학습 목표에 근거한 출제 계획 수립 과정은 평가 도구의 타당도를 높이고 난이도를 조절하는 데 매우 중요하다고 할 수 있다. 그럼에도 불구하고 많은 예비교사들이 이러한 과정을 거치지 않았던 것은 그 과정의 중요성에 대한 인식이 부족했기 때문일 수 있다. 이는 결과적으로 질 높은 문항을 제작하는 데 저해 요소로 작용할 수 있으므로, 이를 개선하기 위한 방안을 마련할 필요가 있다.

2. 수업과 평가의 상호관련성에 대한 인식 및 고려가 부족하다.

연구 참여 예비교사 8명 중 5명은 발성 사고 과정에서 평가와 수업의 연관성에 대하여 전혀 언급하지 않았다. 심지어 예비교사 G는 사후 면담에서 실험 상황을 이용한 평가 문항에 대해 수업 시간에 실험을 수행한 것을 전제로 문항을 출제한 것인지를 묻는 면담자의 질문에 수업 시간에 실험을 실시했는지 여부는 상관없다고 답하였다. 실제 학교 현장에서는 교과서에 제시된 실험을 수행하지 않고 다른 방법으로도 실험 내용에 대한 학습이 이루어질 수 있지만, 예비교사 교육과정에서는 교과서 흐름이나 상황대로 수업이 이루어졌음을 가정하고 문항을 출제하는 것이 바람직하다. 이런 점에서 볼 때 이 예비교사는 평가 문항 제작 과정에서 수업 상황에 대한 고려가 부족했다고 볼 수 있다.

면담자: 직접 실험을 해본거야? 해본 걸 전제하는 거야?

예비교사: 굳이 하지 않았을 때도. 네. 그냥 교과서에서만 딱 실험하지 않고 배웠을 때를 가정하고 문제를 내는 거예요.

면담자: 배운 걸 가정하더라도 똑같고 안 배웠어도 똑같고?

예비교사: 네네.

(예비교사 G와의 사후 면담 내용 중에서)

이와는 대조적으로 3명의 예비교사는 수업과 평가의 연관성을 고려하는 발언을 하였는데, 다음 사례와 같이 주로 수업 시간에 다루었던 내용을 평가 문항에 반영하려고 시도하는 과정에서 나타났다.

평균 운동 에너지까지 나오네. 그렇지 평균 운동 에너지. 아~ 그러면. 그렇지. 교과서에는 중학교 수준이니까 운동 에너지 같은 평균 운동 에너지 같은 말이 안 나왔는데, 내가 이것을 수업 시간에 했나? (교과서를 보고 해당 내용이 없음을 확인한 후) 그러면 이 얘기는 쓰면 안 되나? ... (생략) ... V는 273분의 1, V_0 , T, 그렇지. 이 식을 내가 수업 시간에 했을 거야. 그러면 아이들은 0도씨 때 부피의 273분의 1씩 늘어난다는 걸 알게니까.

(예비교사 D의 발성 사고 내용 중에서)

면 담 자: 그럼. 이 그림은 어디서 가져온 뭐 출처가 있어요?
 예비교사: 여기 실험 내용에 있는데, 이 실험을 했다고 생각하고 냈어요.
 면 담 자: 그러니까 수업에서 다 했었다고 생각하고요?
 예비교사: 네. 수업에서 한번 뭐 설명을 했거나 실험을 직접 했거나 이렇게 해서 했다고 생각하고 설명을 할 수 있는지.
 (예비교사 E와의 사후 면담 내용 중에서)

위 사례에 해당되는 예비교사들은 평가 내용 영역을 결정할 때 수업과 평가의 연계성을 주요 요소로 고려하였는데, 여기서의 연계성은 단순히 수업한 내용을 평가에 반영하는 낮은 수준의 ‘연계성’을 고려한 것이라고 할 수 있다. 하지만 교실 평가는 학생에 대한 이해를 돕기 위한 ‘진단적 의사결정’, 성적 부여 및 피드백을 위한 ‘성적에 대한 의사결정’, 수업 설계 및 진행을 위한 ‘교수적 의사결정’ 등 다양한 측면의 의사결정을 위한 과정으로 수업의 모든 과정에 연계성을 가지며, 교수 도구 자체로써(as instruction) 유용하게 활용될 수 있다(McMillan, 2014). 따라서 평가와 수업이 맺는 다양한 측면의 상호작용에 대해 이해하고 이를 활용할 수 있는 방안에 대해 생각해볼 기회를 예비교사 교육에서 제공할 필요가 있다.

다음은 수업과 평가의 상호관련성에 대하여 또 다른 측면을 고려한 예로, 예비교사 D는 수업 내용을 평가에 반영해야한다는 것과 함께 평가가 수업에 미치는 영향에 대해 언급하였다. 즉 예비교사 D는 사후 면담에서 서답형으로 출제된 문항이 피험자가 답하기에 다소 까다롭지 않겠냐는 면담자의 질문에, 해당 내용이 평가 계획에 포함되어 있었다면 수업 시간에 이 내용을 강조하고 중점적으로 다루었을 것이라고 답하였다.

면 담 자: 애들이 이렇게 쓸까? 이걸 다 쓸 수 있을까? 이렇게 얘기를 했었잖아요?
 예비교사: 네.
 면 담 자: 그거에 대해서 어떻게 생각하세요? 애들이 막 진짜 이거 3~4 줄씩 막 쓰나?
 예비교사: 근데 그러면 제가 만약에 이렇게 문제를 낼 거라고 생각을 하면, 수업 시간에 좀 강조를 했을 것 같아요.
 면 담 자: 어떻게?
 예비교사: 그 단계, 이런 단계가 있어. 온도가 높아졌지, 그럼 활발해질 거야. 그럼 활발해지면 충동을 많이 하니까, 압력이 증가해서 부피가 커졌네? 이런 단계를 잘 이해하고, 잘 기억하고 있어 ~ 이렇게. 이런 식으로 수업시간에 좀 더 강조를 하고, 애들이 충분히 이거에 대해서 집중적으로 공부를 할 수 있도록.
 (예비교사 D와의 사후 면담 내용 중에서)

평가에 대한 이와 같은 접근은 수업 활동 및 내용을 계획하기 전에 평가 계획을 먼저 수립한다는 측면에서 소박한 수준의 백워드 설계(McTighe & Wiggins, 2004; Wiggins & McTighe, 2005) 관점을 가지고 있다고 볼 수 있다. 전통적인 수업 설계 모형들이 수업 목표 확인 후 학습 경험이나 내용의 선정과 조직 후에 평가 방법을 고안했던 것에 반해, 백워드 설계는 수업 목표 확인과 동시에 평가를 고려하여 수업 활동을 계획하는 설계 모형으로서 평가 전문성의 중요성이 강조되는 모형이다(Wiggins & McTighe, 2005). 또한 평가의 수동성을 벗어나 평가를 수업의 모든 측면과 내적으로 연계되어 있는 활동(McMillan, 2014)으로 인식한다는 점에서, 학생들의 학습 향상과 교

수 및 학습 개선을 위해 결과보다는 상황과 과정에 중점을 두는 구성주의적 평가관(Jonassen, 1991)과도 맥을 함께 한다. 이런 관점에서 볼 때 대부분의 예비교사들이 평가와 수업의 역동적인 상호작용 관계를 인식하지 못하는 것으로 나타난 점은 개선되어야 할 것이다.

3. 문항 유형 결정 기준이 다양하다.

예비교사들은 다양한 기준으로 문항 유형을 결정하는 특징이 있었다. 즉 예비교사들이 문항 유형을 결정할 때 주로 고려한 요소로는 문항의 난이도, 평가 시행의 효율성, 검사요령(test wiseness) 효과, 평가를 통한 학습 효과 등으로 나타났으며, 소재의 특성 등에 근거하여 문항 유형을 결정한 경우도 있었다.

구체적으로 살펴보면, 예비교사들은 문항의 난이도를 조절하기 위해 학습자의 특성을 고려하여 문항 유형을 선택형이나 서답형으로 결정하였다. 예를 들어, 예비교사 D는 서답형과 선택형 중에서 고민하다가 선택형으로 출제하면 문항의 난이도가 너무 낮아지고, 이는 열심히 시험을 준비한 학생들의 학습 의욕을 저하시키는 결과를 초래할 수도 있다고 생각하여 서답형을 선택하였다.

면역에 이 문제를 서답형으로 만들었다가는. 그럼, 아니야. 근데 이 문제를 또 객관식으로 만들면 너무 쉬운 문제가 되는데? 어떡하지? 음, 문제를 너무 쉽게 만들었다간 열심히 공부하는 학생들이 억울 할거야. 음, 그러면 그럼 일단 서답형, 단답이 아니라 서답이니까.
 (예비교사 D의 발성 사고 내용 중에서)

예비교사 C는 사후 면담에서 해당 소재를 서답형으로 출제했다면 답을 제대로 쓴 학생들이 많지 않을 것이고 학생들이 서답형을 어려워하기 때문에, 선택형 문항으로 출제했다고 답하였다.

서답형으로 했으면 거의 이것과 거의 유사하게 애가 글을 써야 되는데 그렇게 되면 이걸 제대로 이해하고 있는 학생들만 쓸 것 같고. 사실 그냥 이 기체의 온도와 부피와의 관계만 그냥 알고 있는 학생들은 이걸 쓰기 어려운 것 같아서 서답형을 쓰면 난이도가 되게 높아질 거라는 생각했어요. 그래서 차라리 서답형을 줄 거면 그냥 제 생각에 서답형은 일반적으로 선택형보다 어려운 것 같으니깐? 그래서 서답형을 할 거면 쉬운 소재로 주고 그래서 차라리 애를 이런 식으로 선택하면 그나마 애들이 좀 풀 수 있지 않을까.
 (예비교사 C와의 사후 면담 내용 중에서)

일반적으로 학생들은 서답형보다 선택형 문항을 더 친숙하고 쉽게 생각하며 선택형 문항의 성취도도 더 높게 나타나는 경향이 있으므로(Park, 2003) 두 예비교사의 판단은 타당하다고 볼 수 있다. 하지만 선택지 구성 방식이나 내용에 따라 선택형이 서답형에 비해 난이도가 높게 출제될 수도 있다. 따라서 예비교사 교육과정에서 문항의 난이도와 문항 유형을 단순하게 연계하여 사고하는 것을 지양하고, 평가 상황이나 의도에 적합한 문항 유형을 선택하여 활용할 수 있도록 안내할 필요가 있다.

다음으로 평가 시행의 효율성을 고려하여 문항 유형을 결정한 경우가 3명의 예비교사에게서 나타났다. 예를 들어 예비교사 F는 출제 및 채점의 용이성으로 인하여 서답형 문항을 선택하였다. 그리고 예비교사 B는 서답형 문항의 경우 유사 답안 및 부분 정답 등이 가능함

에도 이에 대한 이해가 부족하여 서술형보다는 답안 작성과 채점이 용이한 단답형으로 출제해야한다고 생각하였다. 또한 예비교사 A는 한 문제로 다양한 내용에 대한 이해를 확인할 수 있어 내용적 포괄성이 큰 선택형 문항으로 출제하기도 하였다.

서답형 문제를 내려면 어떤 걸 하면 좋을까, 음, 빈칸 넣기도 괜찮은 거 같고, 예를 설명하는 것도 좋을 거 같은데. 서답형 문제는 깔끔하게 답이 나오는 단답형 문제로 하면 좋겠다.

(예비교사 F의 발성 사고 내용 중에서)

면 담 자: 서답형으로 이 문제를 만든 이유가 있는지요?

예비교사: 다른 걸 서답으로 내는 것보다 그냥 이거를 하고 싶었어요. 딱히 이유는 크게 없고, 그냥 내기도 쉽고, 채점하기도 쉽고, 그냥 학생들 알고 있는지 아닌지 평가하기도 쉬운 거 같아서.

(예비교사 F와의 사후 면담 내용 중에서)

서답형이 딱 이렇게 정해져 있지 않지만 그 틀은 뭔가 이게 확실히 이게 답이라고 하는 게 있어야 하는데 뭐 이 문제 같은 경우는 쓰기가 애매했어요. 만약에 처음에 문제가 외부 압력이 일정하면 상태의 기체의 온도가 올라가면 일정한 상태에서 기체의 온도가 올라가면? 뭐 활발해진다, 활발해지고 횡수가 증가하고, 기체 압력이 증가하고, 이거를 서답형으로 다 쓰기엔 뭔가 빠질 수도 있는 거 같아요. 뭔가 간단하게 쓸 수도 있는 거 같아요. 근데 그래 가지고 서답형으로 할 수 없다고 생각했어요.

(예비교사 B와의 사후 면담 내용 중에서)

서답형으로 하면 이 많은 걸 물어보기가 조금, 서술형처럼 되니까 서답형 같은 경우에 딱 떨어져서 쓰는 거 같아요. 그래서 (선택형으로 출제하면) 좀 더 많은 내용을 물어볼 수 있을 거 같아서.

(예비교사 A와의 사후 면담 내용 중에서)

한편, 학생들이 개념을 정확하게 이해하지 않고도 추측해서 맞는 검사요령 효과를 고려하여 문항 유형을 결정하는 경우도 있었다. 예를 들어, 예비교사 B는 선택형 문항의 경우 검사요령 효과가 작용할 가능성이 있으므로 선택형보다 서답형 문항이 적용 능력 평가에 더 효과적이라고 생각하여 서답형 문항을 선택하기도 하였다.

(선택형으로 출제하면) 이걸 보기를 보면서 그냥 찍어서도 풀 수 있다고 해야 되나? 사들의 법칙 대충만 알아도 보기 문제를 보고는 그냥 대충 어느 정도 이거구나 풀 수 있을 거 같은데 여기서 제가 하는 거는... (생략)... 이거를 그냥 자기가 뭐 있었던 일이나 자기가 뭐 살면서 이렇게 그런 환경에 적용시켜 가지고 문제를 이렇게 쓰는 거 같아요. 차라리 그게 더 창의적이고 이해를 더 잘 적용을 더 잘 시키지 않았을까 하는 생각?

(예비교사 B와의 사후 면담 내용 중에서)

평가를 통한 학습 효과를 고려하는 경우도 있었다. 예비교사 H는 선택형으로 출제된 이유로 선택지를 구성할 때 학습 효과를 고려하였다고 말하였다. 즉 선택형 문항의 오답지를 제작할 때, 학습 효과를 얻을 수 있는 내용으로 지문을 구성하였다는 것이다. 선택지의 학습 효과에 대해서는 예비교사 C의 생각도 비슷하였다.

면 담 자: 근데 이제 예시를 찾는 문제를 이렇게 선택형 문항으로 하게

된 이유는 뭐예요?

예비교사: 네, 뭐 서답형으로 할까 선택형으로 할까 고민을 했는데 서답형으로 하면 이거 모르는 애들은 아예 못쓰잖아요. 근데 만약 선다형으로 하면 그래도 이 네 개는 확실히 아이들이 알 것 같아서. 그니까 예를 더 많이 알 수 있게 할 거 같아서.

(예비교사 H와의 사후 면담 내용 중에서)

면 담 자: 선택형에서는 어떤 걸 좀 중요하게 고려해서 만들어야 할 것 같아요?

예비교사: 선택형은 일단 문제에서 물어보는 질문이 명확해야 될 것 같고... (생략)... 만약에 옳지 않은 것은?이라고 했으면 맞는 4개가 좀 뭐라 해야 하지? 의미 있는? 예를 들면 교과서에서 배운 것과 거의 유사하게?

면 담 자: 왜 좀 유사하게 내야한다고 생각해요?

예비교사: 문제를 보고 애가 나중에 채점을 하고, 다른 거를 공부할 때, 적어도 1, 2, 4, 5번은 맞는 거니까, 문제 잠깐 봤는데 이걸로 가볍게 복습을 할 수 있는? 그런 용도로 쓰일 수 있지 않을까.

(예비교사 C와의 사후 면담 내용 중에서)

서답형 중에서 괄호 채우기와 같은 단답형 문항은 출제와 채점이 쉬운 장점이 있으나, 학생들의 다양한 능력을 효과적으로 측정하는데 한계가 있다. 반면 서술형 문항과 같은 서답형 문항은 출제 시간을 절약할 수 있는데 반해 채점 시간이 매우 많이 소요되지만 학생들의 개념 이해 수준 및 오개념 파악, 창의성 및 탐구력 등과 같은 다양한 능력을 파악하는 데 유용한 문항 유형이다. 그리고 선택형 문항은 지식과 탐구, 사고력 등을 포함한 다양한 영역에 대한 폭넓은 평가가 가능하고 채점이 쉽지만, 타당하고 매력적인 선택지 구성 능력 및 시간이 요구되고 추측을 통해 맞힐 가능성이 있다(Baek, 2000). 또한 선택형 문항의 선택지를 구성할 때 명확한 정답지와 학습 효과를 기할 수 있는 오답지를 함께 제시한다면 평가 과정에서 학습이 진행되는 효과를 얻을 수도 있다. 위의 사례를 통해 예비교사들은 이러한 문항 유형별 특성에 대해 어느 정도 이해하고 있음을 확인할 수 있다. 특히, 선택형 문항의 학습 효과를 인지하고 활용하려고 하는 것은, '선택형 후평가'라는 전통적 평가관의 직선적 도식에서 탈피하여 학습과 평가의 역동적 상호작용 관계를 구현하려는 구성주의적 평가관에 기초하여 문항 유형을 결정하고 구성하려고 했다는 점에서 매우 유의미한 시도로 생각된다. 반면, 평가 시행의 효율성 측면 중에서 출제 및 채점의 용이성만을 지나치게 고려하여 문항 유형을 결정하는 것은 문항의 질을 확보하는데 저해 요소로 작용할 수 있다. 따라서 예비교사 교육과정에서 해당 기준들의 유의미한 적용 및 선정 방법에 대하여 체계적이고 통합적으로 고민하는 기회를 제공할 필요가 있다. 또한, 유사 답안이나 부분 정답의 발생 가능성에 대한 고려나 허용이 부족했던 점을 개선하기 위한 노력도 필요하다.

소재의 특성에 근거하여 문항 유형을 결정할 경우는 3명의 예비교사에게서 나타났다. 예를 들어, 예비교사 G는 문항 소재로 실험 등의 상황이 주어지는 경우는 서답형, 예시 문제의 경우는 선택형이 적합하다고 하였다. 예비교사 E는 샤를 법칙에 대한 일상생활 속 사례 문제를 서답형으로 출제하기로 계획했었지만, 실제 문항 제작 과정에서 선택형으로 변경하였는데, 그 이유로 소재의 특성을 고려했다고 답하였다. 즉, 소재가 현상을 설명하는 경우는 서답형, 예시가 많은 경우는 선택형으로 출제하였다는 것이다.

첫 번째 문제는 사를의 법칙을 이해하고 실생활에 적용할 수 있는지 (서답형으로 출제하여) 알아보자. . . . (생략) . . . 그러면 교과서에 나온 거는, (교과서와 교사용 지도서의 사례를 보며) 그러면은 아예 여러 가지 예시를 내고 선택형으로 내는 것도 괜찮을 것 같아.

(예비교사 E의 발성 사고 내용 중에서)

면 담 자: 서답형 문항으로 애를 선택한 이유가 특별히 있나요? 이 문제 상황을?

예비교사: 이 기체분자운동에 관련해서 선택형을 내면은 다섯 개 선택지를 보고 이렇게 고르는 거잖아요. 그런데 선택지에 뭐 옳은 설명이든 틀린 설명이든 아니면 이거랑 관련된 예시든 이렇게 해서 하면은 너무 쉬워지지 않을까라는 생각을 했어요. 계속 같은 얘기를 반복해야 되거나 그래서 이거보다는 차라리 알고 있는지 설명을 하는 게 낫지 않을까 해서 애(기체분자운동론)를 설명을 하고, 예시가 좀 더 많다고 느껴지는 거를 앞에 선택형으로 쓰고 했어요.

(예비교사 E와의 사후 면담 내용 중에서)

예비교사 D는 계산 문제의 경우 선택형으로 출제하면 선택지를 활용하여 학생이 자신의 계산 결과에 대한 의미 있는 검토 과정을 거칠 수 있기 때문에 계산 문제를 선택형 문항으로 출제하였다고 하였다.

1번 선택형을 쉽게, 온도와 기체 부피 사이의 관계를 알면 풀 수 있는 문제를 하나 만들고, . . . (생략) . . . 잠깐만. 이거 온도 나타내는 것도 평가하고 싶은데, 절대 온도는 섭씨온도의 273을 더... 아 근데, 그 어차피 계산을 하려면 이거를 알고, 273을 더한 온도 값을 사용해야 하나까 음 그러면 오답 선택지에 켈빈 온도를 사용하지 않은 그냥, 섭씨 온도로 해서, 계산해서 나올 수 있는 오답도 넣어야겠다.

(예비교사 D의 발성 사고 내용 중에서)

면 담 자: 이걸 왜 선택형으로 만드셨어요?

예비교사: 이 문제를 이렇게 선택형으로 만들라고 하지 않아도 이렇게 할 거 같아요. 왜냐면 어... 그냥, 저, 그니까 예를 들어서, 딱, 만약에 잘못 계산을 했어요. 그러면 다른 보기들이 없으면 그냥 예를 들어 처음에 했던 6리터, 이렇게 적으면 이거에 대해 난 맞겠지 하고 넘어갈 텐데, 보기가 있으면 아 잠깐만 애는 왜 이렇게 나왔지? 혹시 내가 잘못 계산한 건가? 하고 맞았던 친구들도 다시 계산할 수도 있고. 틀린 친구들은 어? 이거 잠깐만 좀 아닌 거 같은데 다시 해볼까라고 다시 생각을 고쳐서 할 수 있기 때문에 이걸 선택형으로 만들었을 거 같아요.

(예비교사 D와의 사후 면담 내용 중에서)

소재의 특성 또한 문항 유형을 결정할 때 고려할 수 있는 요소이긴 하나 다른 측면보다 우선적으로 고려해야 하는 요소라고 보기는 어렵다. 즉, 소재의 특성을 문항 유형의 핵심 기준으로 삼기보다는 앞선 사례에서 나타난 문항의 난이도, 평가 시행의 효율성, 평가를 통한 학습 효과 등과 통합하여 고려하는 것이 바람직하다고 할 수 있으므로, 이에 대한 교육이 필요하다.

교사는 평가 도구의 교육적 가치를 이해하고 평가 목적과 내용에 적합한 도구를 개발하거나 선택하여 사용할 수 있어야한다(Nam et al., 2006). 특히 평가 문항의 유형을 결정하는 데는 학습 목표의 성격, 객관적 채점의 중요성, 학생들의 공부와 동기에 주는 영향 등 우선적

으로 고려해야할 요소가 많으므로 이에 관한 능력을 지니고 있어야 한다(McMillan, 2014). 일반적으로 선택형은 평가의 공정성과 객관성이 강조될 때, 문항 제작자가 선택형 문항 제작에 더 자신이 있을 때, 문항을 채점할 시간보다 제작할 시간이 더 많을 때 활용할 수 있는 반면, 서답형은 글 표현 능력을 강조 및 향상시키려고 하거나 학업 성취도보다 태도와 의견에 더 관심이 있을 때, 신속한 결과 확인이 필요치 않아서 채점에 시간적 여유가 많을 때 활용할 수 있는 문항 형태이다(Park, 2012). 이에 비추어 볼 때 이 연구에 참여한 예비교사들이 다양한 요소들을 고려하여 문항 유형을 결정하는 것은 긍정적이거나 이러한 기준이 통합적으로 적용되지 못하고 특정 측면에 한정하여 문항 유형을 결정한 것은 질 높은 평가 문항을 제작하는 데 저해 요소로 작용할 수 있다. 따라서 예비교사들이 다양한 요소들을 통합적으로 고려하여 지필평가 문항을 제작할 수 있도록 예비교사 교육과정에서 강조할 필요가 있다.

4. 문항의 행동 영역과 난이도 판별에 어려움이 있다.

예비교사들은 행동 영역과 난이도를 먼저 결정한 후 문항을 제작하기도 하고, 문항을 먼저 완성한 후에 행동 영역과 난이도를 분석하기도 하였다. 즉, 어떤 경우든지 예비교사들은 학생 수준에의 적합성을 고려하여 난이도를 판단하고, 문항 내용과 행동 영역의 일관성을 고려하여 행동 영역을 구분하려는 모습을 보였으나, 이 과정에서 많은 어려움을 겪는 것으로 나타났다. 예를 들어, 행동 영역 판별의 경우 예비교사들은 주로 지식과 이해, 이해와 적용의 구분을 어려워하거나 각 영역에 대한 이해가 부족하고, 심지어는 명확한 판별 기준이 없이 즉흥적으로 판별하는 모습을 보이기도 했다. 다음 사례에서는 사를 법칙의 생활 속 사례 찾기 문제를 ‘이해’로 판별할 수 있지만 학창 시절의 피험자 경험 및 과외 학생의 문제 풀이 관찰 경험을 통해 암기에 의한 문제 풀이 가능성을 염두에 두고 ‘지식’으로 볼 수도 있어 고민하고 있는 상황을 보여주고 있다.

학생이 사를의 법칙 생활 속의 예를 알고 있는지 평가하겠다. 이걸, 지식의 이해인가? 적용? 적용은 아닌 것 같고, 행동 영역이 지식인가? 이해, 이해인가? 이해? 요부분 다 외워서 풀 텐데? 이런 거 보면 지식인가? 이해인가?

(예비교사 A의 발성 사고 내용 중에서)

사를의 법칙이 이제 생활 속에 적용되는 예를 알고 있는지. 해서 사실 이해라고 했지만 사실은 생각해보면 지식적인 문제도 가까운 거 같아요. 왜냐면 제가 학원 알바도 해봤고, 또 그 학교... 제가 학생일 때도 보면 예 같은 거는 대충 웬만히 정형화된 거는 다 외우고 들어간단 말이에요. 시험 볼 때?

(예비교사 A와의 사후 면담 내용 중에서)

다음은 ‘이해’와 ‘적용’ 영역을 판단하는 것의 어려움을 호소하는 경우이다.

이렇게 해놓고, 지식? 이해? 적용? 이해랑 적용이랑 되게 애매한 문항인데. 이해, 적용. 문제를 만들고 생각을 할까? 이해 적용. 실험 내용을 설명 할 수 있다. 설명을 하려면. 설명하는 건 적용인데? 이해. 개념을 설명하는 거면 이해인가? 이해가 맞나? 아 배운 걸 적용인데. 이해.

이해가 맞는 것 같아.

(예비교사 E의 발성 사고 내용 중에서)

면 담 자: 이거 행동 영역 중에 '이해'로 체크한 특별한 이유가 있어요?

예비교사: 사실 '이해'랑 '적용'이랑 좀 뭐라 해야 되지 애매한 거 같은데 요. 이해해서 설명하는 게 이 예시에 적용을 하는 건지, 이해하고 설명하는 거는 그냥 '이해'로 그친다고 생각해야 되는 건지 모르겠어서 고민을 하다가 이렇게 했거든요. 앞의 문제도 그렇고.

(예비교사 E와의 사후 면담 내용 중에서)

그리고 예비교사 D는 행동 영역과 난이도를 단순하게 연결시켜 동일한 기준으로 판별하였고, 예비교사 C와 G도 명확한 기준 없이 즉흥적으로 행동 영역을 판별하는 모습을 보였다.

지식, 이해, 적용에 대한 이걸 맞춘다기보다는 학습 목표에서 어떤 문제가 들어가야 되는지 일단 정하고 나서 그 문제가 뭐였는지를 체크하는 걸로 끝났던 것 같아요. . . . (생략) . . . 왜냐하면 당연히 지식이나 이해는 '하고 적용은 '상'이니까.

(예비교사 D와의 사후 면담 내용 중에서)

사실 저도 이거 지식이랑 이해랑 적용이랑 그 이원목적분류표에서 본 것 같은데, 정확히 이게 무엇에 대한 단어인지 제가 지금 정확히 기억이 안 나서.

(예비교사 C와의 사후 면담 내용 중에서)

면 담 자: 이해로 판단한 이유는 뭔가요?

예비교사: 그 뭐지? 지식, 이해, 적용이 기억이 안 나서 잘 모르겠는데, 그 적용은 그 생활 속에서 적용하는 사례를 찾는 게 적용 문제인 것 같고, 지식은 단순히 그 뭐지? 음, 그 식에 대입해서 푸는 문제 그런 거라고 생각을 해서, 그러니까 둘 다 아니어서 '이해를 한 거예요.

(예비교사 G와의 사후 면담 내용 중에서)

행동 영역에서 '이해'와 '적용'은 동일한 내용일지라도 학생의 선행 지식이나 교수·학습 내용에 따라 판단 결과가 달라질 수 있다. 이로 인하여 경력 교사조차도 두 영역을 구분하는 데 어려움을 겪는 것으로 나타났으므로(Park *et al.*, 2011), 예비교사들이 각 영역에 대한 이해가 부족하고 두 영역을 어려워하는 것은 어찌 보면 당연한 일이다. 따라서 예비교사의 행동 영역에 대한 판단 능력을 제고하기 위한 실천적인 방안이 필요하다.

한편 예비교사들은 난이도 판단 과정에서도 어려움을 겪었는데, 그 주된 원인은 학생의 특성에 대한 이해 부족인 것으로 확인되었다. 예를 들어, 다음은 학생에 대한 이해 부족으로 샤를 법칙에 대한 계산 문제의 난이도를 판단하는 것을 어려워하는 사례이다.

면 담 자: 중간에 이런 얘기를 했었어요. 하다가 아 애들 수준을 모르겠네 라고 하면서 연관 지어서 어떤 문제를 내야 되는지 모르겠다 막 이렇게 했는데 기억나요?

예비교사: 아, 네. 난이도예요. 아까 이 문제 푸는 것도 그냥 식에 대입만 하면 되는 거니까 쉬운 건데 그것을 잘 이걸 하라고 해야 될지 모르겠는 거예요. 제 기준에는 그냥 저도 못 풀어가지고 당황 하긴 했는데 너무 떨려가지고 그랬긴 했는데 이 문제가 쉬운

건지 어려운건지 모르겠더라고요. 제가 생각하기엔 정말 쉬운 문젠데 이걸 또 못 푸는 음, 저걸 잘 모르겠어서 애들 수준이 어떻게. 하가 되는지 중이 되는지 상이 되는지 모르겠더라고요.

(예비교사 B와의 사후 면담 내용 중에서)

학생의 특성에 대한 이해는 문항의 난이도 및 변별도 등을 평가하고 조절하는 데 중요한 요소이므로(McMillan, 2014), 학생의 특성에 대한 이해 부족은 문항의 완성도를 저해하는 결과를 초래한다. 따라서 예비교사 교육과정에서 학생의 특성, 특히 평가와 관련된 특성에 대한 예비교사의 이해를 높이기 위해 체계적으로 교육할 필요가 있다.

5. 교사용 지도서 활용이 제한적이고 교사용 지도서에 대한 오해가 있다.

교사용 지도서는 교육과정과 교과서 내용을 안내하고 해설하며, 학습자의 특성과 교수·학습 방법에 관한 정보를 제공하며, 가르치는 내용과 방법의 기준 및 평가의 준거와 자료까지 제시하고 있다(Bismack *et al.*, 2014; Choi, 2006; Davis *et al.*, 2014; Lee & Choi, 2013; Park, 2008). 즉 교사용 지도서는 PCK의 다양한 구성 요소인 교육과정에 관한 지식, 학생에 관한 지식, 교수 전략에 관한 지식, 평가에 관한 지식 등에 관한 다양한 실천적 지식 정보를 제공하고 있다. 그리고 교사의 학생 평가 수행 과정에서는 평가 전문성을 포함한 모든 PCK 구성 요소들의 종합적인 활용이 요구되므로(Falk, 2012; Min, 2012), 예비교사들이 지필평가 문항을 제작하는 과정에서도 교사용 지도서는 매우 유용한 자료라 할 수 있다. 실제로 이 연구에 참여한 예비교사 대부분은 지필평가 문항 제작 과정에서 교사용 지도서에 대한 의존도가 매우 높게 나타났는데, 이는 현직 교사들이 본인이 전공하지 않은 영역을 가르치거나 새로운 내용을 가르칠 때 이런 경향을 나타내는 것(Grossman & Thompson, 2004; Mulholland & Wallace, 2005)과 유사한 결과이다. 하지만 예비교사들은 특정 측면에 제한하여 지도서를 활용하는 경향이 있었다. 예를 들어, 가장 빈번한 경우는 다음 사례와 같이 선택형 문항의 선택지를 제작하기 위한 참고 자료로 교사용 지도서 활용하는 것이었다.

냉장고 관련된 예를 한번 볼까. 어, 이게 어디 있지? 이 책에는 예가 별로 없네. 저 책을 볼까? 예를 보면 어디 있을까. 지도서를 보는 게 예가 더 많이 나올 거 같으니까 지도서를 봐야겠다. 온도와 변화는 기체의 부피 변화와, 여기 있다. 온도에 따른 기체의 부피 변화.

(예비교사 F의 발성 사고 내용 중에서)

또한 예비교사들은 출제하려는 내용이나 용어가 교육과정에 적합한지를 확인하거나, 해당 학년에서 특정 개념을 어느 수준까지 다루어야 하는지를 확인하기 위한 목적으로 교사용 지도서를 활용하기도 하였다. 다음은 교사용 지도서의 '지도상 유의점'에 제시된 내용을 보고 중학교 수준에서는 샤를 법칙 공식을 다루지 않아야 함을 깨달은 사례와 교사용 지도서의 '수업 노하우'를 통해 중점적으로 지도해야 하는 내용을 파악하고 이를 문항 소재 선택에 반영한 사례이다.

(교사용 지도서를 보며) 온도와 부피의 관계, 기체 온도와 부피의 관계를 실험을 통해 이해하도록 해야 하며, 공식은, 어머, 공식 위주로 암기하지

않도록 유의해야 해? 그렇구나.

(예비교사 D의 발성 사고 내용 중에서)

면담자: 생활 속에 적용한 예. 이걸로 이 상황을 문제로 내야 되겠다고 생각한 이유가 뭐죠?

예비교사: 그 아까 보일의 법칙 낼 때 보니까 이 보일의 법칙에서도 그 법칙을 아는 것도 중요한데 그 예시를 아는 것도 중요하니까 그래서 저도 그리고 개인적으로도 과학을 배우는 거는 그 이런 이론을 배우고 이 이론이 우리 생활에서 어떻게 나타나는 지가 중요하다고 생각하기 때문에 이 문제를 내야겠다고 생각했어요.

면담자: 보일의 법칙에서 그게 중요하다고 한다고 했는데 어디서 누가 그렇게 한다고 했던 거예요?

예비교사: 교사용 지도서에 나와 있더라구요. 그 보일의 법칙에 대한 개념을 아는 것도 중요하지만 보일의 법칙이 어떻게 적용되고 이용되는지 이해하고 예를 쓸 수 있게 하는 것도 중요하므로 생활에서 느끼는 보일의 법칙의 예를 다양하게 제시하는 것이 좋다. 이렇게 수업 노하우에 나와 있어서.

(예비교사 G와의 사후 면담 내용 중에서)

교육과정 및 학생에 대한 이해가 부족한 예비교사의 특성상 교사용 지도서를 참고하여 평가 계획을 수립하고 문항을 제작하는 것은 평가 문항의 타당도를 확보하기 위해 예비교사들이 취할 수 있는 합리적인 선택이다. 하지만 선택지 작성을 위한 사례 찾거나 교육과정의 부합성 확인 용도에 제한하여 교사용 지도서를 활용한 점은 추후 개선이 필요하다. 즉 교사용 지도서에는 교육과정 내용과 실생활 적용 사례 이외에도 학생의 오개념, 학생들이 자주하는 질문, 교수·학습 전략 활용 시 유의점과 적용 Tip 및 학생 반응, STS 자료, 과학사 자료, 전문적인 교과 내용 지식, 평가 전문성 관련 지식 등과 같이 지필평가 문항 제작에 필요한 정보들이 다양하므로, 이에 대한 안내와 함께 효과적인 활용 방안에 대하여 지도할 필요가 있다.

한편, 교사용 지도서에 대한 오해가 평가 문항의 타당도를 저해하는 경우도 발견되었다. 예를 들어, 교사용 지도서에 수록된 내용이라면 교육과정의 부합 여부를 확인해보지도 않고 학생들이 모두 이해할 것이라고 가정하거나 수업 시간에 가르쳤을 것이라고 가정하여 지필평가 문항 제작 시 반영하는 경우가 있었다.

이거(교사용 지도서에 제시된 내용)를 한번 괄호를 채워서 써보라고 할까? 아 좀, 중학교 1학년한테 어려운 내용인가? 아닌데, 지도서에 나와 있는 걸로 봐선 중학교 1학년도 충분히 알 수 있는 개념일 것 같다.

(예비교사 H의 발성 사고 내용 중에서)

난이도 '하' 문제로 생각해보면, 압력과 온도가 주어졌고 처음에 제 부피를 주어진 다음에 온도를 변화시켜서 그 부피를 구하게 하는 문제를 한번 만들어 봐야겠다. . . . (생략) . . . 평가 기준을 보면 평가 기준에 사람의 법칙을 이해해야지 이걸 풀 수 있다. 사람의 법칙 식이 (교과서를 보다가 교사용 지도서를 보며) 아! (교사용 지도서에서 샤를 법칙 공식을 발견하고 문항 해설에 옮겨 적으며) 샤를의 법칙 이제 식을, 이걸 보면 $V=V_0+V_0*t/273$ 의 $1+t/273$ 이거인데.

(예비교사 B의 발성 사고 내용 중에서)

이와는 반대로 교사용 지도서는 교과서보다 상위 개념만이 포함되어 있다고 생각해서 교과서 위주로 참고하여 지필평가 문항을 제작한 경우도 있었다.

면담자: 교과서를 주로 활용하셨어요?

예비교사: 네.

면담자: 지도서는 거의 안 보신 건가요?

예비교사: 지도서는 보긴 했는데 일단 아이들이 이걸 자주 선생님이 보는 거잖아요. 선생님이 보고 이걸로 가르치는 거잖아요. 근데 그냥 그 여기는 이것보다 조금 더 상위 개념이 들어 있을 거라고 생각을 해서 여기에 대한 내용을 잘 안 봤어요. 이걸(교과서를) 더 많이 봤어요.

(예비교사 A와의 사후 면담 내용 중에서)

교사용 지도서에는 교육과정의 위계에 대한 이해 및 해당 개념에 대한 교사의 심층적인 이해를 돕기 위하여 해당 교육과정 수준을 넘어선 이론적 배경 지식과 새로운 과학 동향 등에 대한 자료 등을 포함한 심화·보충 자료가 제시되어 있다(Han & Noh, 2003). 또한 비록 교사용 지도서가 철저한 검토 및 검증 과정을 거쳤다고 하더라도 현실적으로 교과내용지식 및 PCK 구성 요소 측면에서 오류가 있을 수 있음을 감안할 때, 예비교사들이 보여준 교사용 지도서에 대한 무비판적 수용이나 거부하는 예비교사들의 지필평가 문항 제작 능력을 저해하는 요소로 작용할 수 있다. 따라서 교사용 지도서의 특성에 대한 이해 및 평가 문항 제작 시 효과적인 활용 방안에 대하여 지도할 필요가 있다.

6. 검토 관점과 내용의 제한성 및 검토시기에 따른 차이점이 있다.

평가 문항의 타당도와 신뢰도를 확보하기 위해서는 문항 제작 과정에서 문항의 전반적인 사항에 대하여 적절하고 체계적인 검토 과정을 반드시 거쳐야 한다. 본 연구에 참여한 예비교사들의 검토 과정은 검토시기에 따라 크게 세 가지로 구분할 수 있었다. 즉 개별 문항 제작 직후 검토하는 경우, 문항의 정답과 해설 및 채점 기준을 작성하면서 검토하는 경우, 전체 문항 완성 후 종합적으로 문항 전반을 다시 검토하는 경우가 있었으며, 검토시기에 따라 검토 내용은 약간 차이점을 보였다. 먼저 두 명의 예비교사는 개별 문항 제작 직후에 검토 과정을 거쳤는데, 교육과정의 일치 여부 및 발문의 명료성 등을 주로 검토하였으며 다음이 각각의 사례이다.

다음은 일정한 압력에서 온도가 증가하면 기체 부피가 커지는 이유를 찾아봅시다. 다음 칸에 들어갈 말을 순서대로 짝지은 것은? 보기를 챙겨줘서 (교과서와 교사용 지도서의 내용을 보며) 분자 운동이 증가하는 거를 배웠나? 분자 운동 설명, 배우긴 배웠네. 여기 부분이란 여기 온도가 증가하면이란 말을 바꿔서 기체를 가열한다고 해야겠다.

(예비교사 C의 발성 사고 내용 중에서)

예비교사: 크기가 커진다, 작아진다, 아 좀 그러네? 말이 중복돼서 어떡하지? 음.

면담자: 무슨 생각해요?

예비교사: 지금 표현이 겹쳐서 학생들 이 문제를 풀 때 헷갈릴 수 있을 거 같아서 좀 시험 시간에 질문이 나올 것 같아요. 그래서 좀 명확하게 문제를 바꿔야 될 거 같아요.

(예비교사 G의 발성 사고 내용 중에서)

일부 예비교사들만이 개별 문항 제작 직후에 검토 과정을 거친 것과 달리, 문항의 정답과 해설 및 채점 기준을 작성하는 과정에서는

모든 예비교사들이 검토 과정을 거쳤고, 이때 검토 관점은 좀 더 다양하게 나타났다. 그러나 여러 측면을 통합적으로 검토하기보다는 특정 측면에 국한하여 검토하는 경향이 있었다. 예를 들어 예비교사 D는 채점 기준을 작성하면서 문항의 명료성이 부족함을 깨닫고 좀 더 구체적인 발문으로 수정하였으며, 예비교사 H는 채점 기준 및 문항 해설을 작성하다가 본인이 만든 선택지가 샤를 법칙의 예시가 아닌 고체의 열팽창에 대한 예시임을 깨닫고 다른 예시로 수정하였다.

그러면 채점 기준. 첫 번째, 분자 운동을 설명하라고 했으니까, 일단은, 기체 온도가 높아졌다. 그래, 기체 분자의 운동이 활발해졌다는 건 들어 가야 돼. 운동이 활발해 이젠 꼭 들어가야 돼. 아 그럼 이렇게~ (문항을 수정하면서) 기체 분자 운동을 '자세히', '자세히' 설명하여라. 아! 그럼 활발해진다하고 끝날 수도 있으니까 탁구공 안에서 일어난 분자 운동을 설명하, 아! 그럼 원래대로 돌아온 이유를 공의, 탁구공의 부피가 증가한 이유를 ('자세히'를 지우고) '분자 운동으로' 설명하여라.

(예비교사 D의 발성 사고 내용 중에서)

'채점 기준 및 문항 해설은, 1번은 철도에 틈을 조금 두어서 가는 이유는 여름철이, 여름에 온도가 높아. 근데, 이 기체에 관련된 내용인데, 철도는 아닌 것 같은데? 철도의 부피가 증가하는 것도 샤를의 법칙인지 찾아봐야겠다. 아니 샤를의 법칙이 아니지. 기체, 샤를의 법칙은 이상 기체에서 성립하는 법칙이니까. 온도에 따라 부피가 증가한다고 해서 무조건 샤를의 법칙은 아니지. 그러면, 열기구에 대한 내용을 1번에 넣어야겠다.

(예비교사 H의 발성 사고 내용 중에서)

전체 문항 완성 후 문항 전반을 다시 검토하는 경우는 4명의 예비교사에게 나타났는데, 검토 내용과 수준은 개인별로 약간 차이가 있었다. 예를 들어 예비교사 A는 평가 기준, 내용 영역과 난이도, 과학적 오류 여부, 답지 배열의 일관성 등 문항의 체제와 정확성을 총체적으로 검토하였다. 특히 문항의 오류를 발견하고 수정하는 과정에서 답지의 배열순서가 어긋나게 되자, 답지 배열의 일관성을 고려하여 평가 내용을 다시 수정하는 등 정확성과 타당도 및 문항의 체제까지 다양한 요소를 고려하는 꼼꼼함을 보였다.

한 번 검토를 해볼까? 1번. 학생이 평가 기준은 학생이 샤를의 법칙과 온도에 따른 분자의 분자 수, 부피, 압력, 분자 운동. 이 어떻게 변화하는지 알고 있는지를 평가하겠다. 문항 유형, 적용 단계, 난이도는 상이고. 아래 그림은 피스톤의 부피가, 피스톤의 부피가 라는 말이 맞나? '실린더'라고 해야 되나? '실린더의 부피'라고 해야지. 그치? ... (생략) ... 5번 부피는 '가' 보다 '나' 에서. 어, 아이고! 큰일 날 뻔 했다. 바뀌어야지. '나' 보다 '가'에서 이렇게 바뀌어야지. 아, 이러면 '가나', '가나' 하다가 '나가' 하면 이상한데? 음, 통일성이 맞지 않은데? 음, '나'에서 더 크다. 이렇게 적어야 하나?

(예비교사 A의 발성 사고 내용 중에서)

두 명의 예비교사의 경우에는 문항의 오류 여부, 난이도와 행동 영역 판별의 타당도 등과 같은 문항의 타당도와 정확성을 검토하긴 했으나 체계적인 고민이나 고려는 부족한 측면이 있었으며, 다음이 그 사례이다.

틀린 게 있나 한번 보면, 틀린 문항이 없고, 문제는, 음 1번 문항은 없고, 이걸 헛갈릴 수 있으니까, 이걸 난이도 상. 예니까 적용 문제가 맞고.

그냥 샤를의 법칙을 알고 있으면 당연히 알 수 있는 문제니까 난이도는 하. 지식 문제.

(예비교사 F의 발성 사고 내용 중에서)

다른 예비교사들이 거친 검토 과정은 그 내용과 수준이 매우 단순한 편이었다. 예를 들어, 다음 사례에서는 문항 해설의 부족한 부분만을 보충하고 있음을 보여주고 있다.

이번 거는 여기서 끝내고, 전에 거 살펴보면 괜찮은 거 같고. 4번 문제 'ㄱ'의 해설을 써야겠다. 그래프에서 기체의 온도가 증가하면 부피도 증가함을 알 수 있다. 그래서 끝내야지.

(예비교사 C의 발성 사고 내용 중에서)

이와 같이 대부분의 예비교사들이 검토 과정을 거치는 것으로 나타났지만, 검토 시기와 내용 및 수준에서 다소 차이가 있음을 알 수 있었다. 또한 각 문항 제작 직후에 검토 과정을 거치는 경우에는 특정 측면에 한정하여 세밀한 검토 과정이 이루어진 반면, 문항의 정답과 해설 및 채점 기준을 작성할 때 모든 문항을 제작한 후에 문항 전체에 대한 검토 과정을 거치는 경우에는 보다 다양한 측면을 총체적으로 검토하는 경향이 있음을 알 수 있었다. 여러 문항으로 이루어진 평가지를 성공적으로 제작하기 위해서는 각 문항에 대한 검토 과정과 전체 문항에 대한 검토 과정을 분리하고 주요 검토 관점도 달리할 필요가 있다. 즉, 각 문항에 대한 검토 과정에서는 과학적 오류 여부, 교수 전략과의 연계성, 문항의 체제 및 편집 체계, 선택지, 배점 등이 적절한지에 초점을 두는 반면, 전체 문항에 대한 검토 과정에서는 모든 문항들의 교육과정 및 교수·학습 과정과의 부합성 판단, 난이도 및 변별도 판단과 조정, 풀이 소요 시간 판단 및 조정 등에 초점을 두고 검토할 필요가 있다. 그러나 이 연구에서는 이와 관련된 예비교사들의 문항 검토 과정의 취약한 부분들이 나타났으므로, 이를 개선하기 위한 체계적인 교육과 지도가 필요하다.

IV. 결론 및 제언

이 연구에서는 발성 사고법을 이용하여 예비 화학 교사의 지필평가 문항 제작 과정의 특징을 분석하였다. 8명의 예비 화학 교사가 발성 사고법을 통해 개별적으로 지필평가 문항을 제작하는 과정을 분석한 결과, 크게 6가지 특징이 나타났다. 즉, 예비교사들은 지필평가 문항 제작 과정에서 문항 제작 계획 단계의 체계성이 부족하였고, 수업과 평가의 상호관련성에 대한 인식 및 고려가 부족하였다. 그리고 문항 유형 결정 기준이 다양하였으며, 문항의 행동 영역과 난이도 판단을 어려워하는 것으로 나타났다. 교사용 지도서 활용의 제한성과 교사용 지도서에 대한 이해, 검토 관점과 내용의 제한성 및 검토시기에 따른 차이점이라는 특징도 보였다.

이상의 결과는 예비교사들이 지필평가 문항을 제작하는 과정에서 주로 고려하는 측면과 부족한 측면에 대한 정보를 제공하므로, 예비교사의 지필평가 문항 제작 능력을 향상시키는 방안을 모색하는 데 의미 있는 시사점을 제공할 수 있다. 예를 들면 예비교사의 학습 목표에 근거한 출제 계획 단계의 체계성 부족 현상을 개선하기 위해서는, 예비교사에게 출제 계획 단계가 출제 과정 전반과 문항의 질을 확보하는데 중요한 영향을 미침을 사례를 들어 강조할 필요가 있다. 또한

출제 계획 단계를 체계화하고 각 단계에서 해야 할 사항들을 명시화하여 예비교사의 지필평가 문항 제작 과정과 결과물의 질 제고에 도움이 되는 전략을 구안하고 교육할 필요가 있다.

수업과 평가의 상호관련성에 대한 인식과 구현 능력을 향상시키기 위해서는, 예비교사 교육과정에서 수업과 평가의 역동적인 상호관련성을 중시하는 구성주의적 평가 관점에 대하여 이론적 및 실제적 관점에서 보다 강조할 필요가 있다. 또한 수업과 평가의 상호작용 측면에서 다양한 수준의 사례를 제시하여 구성주의적 평가 관점에서 분석해보고, 이를 바탕으로 수업과 평가를 함께 개선하는 방안에 대하여 생각해보고 반성적으로 실천할 수 있는 기회를 제공하는 것도 유용할 수 있을 것이다.

예비교사들이 문항의 난이도, 평가 시행의 효율성과 평가를 통한 학습 효과 및 자신의 평가 전문성 등과 같은 다양한 요소들을 통합적으로 고려하여 문항 유형을 결정할 수 있도록 지원하기 위한 노력도 필요하다. 가령, 예비교사에게 문항 유형 결정에 영향을 미치는 주요 요인들을 구체적인 사례를 들어 소개하고 강조할 뿐만 아니라 반성적인 분석 및 실천 기회를 제공할 필요가 있다. 이를 위해서는 문항 유형 결정에 영향을 미치는 요인과 사례들을 보다 체계적으로 조사하는 연구가 선행되어야 할 것이다. 또한 특정 요인만을 제한적으로 고려하여 문항 유형을 결정하는 사례와 다양한 측면을 통합적으로 고려하여 문항 유형을 결정하는 사례를 조사 및 발굴하고, 각 사례의 긍정적인 측면과 부정적인 측면을 보다 체계적으로 분석하기 위한 연구도 필요하다.

예비교사의 문항의 행동 영역과 난이도 판단 능력을 제고하기 위해서는, 행동 영역 및 난이도에 대한 명확한 이해를 바탕으로 문항의 행동 영역 및 난이도를 실천적으로 판단해볼 수 있는 기회를 가능한 다양하고 반복적으로 제공할 필요가 있다. 특히 문항의 난이도 판단 능력의 경우 학습자 특성에 대한 이해도를 높이는 방법이 효과적일 수 있으므로, 예비교사에게 학생과의 접촉 기회를 보다 풍부하게 제공하기 위한 제도적 장치가 마련되어야 할 것이다. 물론 국내 예비교사 교육과정의 현실상 다양한 제약이 있어 이를 구현하는 데에는 여러 가지 어려움이 있을 수 있다. 따라서 지필평가 문항 제작 과정에서 예비교사에게 평가와 관련된 학생의 특성, 가령 관련 개념에 대한 학생의 이해 수준 및 오개념, 사고력, 유사 문항에 대한 문항 이해 수준 및 응답 특성, 흥미도 등에 대한 정보를 확보하여 체계적으로 제공하고, 이를 바탕으로 난이도를 판단하는 기회를 제공하는 방안이 보다 효율적이고 현실적일 수 있다.

예비교사의 평가 문항 제작 시 교사용 지도서 활용의 제한성을 개선하기 위한 방안으로는, 교사용 지도서에 포함된 PCK 구성 요소별 정보 및 평가 문항 제작 과정에서 이 정보의 효과적인 활용 방안에 대하여 예시하고 강조하는 방안이 유용할 수 있다. 또한 교사용 지도서의 제작 절차와 특성 및 교사용 지도서의 제한점 등에 대해서도 함께 안내한다면 평가 문항 제작 과정에서의 교사용 지도서에 대한 무비판적 수용이나 거부 현상 등을 개선하는 데 도움이 될 수 있을 것으로 기대한다. 이를 위해 예비교사의 지필평가 문항 제작 과정에서 교사용 지도서의 활용 방법 및 제한점을 보다 체계적으로 분석하는 연구가 필요하다.

마지막으로 지필평가 문항의 완성도와 관련된 다양한 측면들을 통합적으로 검토하는 것과 함께 검토시기에 따라 주안점을 달리하여

검토할 필요성 및 실질적인 검토 내용과 방법에 대하여 안내하고 교육함으로써, 예비교사의 지필평가 문항 검토 능력을 제고할 수도 있을 것이다.

이러한 다양한 방안들을 구현하기 위한 구체적인 사례로서 이 연구의 결과를 예비교사교육에 체계적으로 활용한다면 보다 실제적인 교육이 될 것으로 기대한다. 그리고 지필평가 문항 제작 전에 예비교사에게 관련 내용을 단순히 소개하는 수준을 넘어, 실제 지필평가 문항 제작 과정에서 멘토링, 컨설팅, 협의회 등의 전략을 통해 지속적인 피드백을 제공한다면 그 효과는 가중될 것이다. 또한 이러한 방안들을 예비교사 교육과정뿐만 아니라 현직 교사 연수 과정과도 연계하여 실행한다면 예비교사 및 현직교사의 평가 전문성 향상에도 기여할 것으로 기대한다.

국문요약

이 연구에서는 예비 화학 교사가 개별적으로 지필평가 문항을 제작하는 과정의 특징을 분석하였다. 이를 위하여 8명의 예비 화학 교사가 기체 법칙에 대한 6개의 지필평가 문항을 제작하는 발성 사고 과정과 개별 심층 면담 전사본을 분석하였다. 분석 결과, 예비교사들은 지필평가 문항 제작 과정에서 문항 제작 계획 단계의 체계성 부족, 수업과 평가의 상호관련성에 대한 인식 및 고려 부족, 문항 유형 결정 기준의 다양성, 문항의 행동 영역과 난이도 판단 곤란, 교사용 지도서 활용의 제한성과 교사용 지도서에 대한 오해, 검토 관점과 내용의 제한성 및 검토시기에 따른 차이점 등의 6가지 특징을 보였다. 이는 예비교사의 지필평가 문항 제작 능력을 향상시키는 방안을 모색하는 데 의미 있는 시사점을 제공할 수 있다.

주제어 : 지필평가 문항 제작 과정, 발성 사고법, 예비 화학 교사

References

- 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). London, UK: Springer.
- Atjonen, P. (2014). Teachers' views of their assessment practice. *The Curriculum Journal*, 25(2), 238-259.
- Baek, S.-G. (2000). *Principles of performance assessment [수행평가의 원리]*. Seoul: Kyoyookbook Publication Co.
- Bismack, A. S., Arias, A. M., Davis, E. A., & Palincsar, A. S. (2014). Connecting curriculum materials and teachers: Elementary science teachers' enactment of a reform-based curricular unit. *Journal of Science Teacher Education*, 25(4), 489-512.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12.
- Cho, H., & Ko, Y. (2008). Re-conceptualization of secondary science teacher's pedagogical content knowledge (PCK) and its application. *Journal of the Korean Association for Research in Science Education*, 28(6), 618-632.
- Cho, K.-W. (2004). A critical examination of secondary teacher education for secondary school teachers. *Journal of Educational Studies*, 35(1), 1-19.
- Choi, H. S., & Kim, J. B. (2013). A study on performance level of pre-service physics teachers in constructing questions for classroom assessment - Focused on analysis of multiple choice question about physics concept for formative assessment. *Journal of Science Education*,

- 37(3), 458-475.
- Choi, J.-I., & Paik, S.-H. (2016). An analysis of content validity of behavioral domain of descriptive tests and factors that affect content validity: Focus on the fifth and sixth grade science. *Journal of the Korean Association for Science Education*, 36(1), 87-101.
- Choi, J.-Y. (2006). An qualitative case study of pre-service teacher's planning for the practical arts instruction in the elementary school. *The Journal of Elementary Education*, 19(1), 477-499.
- Davis, E. A., Palincsar, A. S., Arias, A. M., Bismack, A. S., Marulis, L., & Iwashyna, S. (2014). Designing educative curriculum materials: A theoretically and empirically driven process. *Harvard Educational Review*, 84(1), 24-52.
- 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.
- Ferguson, L. E., Braten, I., & Stromso, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learning and Instruction*, 22(2), 103-120.
- Grossman, P., & Thompson, C. (2004). Curriculum materials: Scaffolds for new teacher learning? (A Research Report No. R-04-1). Seattle, WA: Center for the Study of Teaching and Policy.
- Han, K.-A., & Noh, S.-G. (2003). An analysis on the utilization of teacher's guides for science in elementary school. *Journal of Korean Elementary Science Education*, 22(1), 51-64.
- Hashweh, M. Z. (2005). Teacher pedagogical constructions: A reconfiguration of pedagogical content knowledge. *Teachers and Teaching*, 11(3), 273-292.
- Jang, S.-M., & Kim, J.-Y. (2002). Analysis on the status of performance assessment in science based on the elementary teachers' concerns. *Journal of Korean Elementary Science Education*, 21(2), 227-240.
- Jeong, E., & Choi, W. (2014). A survey on evaluation in science education at primary and secondary school in Korea. *Journal of Science Education*, 38(1), 168-181.
- Jonassen, D. H. (1991). Evaluating constructivistic learning. *Educational Technology*, 31(9), 28-33.
- Kang, H., & Kang, S. (2015). Analysis of elementary school teachers' self-diagnosis on their competency for assessment in science. *Journal of Korean Elementary Science Education*, 34(2), 153-163.
- Kim, H. J., & Yoo, J. (2012). An analysis on rater error in holistic scoring for performance assessments of middle school students' science investigation activities. *Journal of the Korean Association for Research in Science Education*, 32(1), 160-181.
- Kim, H.-J., Kwack, D.-O., & Sung, M.-W. (2000). An investigation on science teachers' evaluation practices in the secondary schools. *Journal of the Korean Association for Science Education*, 20(1), 101-111.
- Kim, K., Park, E., Song, M., Sang, K., Kim, S., Kim, H., Shin, J., Seo, J., Lee, C., Kim, J., Kim, K. & Choi, S. (2012). A study on management of standards-based assessment in secondary schools [중등학교의 성취평가제 운영 방안 연구]. Seoul: Korea Institute for Curriculum and Evaluation (CRE 2012-8).
- Kim, K.-M. & Kim, S.-W. (2002). A study on the weight of assessment domains in science education focused on the teacher's view points. *Journal of the Korean Association for Science Education*, 22(3), 540-549.
- Kim, S. (2002). A study on the teacher's competence for classroom assessment. *Journal of Educational Evaluation*, 15(1), 51-171.
- Kim, S.-W., & Hyun, M.-S. (2005). The study on the recognition of science teachers about the general matters of performance assessment and the appropriate performance assessment methods in middle school science curriculum. *Journal of Research in Curriculum Instruction*, 9(2),
- Kim, Y., Park, J., Park, J., Lee, H., & Kim, Y. (2010). Science teachers' perceptions and needs for courses in science education subjects for science teacher preparation program in Korea. *Journal of the Korean Association for Research in Science Education*, 30(6), 785-798.
- Ko, M.-S., Kim, E.-A., Heo, J.-M., & Y. I.-H. (2013). Elementary school teachers' beliefs of inquiry and practice of science performance assessment. *Journal of the Korean Society of Earth Science Education*, 6(2), 124-135.
- Kwon, J.-S., Kim, B. K., Choi, B.-S., Kim, H.-N., Paik, S.-H., Yang, I.-H., Kwon, Y.-J., Cha, H., Woo, J.-O., & Jeong, J.-W. (2012). Theories in science education [과학교육론]. Seoul: Kyoyookbook Publication Co.
- Lee, I., Lee, B., Kim, S., Park, J., Jin, J., Kim, O., & Seo, S. (2004). Analyzing realities of teacher's student assessment and educational needs for their professional development [교사의 학생 평가 실태 조사 및 전문성 신장에 대한 요구분석]. Seoul: Korea Institute for Curriculum and Evaluation (RRE 2004-5-1).
- Lee, K. H., & Choi, M. Y. (2013). Critical examination of 2009 revised teacher's guide book by an aspect of pedagogical content knowledge. *Journal of Learner-Centered Curriculum and Instruction*, 13(6), 667-689.
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, source, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome, & N. Lederman (Eds.), *Examining pedagogical content knowledge*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Marzano, R. J. (2006). *Classroom assessment and grading that work*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction* (6th Ed.). Boston, MA: Pearson.
- McTighe, J., & Wiggins, G. (2004). *Understanding by design: Professional development workbook*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Min, H. J. (2012). Development of assessment expertise model through analyzing realities of science teacher's student assessment and teacher training. (Doctoral dissertation). Korea National University of Education, Cheongju.
- Mulholland, J., & Wallace, J. (2005). Growing the tree of teacher knowledge: Ten years of learning to teach elementary science. *Journal of Research in Science Teaching*, 42(7), 767-790.
- Nam, M., Park, S., Song, M., Kim, K., Kim, S., Cho, I., Lim, W., Lee, K., Oh, S., Kang, M. & Kang, J. (2006). A study on improvement of teacher's competence in student assessment(III) [교사의 학생평가 전문성 신장 연구(III)]. Seoul: Korea Institute for Curriculum and Evaluation (RRE 2006-5).
- Noh, T., Lee, J., Kang, S., & Kang, H. (2015). Secondary school science teachers' actual and preferred types of assessment. *Journal of the Korean Association for Science Education*, 35(4), 725-733.
- Noh, T., Yoon, J., & Kang, S. (2009). The investigation of elementary school teachers' perceptions toward constructivist science assessment and their relationship with related variables. *Journal of Korean Elementary Science Education*, 28(3), 352-360.
- Park, C. (2003). The relationship between item types and student's achievement. *Korean Journal of Educational Research*, 41(4), 71-86.
- Park, D. (2012). *Educational evaluation: Understanding and application*. Seoul: Kyoyookkwahaksa.
- Park, H. (2016). A survey on the conditions of middle school science evaluation. *Teacher Education Research*, 55(3), 389-398.
- Park, H. J., Jeong, D. H., & Choi, W. H. (2011). Science teachers' perceptions of science practices. *Journal of the Korean Association for Science Education*, 31(1), 61-77.
- Park, K. (2008). A case study on instructional planning process of teachers. *The Journal of Korean Teacher Education*, 25(3), 379-405.
- Park, S., & Oliver, J. S. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38(3), 261-284.
- Park, S., Kang, C., Kim, K., Lee, G., Lee, M., Shon, M., Lee, H. (2003). An Evaluative Study on the 7th Primary Curriculum(III) [제7차 초·중등학교 교육과정 평가연구(III)]. Seoul: Korea Institute for Curriculum and Evaluation (RRC 2003-2).
- Sato, M., Coffey, J., & Moorthy, S. (2005). Two teachers making assessment for learning their own. *Curriculum Journal*, 16(2), 177-191.
- Schellings, G. L. M., & Broekkamp, H. (2011). Signaling task awareness in think-aloud protocols from students selecting relevant information from text. *Metacognition and Learning*, 6(1), 65-82.
- Shin, J., Ryu, S., & Yang, I. (2016). Teachers' needs and utilization of science performance assessment in the elementary schools. *KNUE Journal of Research in Science Education*, 22(2), 18-29.
- Song, S.-C., Lee, C.-H., & Shim, K.-C. (2014). A study on science teachers' perception about science teacher training curriculum. *Teacher Education Research*, 53(1), 15-27.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design*. (2nd ed.). Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Yang, I.-H., Na, J.-C., Lim, S.-M., Lim, J.-K., & Choi, H.-D. (2008). An analysis of elementary schools' science test items by Klopfer's taxonomy of educational objectives: Focusing on the first term of the 5th grade. *Journal of Korean Elementary Science Education*, 27(3), 221-232.