

# 초등학교 과학과 5, 6학년 서술형 평가문항의 행동영역 내용타당도 및 이에 영향을 미치는 요인 분석

최정인<sup>1</sup>, 백성혜<sup>2\*</sup>

<sup>1</sup>인천가림초등학교, <sup>2</sup>한국교육대학교

## 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

Jung-In Choi<sup>1</sup>, Seoung-Hye Paik<sup>2\*</sup>

<sup>1</sup>Incheon Garim Elementary School, <sup>2</sup>Korea National University of Education

### ARTICLE INFO

#### Article history:

Received 14 December 2015

Received in revised form

22 December 2015

11 January 2016

Accepted 12 January 2016

#### Keywords:

elementary science, descriptive test, content validity, two-dimensional classification of educational objectives, PCK

### ABSTRACT

This study analyzes the content validity of descriptive tests developed for elementary schools, in order to acquire basic data to improve them. Various descriptive tests were collected and tested for differences in proportions between two-dimensional classification of educational objectives and the level of behavioral objectives. Results show that the descriptive tests developed by elementary school teachers mainly focused on “knowledge” and “understanding,” and that content validity for behavioral levels to be low. Nine elementary school teachers were interviewed to understand the result. From the interviews, we found both internal and external factors that cause low content validity. The main internal factors were teachers’ ability to make two-dimensional classification of educational objectives, the teachers’ consideration of students’ level, item level of difficulty, the ease of scoring, and path dependence. The main external factors were curriculum, parents, and administration. Based on the results, we suggested the factors related to elementary school teachers’ PCK of descriptive tests.

## 1. 서론

국가의 경쟁력은 창의적 문제해결력을 지닌 인적 자원에 있다. 이에 따라 세계의 여러 나라들은 창의성의 함양이 미래 사회에서 개인과 국가의 경쟁력과 직결됨을 인식하고, 학생들의 창의성을 신장시키기 위한 교육으로 미래 교육의 방향을 결정하고 있다. 우리나라 역시 창의성 신장을 학교교육의 핵심으로 삼고 있음을 교육과정에서 천명하였다. 2009 개정 교육과정의 총론에서 우리 교육이 추구하는 인간상은 ‘기초 능력의 바탕 위에 새로운 발상과 도전으로 창의성을 발휘하는 사람’임을 규정하고 있으며, 창의적인 인재를 기를 수 있도록 교육과정을 구성한다는 방침을 제시하고 있다(Ministry of Educational and Science Technology, 2009). 이와 더불어 2009 개정 과학과 교육과정에서도 ‘과학적 사고력과 창의적 문제해결력을 길러 일상생활의 문제를 해결할 줄 아는 과학적 소양을 기르는 것’이 과학교육의 목표임을 밝히고 있다(Ministry of Educational and Science Technology, 2012).

교육과정에서 지향하는 바가 창의성, 문제해결력 등의 고등사고능력 신장이라면, 그 성과를 가늠하는 것이 교육 평가이다. Tyler(1949)는 “평가의 과정이란 본질적으로 교육과정 및 수업의 프로그램에 의하여 교육목표가 실제로 어느 정도나 실현되었는지 밝히는 과정이다.”라고 정의했다. 이에 비추어보면 과학적 사고력 및 창의적 문제해결력 신장이라는 과학과의 목표가 어느 정도 달성되었는지 평가를 통해 점검하는 것은 학생 그리고 교육과정 및 교수프로그램의 질적

향상을 위한 중요한 토대가 된다. 교육 평가는 학습자의 행동변화 및 학습과정에 관한 정보를 수집하고 그 결과를 학습의 극대화를 위해 이용할 수 있기 때문이다(Kim & Kang, 2010). 그러나 창의적 문제해결력과 같은 고등사고능력을 평가하기 위해서는 선다형으로 대표되는 선택형 문항 등과 같은 평가도구로는 불충분하다. 선택형 평가 문항들은 기존의 지식을 단순히 기억하고 재생하는 능력을 측정하는데 유용하며, 창의성, 문제해결력 등의 고등사고능력을 평가하는데 한계가 있기 때문이다. 이에 대한 대안으로 제안되고 있는 서술형 평가는 단순한 지식의 암기에서 벗어나, 구체적인 부가 가치 창출능력, 수행 능력, 실천적 지능이 중요해진 현대 사회의 인재 육성을 위해 주목받고 있는 평가 방법이라 할 수 있다(Yang, 2006).

범교과적인 수준 및 과학교과의 수준에서 서술형 평가의 정의를 살펴보면, ‘학생이 이미 알고 있는 지식을 활용하여 주어진 문제 상황을 논리적으로 분석, 설명, 해석하거나 창의적으로 해결하는 능력을 측정하는 평가’(Kim *et al.*, 2010), ‘주어진 질문에 대해 학생들이 높은 반응 자유도를 보이면서 여러 개의 문장으로 응답하는 평가 형태’(Kyungkido Education Office, 2012), ‘지식에 대한 적용 능력, 비판 능력, 감상 능력, 창의력, 종합적 판단능력, 문장 표현능력 등과 같은 고등정신능력을 측정하기 위한 평가’(Hahn *et al.*, 1997), ‘창의성과 같은 고등 사고 기능을 묻거나 문제해결과정, 과학탐구 과정의 이해도를 측정할 수 있도록 구성된 평가’(Kang *et al.*, 2007) 등으로 제시되어 있다.

\* 교신저자 : 백성혜 (shpaik@knue.ac.kr)  
http://dx.doi.org/10.14697/jkase.2016.36.1.0087

서술형 평가는 문제해결력 등의 고등사고능력을 직접적으로 평가함으로써 궁극적으로 창의적인 능력을 신장시킬 수 있으며, 교수학습의 결과뿐만 아니라 그 과정에 대한 자세한 정보의 수집, 학생 개인의 특성과 수준을 지속적이면서도 종합적으로 평가할 수 있다는 특성이 있어 서술형 평가를 통해 학생이 인지적으로 ‘잘 아는 것’뿐만 아니라 ‘아는 것을 실제로 적용할 수 있는지’에 대한 평가가 가능하다 (Paik & Ryu, 2014)는 장점이 있다. 따라서 서술형 평가는 기존의 암기식, 주입식 위주의 학교수업 방법을 개선하는 한편, 다양한 문제 상황에 능동적으로 대응하고 새로운 방식으로 문제를 해결하는 능력을 갖춘 인재육성을 위한 목적에서 도입되어 그 중요성이 점차 커지고 있으며, 2009 개정 교육과정 총론에서도 “교과의 평가는 선택형 평가보다는 서술형이나 논술형 평가 그리고 수행 평가의 비중을 늘려서 교과별 특성에 적합한 평가를 실시하도록 한다.”라고 하여 서술형 평가의 도입을 권장하고 있다 (Ministry of Educational and Science Technology, 2009). 그리고 그에 따라 각 급 학교에서는 선택형 문항이 추가 되는 기존의 평가형식을 지양하고 서술형 평가를 실시하고 있으며 점차 그 비중을 확대해 가고 있는 실정이다.

하지만 서술형 평가는 교사나 학교로부터의 자발적인 평가 개혁이 아닌 상위행정기관으로부터 상의하달(Top-Down) 방식으로 학교 현장에 도입이 되었으며, 교사들에게 익숙한 선다형, 진위형, 배합형 등의 선택형 문항과 단답형, 완성형 등의 서답형 문항들과는 다른 평가문항 제작 기술을 요구하기 때문에 실제로 서술형 평가 도입의 취지에 맞는 평가문항이 개발되었는지 아니면 서술형 평가문항의 형식만을 취하고 있을 뿐인지 면밀한 검토가 필요하다. 올바르게 정확한 평가를 위하여서는 양질의 평가문항을 만드는 것이 우선되기 때문이다. 흔히 양질의 평가문항은 양호도가 높다고 말할 수 있다. 평가도구의 양호도는 타당도, 신뢰도, 객관도, 실용도로 구성되는데 이 중 타당도는 검사도구가 측정하고자 하는 것을 얼마나 충실히 측정하였는가를 의미하는 것으로 검사도구 목적의 적합성에 해당한다 (Seong, 2014). 서술형 평가는 ‘지식’이나 ‘이해’ 등 저차원적인 인지과정을 측정하고 평가하는 것이 아닌 ‘적용’, ‘분석’, ‘종합’, ‘평가’ 등의 고등사고능력을 측정하고 평가하기 위해 도입되었음을 생각하여 볼 때 학교 현장에서 개발된 서술형 평가문항 역시 고등사고능력을 겨냥하고 있어야 할 것이다. 이러한 경우에 우리는 학교의 서술형 평가도구의 타당도가 높다고 말할 수 있다. 그리고 타당도가 높은 평가도구에 의한 교육 평가가 이루어졌을 때 교육 평가의 의미와 가치를 찾을 수 있다 (Kim & Kang, 2010).

서술형 평가를 포함한 평가와 관련된 과학과의 선행연구들은 문항 개발 (Hong & Chung, 2006; Kim & Kim, 2012; Park & Kang, 2012; Song *et al.*, 2004), 분류 틀을 사용한 문항 분석 (Kim & Han, 2010; Lee & Jeong, 2014; Paik *et al.*, 2008; Park, 2012; Yang *et al.*, 2008), 평가에 대한 실태 및 인식 분석 (Choi, 2011; Lee, 2013; Shim, 2008), 평가문항의 타당도 분석 (Paik *et al.*, 2008) 및 신뢰도 분석 (Cho *et al.*, 2013; Kim & Yoo, 2010; Lee & An, 2005) 등의 연구 주제를 보이고 있다.

Lee & Jeong (2014)은 Bloom의 신교육목표분류체계와 국가수준 학업성취도 평가의 과학과 평가 틀에 근거하여 고등학교 생명 과학의 문항을 분류하였으며, 그 결과 서술형을 포함한 서답형 문항의 지식 차원은 사실적 지식과 개념적 지식에 편중되어 있음을 보고하였다.

또한 인지과정 분석결과 ‘기억’과 ‘이해’를 측정하는 평가문항이 전체의 97.5%에 달하였으며, ‘분석’과 ‘평가’는 드러나지 않았음을 결과로 제시하였다. Kim & Han (2010)은 3차원 과학 평가 틀을 활용하여 국가수준 학업성취도 평가문항을 분석하였는데 탐구과정 요소 중 ‘문제 인식’ 및 ‘가설 설정’의 문항수가 적음을 밝히고 있다. 그리고 사회적, 기술·산업적 상황과 같은 다양한 과학상황이 문제 속에 반영되어야 함을 제언하고 있다. Yang *et al.* (2008)은 교사가 제작한 초등학교 5학년 과학과 지필평가문항을 Klopfer의 교육목표분류체계로 분석한 결과 ‘지식’과 ‘이해’ 항목의 비율이 너무 높게 나타났으며, 과학적 탐구과정은 상대적으로 낮은 비율임을 보고하고 있다. Park (2012)은 중학교 1학년 과학과 총괄 지필평가문항을 분석한 결과 행동영역의 경우 ‘지식’, ‘이해’ 수준에 편중되어 있으며, ‘적용’, ‘분석’, ‘종합’, ‘평가’를 측정하는 문항은 거의 출제가 되지 않고 있다고 말하였다.

한편, Lee (2013)는 고등학교 화학과 서술형 평가에 대한 현황 분석 연구를 통해 교사들은 학생의 고등정신능력과 논리적 표현력의 측정 및 향상을 위해 서술형 평가가 필요하다고 인식하고 있음을 보고하였다. Shim (2008)은 중학교 과학과 서술형 평가에 대한 인식 및 실태 조사를 통해 중학교 과학 교사들은 학생들의 사고력을 길러주기 위해 서술형 평가가 필요하다고 생각하며, 학생들은 공부를 더 확실하고, 열심히 그리고 깊게 하게 되기 때문에 서술형 평가가 필요하다고 응답하였음을 보고하였다. 그리고 서술형 평가문항 출제 단계에서 단순 지식보다는 고등정신능력을 측정할 수 있는 문항을 제작하려는 노력이 필요함을 제언하고 있다. Choi (2011)는 중학교 과학과 서술형 평가에 대한 문항 특성 및 교사와 학생의 인식을 조사하였다. 그녀는 서술형 평가문항이 채점이 용이한 내용을 중심으로 출제되는 경향이 있음을 밝혔고, 서술형 평가문항의 수행 수준은 단순정신능력에 초점을 둔 문항이 다수 출제되어 학생들의 고등정신능력과 창의적 문제해결력을 측정하고자 하는 평가의 목적에 어긋남을 지적하였다. 또한 중학교 과학교사들은 서술형 평가문항의 제작의 전반적인 과정에서 평가 전문성이 높지 않음을 확인하였다.

Paik *et al.* (2008)은 고등학교 화학 I 과정 ‘물’ 단원의 학업성취도 평가문항의 내용타당도를 검증하였다. 연구결과 고등학교에서 실시하는 학업성취도 평가에서 내용영역의 내용타당도는 높으나 행동영역의 내용타당도는 낮은 것으로 나타났다. 이러한 연구 결과를 통해 행동영역 중 분석력, 종합력, 평가력 등 고등정신능력을 기르는 평가문항의 개발에 노력을 기울여야 함을 제언하고 있다.

이상의 선행연구들은 과학교육에서 양질의 평가가 정착되고 시행되기 위한 중요한 시사점들을 제공하여 주고 있으나 초등학교 과학과 서술형 평가를 대상으로 한 내용타당도 분석 연구는 아직까지 없다. 초등학교 교육현장에서 서술형 평가가 확대되어 실시되고 있는 현실에서 학생들의 고등사고능력을 측정 및 평가할 수 있는 문항이 출제되고 있는지 내용타당도 분석을 통해 확인하는 것은 매우 중요한 일이라 할 수 있다. 만약 학생의 ‘지식’이나 ‘이해’를 측정 및 평가하는 문항이 주로 출제되고 있다면 이것은 서술형 평가문항이 갖는 장점을 살리지 못하는 것이며, 서술형 평가문항의 도입 취지와도 맞지 않는 것이다. 게다가 ‘지식’이나 ‘이해’ 등의 인지과정은 서술형이 아닌 선택형이나 서술형을 제외한 서답형 문항으로도 효율적으로 측정 및 평가가 가능하다. 따라서 본 연구는 초등학교 교사들이 제작한 서술형 평가문항들이 서술형 평가 도입의 취지에 맞게 고등사고능력 평가

를 위한 도구로서의 기능을 충분히 하고 있는지 내용타당도 분석을 통해 알아보는 데 목적이 있다. 또한 서술형 평가 문항 제작 및 실행 경험에 대한 초등 교사들과의 면담을 통해 서술형 평가문항의 내용타당도 분석 결과를 다각도로 조명함으로써 교사의 평가전문성 신장 및 평가도구의 개선을 위한 기초 자료를 제시하고자 한다.

## II. 연구방법

### 1. 연구 대상

이 연구는 A광역시교육청에서 운영하는 교수학습 지원센터 누리집에 탑재된 2014학년도 초등학교 5, 6학년 과학과 총괄평가 문항 중 서술형 평가문항을 내용타당도 분석의 대상으로 선정하였다. 연구의 목적에 비추어 Bloom의 교육목표분류학을 이원분류표에 적용하지 않는 학교의 평가문항은 분석대상에서 제외하였다. 이원분류표란 문항을 제작하기 전에 문항개발자가 참고하는 표로써, 이를 통해 어떤 내용을 어떤 정신능력수준까지 측정할 것인가를 결정한다(Seong, 2014). 따라서 문항출제자가 문항을 개발할 때 ‘분석’, ‘종합’, ‘평가’와 같은 고등사고능력을 얼마나 반영하고자 의도하였는지 그리고 제작된 평가문항이 실제로 그러한 고등사고능력을 얼마나 반영하고 있는지를 비교분석하기 위해선 이원분류표가 Bloom의 교육목표분류학을 기초로 작성되어야 하므로 이원분류표의 행동영역을 ‘지식-이해-적용’으로 약식화한 경우나 ‘과학지식-탐구능력-과학적 태도’로 작성한 경우 그리고 교육과정문서의 성취기준만을 평가목표로 제시하고 행동영역을 별도로 기재하지 않는 경우 등은 분석대상에서 제외하였다. 그 결과 A광역시 소재의 5학년 과학과 평가문항은 26개교의 277 문항, 6학년 과학과 평가문항은 25개교의 226문항이 표집 되었다. 5학년 2학기 1차 평가, 5학년 2학기 2차 평가, 6학년 2학기 2차 평가의 경우 학급별 평가가 시행된 학교가 포함되어 학교 수에 비하여 분석된 서술형 평가문항의 수가 많다. 시험 시기 및 학교별 문항 수는 table 1과 같다.

선행연구(Kim & Kang, 2010; Paik *et al.*, 2008; Song, 2003)에서는 지도서에 제시된 수업목표와 평가도구의 평가목표를 내용영역과 행동영역으로 나누어 내용타당도 분석을 하였다. 그러나 서술형 평가문항은 선택형 문항에 비해 출제되는 문항의 수가 상대적으로 적기 때

문에 1회의 평가 속에 많은 수업목표를 반영하기 어렵다. 따라서 서술형 평가문항의 경우 평가문항의 내용영역이 수업목표를 얼마나 잘 반영하고 있는지를 통해 타당도를 확인하는 것은 연구의 목적에 비추어 의미 있는 근거를 제공하지 못하므로 인지과정에 해당하는 행동영역만을 내용타당도의 분석 대상으로 삼았다.

Table 1. Numbers of items for content validity analysis

Year-Grade-Semester-Type	Number of analyzed items
2014-5-1-midterm	72 items of 15 school
2014-5-1-finals	100 items of 19 school
2014-5-2-midterm	53 items of 6 school
2014-5-2-finals	52 items of 5 school
subtotal	277 items of 26 school
2014-6-1-midterm	72 items of 16 school
2014-6-1-finals	97 items of 21 school
2014-6-2-midterm	12 items of 3 school
2014-6-2-finals	45 items of 3 school
subtotal	226 items of 25 school
total	503 items of 33 school

서술형 평가문항의 내용타당도 분석과 함께 분석 결과에 대한 심층적인 이해를 도모하기 위하여 현직 초등교사 9명을 대상으로 반구조화된 면담을 진행하였다. 면담의 대상자들은 Table 2와 같다. 경력 5년 이하의 교사가 3명, 5년 초과~10년 이하 경력의 교사가 3명, 10년 초과~15년 이하 경력의 교사가 3명이었으며, 남자 교사 3명과 여자 교사 6명이었다. 또한 학사 학위자가 5명, 석사 학위자가 2명, 박사 학위자가 2명이었으며, 대도시 학교에 근무하는 교사가 8명, 지방 학교에 근무하는 교사가 1명이었다. 이들은 모두 과학과 서술형 평가문항 출제 경험이 있으며, 교육과정 운영 및 교수학습 지도, 교육 평가에 대한 관심이 높은 교사들로 면담에 적극적인 참여 의사를 밝혔다.

### 2. 연구 방법 및 절차

이 연구에서 서술형 평가문항의 내용타당도 분석절차는 Figure 1과 같다.

Table 2. The background informations of participants

Participants	Gender	Career of Education	Major Course	Highest academic degree	Region
C	Male	2years and 2months	Elementary education	Bachelor	A metropolitan city
J	Female	7years and 6months	Elementary education	Bachelor	A metropolitan city
K1	Male	14years	Educational Administration	Doctor	A metropolitan city
K2	Female	10years and 7months	Elementary Education	Bachelor	A metropolitan city
K3	Female	9years and 5months	Elementary Science Education	Master	B metropolitan city
K4	Female	12years	Language Education	Master	C provincial city
N	Female	9years	Elementary Science Education	Doctor	D metropolitan city
P1	Female	5years	Elementary Education	Bachelor	A metropolitan city
P2	Male	3years and 7months	Elementary Education	Bachelor	A metropolitan city

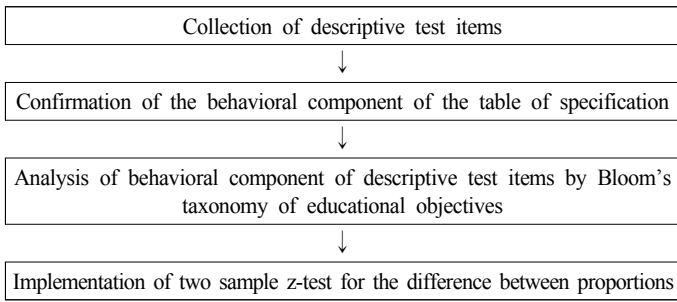


Figure 1. Procedure of content validity analysis of items

평가문항의 평가목표는 내용과 행동으로 이루어져 있다. 평가할 내용이 정해지더라도 그에 대한 행동준거를 어떻게 설정하느냐에 따라 문항이 요구하는 인지적 수준이 달라질 수 있다(Kwon *et al.*, 2012). 교사가 작성한 이원분류표의 행동영역은 제작한 평가문항이 측정하고자 하는 학생들의 인지적 수준을 의미하기 때문에 학업성취도 검사의 내용타당도 검증은 문항들이 검사 제작 전 작성한 이원분류표에 의하여 제작되었는지를 확인하는 방법을 사용할 수 있다(Kim & Kim, 2011; Seong, 2014). 이 연구에서는 행동영역의 내용타당도 분석을 위한 준거로 Bloom의 교육목표분류학을 사용하였다. Bloom의 교육목표분류학은 지식의 구분, 유목의 구인타당도, 유목의 모호성, 위계의 비 타당성, 일차원적 단일성 등의 문제가 있음을 지적받고 있으나(Kang *et al.*, 2005), 그것이 가지는 권위 및 범용성으로 인해 여전히 다수의 학교에서 이원분류표의 준거로써 사용이 되고 있다.

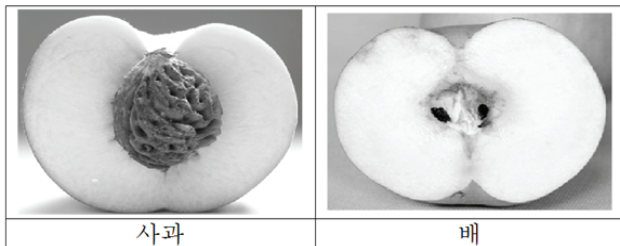
Bloom(1956)은 인지적 행동특성을 서술하기 위하여 인지적 영역을 체계화한 교육목표분류학을 제시하였다. ‘지식’, ‘이해’, ‘적용’, ‘분석’, ‘종합’, ‘평가’ 중 ‘평가’가 가장 복합적인 지적능력이라 규정하였으며, 단순정신능력에서 고등정신능력으로 인지과정을 위계화하였다. ‘지식’은 이미 배운 내용, 즉 사실, 개념, 원리, 방법, 유형, 구조, 이론 등에 대해 기억(회상 및 재인)하는 능력을 말하고, ‘이해’는 이미 배운 내용의 의미를 파악하는 능력을 뜻하며 단순히 자료를 기억하는 수준을 넘어 자료가 다소 치환되어도 의미를 파악하고 해석하고 추론하는 능력을 말한다. ‘적용’은 이미 배운 내용, 즉 개념, 규칙, 원리, 이론, 기술, 방법 등을 구체적인 또는 새로운 장면에서 활용하는 지적 기능을 의미하며, ‘분석’은 조직, 구조 및 구성 요소의 상호 관계를 이해하기 위하여 주어진 자료의 구성 및 내용을 분석하는 능력을 뜻한다. ‘종합’은 비교적 새롭고 독창적인 형태, 원리, 관계, 구조 등을

만들어 내기 위하여 주어진 자료의 내용 및 요소를 정리하고 조작하는 능력이며, ‘평가’는 아이디어, 작품, 해결책, 방법, 자료 등의 가치를 내적·외적 준거로 판단하는 능력을 말한다(Kim & Kim, 2011). Bloom의 교육목표분류학을 사용하여 문항이 요구하는 인지적 수준을 결정하는 과정의 예시는 다음과 같다.

Figure 2는 G초등학교 5학년 1학기 2차 평가에 수록된 서술형 평가 문항으로 이원분류표에서 ‘적용’의 인지과정을 측정하고자 목표하였다. ‘적용’은 이미 배운 내용, 즉 개념, 규칙, 원리, 이론, 기술, 방법 등을 구체적인 또는 새로운 장면에서 활용하는 능력을 말한다. 그러나 위 문항 속의 열매의 종류 및 열매의 기능에 관한 내용은 교과서에 수록된 것들로 이 문제를 풀기 위해선 학생들은 이미 배운 내용을 기억해내는 능력이 요구된다. 따라서 이 문항은 ‘지식’을 측정하고 평가하는 문항이 된다. ‘적용’을 평가하기 위해서는 주어지는 문항 사태가 새롭고 독창성이 있어야 하는데 학습사태에서 이미 접해 보았던 문제 사태로는 지식이나 이해 이상의 능력을 평가하기 힘들다(Kim & Kim, 2011). Figure 3은 S초등학교 6학년 1학기 2차 평가에 수록된 서술형 평가문항으로 이원분류표에서는 ‘종합’의 인지과정을 측정하는 것으로 표기가 되어 있었다. ‘종합’이라는 인지과정은 비교적 새롭고 독창적인 형태, 원리, 관계, 구조 등을 만들어 내기 위하여 주어진 자료의 내용 및 요소를 정리하고 조작하는 능력을 의미한다. 그러나 이 문항을 풀기 위해서 학생에게는 그저 산, 염기와 반응하는 지식의 색변화에 대한 ‘지식’만이 요구될 뿐이다. 이와 같이 Bloom의 교육목표분류학을 사용하여 서술형 평가문항의 행동영역의 인지적 수준을 결정하는 작업에 과학교육전문가 1인, 초등과학교육 박사과정 1인, 초등과학교육 석사학위 소지자로 서술형 평가도구의 제작 경험이 있는 초등학교교사 2인이 평가문항분류자로 참여하였다. 먼저 연구자의 분류 결과의 신뢰도를 확보하기 위하여 검사-재검사를 일정기간 간격으로 3회 반복하였으며, .90 이상의 일치도를 얻어내었다. 분류자간 일치도를 검증하기 위해 연구자, 분류자들이 각각 분류한 결과를 명목 척도로 처리하고 신뢰도를 분석하였으며, 결과는 Cronbach  $\alpha$  값이 0.825로 분류자간 신뢰도가 확보되었다. 분류 결과가 상이한 문항에 대해서는 상호 조율을 통해서 재분류를 한 후 행동영역의 인지적 수준을 확정지었다.

서술형 평가문항의 행동영역의 인지적 수준을 결정 한 후 내용타당도를 판단하기 위하여 교사가 작성한 이원분류표의 행동영역별 행동소와 연구자가 분류한 문항의 행동영역별 행동소의 빈도수와 백분율

7. 아래 사진은 열매의 모습입니다. 식물이 열매를 만들면 어떤 이로운 점이 있는지 쓰시오. (4점)



( )

Figure 2. Example of descriptive items for fifth grade

22. 산과 염기가 무엇인지 친구 상우에게 알려주려고 합니다. 아래의 <보기>의 단어를 모두 사용하여 산과 염기에 대해 설명하십시오. (8점)

<보기>  
 붉은색 리트머스 종이, 푸른색 리트머스 종이  
 페놀프탈레인 용액

Figure 3. Example of descriptive items for sixth grade

을 각각 구한 후 비율차를 검정하였다. 내용타당도 분석을 위한 비율 차 검정은 다음과 같은 방법으로 계산한다(McAfee, 2011; Paik *et al.*, 2008).

$$t(Z) = Dp / \sigma Dp$$

$t(Z)$  : 비율차 지수,  $Dp$ : 비율차,  $\sigma Dp$ : 비율의 표준오차

$$Dp = p_1 - p_2$$

$$\sigma Dp = \sqrt{p(1-p)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

$$p = \frac{n_1 p_1 + n_2 p_2}{N_1 + N_2}$$

$N_1$  = 이원목적분류표상의 행동영역의 전체 행동소 수

$N_2$  = 서술형 평가문항이 측정하는 행동영역의 전체 행동소 수

$n_1$  = 이원분류표 상의 행동영역별 하위 행동소 수

$n_2$  = 서술형 평가문항이 측정하는 행동영역별 하위 행동소 수

$p_1$  = 이원분류표 상의 행동영역별 하위 행동소 비율

$p_2$  = 서술형 평가문항이 측정하는 행동영역별 하위 행동소 비율

통계처리 결과 비율차의 유의도가  $p > .05$ 이면 이원분류표 상의 인지수준과 문항의 인지수준이 동질적이며 내용타당도가 높은 것으로 해석하였고,  $p < .05$ 이면 이원분류표의 인지수준과 평가문항의 인지수준 사이에 유의미한 차이가 있는 것이므로 내용타당도가 낮은 것으로 해석하였다.

그리고 내용타당도 분석 결과에 대한 심층적인 이해를 위하여 9인의 교사를 대상으로 반구조화된 면담을 진행하였다. 면담에서는 초등 교사들이 과학과 서술형 평가에 대해 어떠한 개념을 보유하고 있으며, 어떠한 방법과 절차로 서술형 평가문항을 출제하는지 그리고 서술형 평가를 시행하는데 어려움을 주는 요소는 무엇인지를 중심질문으로 물었다(Table. 3). 면담은 자유롭고 허용적인 분위기에서 시행되었으며, 3인의 교사는 시간과 거리 등의 제약으로 전화 면담으로 대체하였다.

Table 3. Question contents of semi-structured interview

Q 1	What does 'descriptive test' mean in elementary school science?
Q 2	What methods and procedures do you have to create descriptive test items?
Q 3	What difficulties do you have to create and conduct the descriptive test?

Table 4. The frequency analysis of behavioral component(cognitive levels) of descriptive test items in 5<sup>th</sup> grade

	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	Total
Table of specification	21 (7.58%)	51 (18.41%)	83 (29.96%)	28 (10.11%)	79 (28.52%)	15 (5.42%)	277 (100%)
Descriptive test items	116 (41.88%)	85 (30.69%)	29 (10.47%)	24 (8.66%)	23 (8.30%)	0 (0.00%)	277 (100%)

Table 5. The frequency analysis of behavioral component(cognitive levels) of descriptive test items in 6<sup>th</sup> grade

	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	Total
Table of specification	20 (8.85%)	41 (18.14%)	55 (24.34%)	28 (12.39%)	80 (35.40%)	2 (0.88%)	226 (100%)
Descriptive test items	78 (34.51%)	76 (33.63%)	24 (10.62%)	43 (19.03%)	5 (2.21%)	0 (0.00%)	226 (100%)

이 연구는 A광역시 소재의 초등학교에서 출제된 서술형 평가문항 중 Bloom의 교육목표분류학을 이원분류표에 사용한 경우로 연구대상을 한정하였으므로 내용타당도 분석의 결과를 일반화하는데 유의하여야 한다. 또한 의미 있는 면담 결과를 도출하기 위하여 평소 교육과정 운영이나 교수학습 지도 및 평가에 관심이 큰 교사들 중 면담에 적극적인 참여 의사를 밝힌 교사들을 대상으로 면담을 진행하였으므로 그 결과를 전체 교사로 확대하여 해석하는데 유의하여야 한다.

#### IV. 연구 결과

##### 1. 행동영역의 내용타당도 분석

이원분류표에서 교사가 서술형 평가문항을 통해 측정하고자 하는 행동영역의 행동소를 확인하고, 이에 대응하는 서술형 평가문항을 Bloom의 교육목표 분류학에 기초하여 '지식', '이해', '적용', '분석', '종합', '평가'의 6개 하위 행동소로 분류하였다. 5학년 분석 결과와 6학년 분석 결과를 각각 Table 4, Table 5에 제시하였다.

5학년 서술형 평가의 이원분류표 상의 행동영역은 '적용', '종합', '이해', '분석', '지식', '평가'의 순으로 비중을 차지하였으며, '적용' 이상의 행동소들은 전체 문항의 74.01%를 점유하였다. 그리고 6학년 서술형 평가의 이원분류표 상의 행동영역은 '종합', '적용', '이해', '분석', '지식', '평가'의 순으로 비중을 가졌으며, '적용' 이상의 행동소들은 전체 문항의 73.01%를 차지하였다. '적용', '분석', '종합', '평가' 등은 고등정신능력, 창의력, 문제해결력 등으로 지식을 획득한 후 새로운 문제 상황에 직면해서 그것을 응용하거나, 분석하거나, 종합하거나, 평가한다는 것이 공통점이라 할 수 있으며, 이러한 수준은 학습자 개인이 획득한 지식과 이해를 독립적으로 사용할 수 있는 수준을 일컫는다(Kim & Kim, 2011). 학교현장에서의 교수학습의 궁극적인 목표도 이와 같은 학생 개별적 능력의 개발에 있으며, 이러한 인지 과정을 측정하고 평가하기 위하여 서술형 평가가 도입된 만큼 이원분류표의 행동소 분류 결과는 현장의 교사들이 서술형 평가의 특성을 잘 이해하고 있는 것으로 보여 진다.

그러나 출제된 서술형 평가문항을 Bloom의 교육목표분류학에 의거하여 분류한 결과는 이원분류표의 분류 결과와 크게 달랐다. 5학년의 경우, 서술형 문항이 측정하게 되는 학생의 인지적 과정은 '지식', '이해', '적용', '분석', '종합'의 순으로 나타났으며 '평가'를 측정하는

문항은 발견되지 않았다. 교사들은 이원분류표 상에는 ‘적용(29.96%)’ 및 ‘종합(28.52%)’을 측정하는 서술형 평가문항을 큰 비중으로 계획하였으나 실제 출제된 평가문항들은 ‘지식(41.88%)’을 묻는 문항이 가장 많았으며, ‘지식’과 ‘이해’를 측정하는 문항이 전체 문항의 72.57%에 달하였다. 6학년 서술형 평가문항의 경우, 문항이 측정하게 되는 학생의 인지적 과정은 ‘지식’, ‘이해’, ‘분석’, ‘적용’, ‘종합’의 순으로 나타났으며, 역시 ‘평가’를 측정하는 문항은 발견되지 않았다. 교사들은 이원분류표 상에서 ‘종합(35.40%)’과 ‘적용(24.34%)’을 측정하는 서술형 평가문항을 비중 있게 계획하였으나 출제된 서술형 평가문항들은 ‘지식’과 ‘이해’를 묻는 문항이 전체 문항의 68.14%를 차지하였다. 이원분류표의 계획과는 상반되게 출제된 서술형 평가문항들이 학습자의 ‘지식’과 ‘이해’를 주로 묻는 것은 교사들이 서술형 평가문항이 갖는 기능을 잘 알고 있음에도 불구하고, 교사들의 서술형 평가문항 제작 현실은 여전히 선택형 및 서답형 평가 문항의 프레임 벗어나지 못하고 있음을 보여준다.

교사가 서술형 평가문항을 개발하면서 측정하고자 의도하였던 Bloom의 행동영역이 실제 평가문항에 얼마나 잘 반영되었는지 즉, 이원분류표의 평가목표와 평가문항의 행동영역 사이에 유의미한 차이가 있는지 알아봄으로써 내용타당도를 확인하기 위하여 비율차 검정을 실시하였다. 비율차 검정의 결과 유의확률 .05의 수준에서 5학년 서술형 평가문항은 ‘분석’을 제외한 모든 행동소에서 유의미한 차이를 보였다(Table 6). 그리고 ‘지식’, ‘적용’, ‘종합’은 유의미한 차이가 ‘이해’와 ‘평가’에 비하여 크게 나타났다. 초등학교 5학년 서술형 평가문항의 비율차 검정의 결과는 Bloom의 행동영역의 6개 하위 행동소 중 5개의 행동소에서 유의미한 차이를 보이고 있어 내용타당도가 낮았다.

Table 6. The test results for difference in proportions between objects classification and behavioral domain of items in 5<sup>th</sup> grade

Cognitive levels	$P_1$	$P_2$	Dp	$\sigma Dp$	t(z)	P-value
Knowledge	0.0758	0.4188	-0.3430	0.0423	-8.0986	0.0001*
Comprehension	0.1841	0.3069	-0.1227	0.0423	-2.9001	0.0037*
Application	0.2996	0.1047	0.1949	0.0412	4.7369	0.0001*
Analysis	0.1011	0.0866	0.0144	0.0322	0.4478	0.6543
Synthesis	0.2852	0.0830	0.2022	0.0403	5.0144	0.0001*
Evaluation	0.0542	0.0000	0.0542	0.0186	2.9148	0.0036*

p\* < .05

6학년 서술형 평가문항의 경우, ‘분석’과 ‘평가’를 제외한 나머지 행동소에서 유의미한 차이를 보였다(Table 7). 그리고 ‘지식’, ‘이해’, ‘종합’의 유의미한 차이는 ‘적용’의 경우보다 큰 것으로 드러났다. 초등학교 6학년 서술형 평가문항의 비율차 검정의 결과는 Bloom의 행동영역의 6개 하위 행동소 중 4개의 행동소에서 유의미한 차이를 보이고 있으므로 내용타당도가 낮다고 볼 수 있다.

이상의 내용타당도 분석의 결과는 A광역시 소재의 총 33개교에서 시행된 2014학년도 1, 2학기 서술형 평가문항들이 이원분류표에서 규정한 행동영역을 잘 대표하지 못하여 본래 측정하고자 하였던 학생의 인지과정을 충실히 측정하지 못하는 근본적인 문제가 있음을 알려

준다. 교사들이 ‘적용’, ‘종합’ 등의 고등사고기능을 측정하고 평가하기 위한 계획으로 이원분류표를 작성하였으나 서술형 평가문항은 ‘지식’이나 ‘이해’를 주로 측정하고 있기 때문이다. 과학성취도 검사가 단순히 선언적 지식이나 사실을 상기하고 기술하게 하는 등 낮은 수준의 인지적 능력을 요구하는 평가현실에 관한 문제제기는 이전에도 있어 왔다. (Gullickson, 1984; Lee & Jeong, 2014; National Research Council, 2006; Paik *et al.*, 2008; Stiggins & Bridgeford, 1985; Yang *et al.*, 2008). 그리고 이를 해소하기 위해 서술형 평가가 도입된 지금의 평가 현실에서도 이러한 문제제기는 여전히 유효한 것으로 보인다.

Table 7. The test results for difference in proportions between objects classification and behavioral domain of items in 6<sup>th</sup> grade

Cognitive levels	$P_1$	$P_2$	Dp	$\sigma Dp$	t(z)	P-value
Knowledge	0.0885	0.3451	-0.2566	0.0461	-5.5725	0.0001*
Comprehension	0.1814	0.3363	-0.1549	0.0470	-3.2964	0.0010*
Application	0.2434	0.1062	0.1372	0.0439	3.1229	0.0018*
Analysis	0.1239	0.1903	-0.0664	0.0426	-1.5571	0.1194
Synthesis	0.3540	0.0221	0.3319	0.0447	7.4183	0.0001*
Evaluation	0.0088	0.0000	0.0088	0.0084	1.0476	0.2948

p\* < .05

## 2. 행동영역의 내용타당도에 영향을 미친 요인 분석

‘초등학교 과학과 서술형 평가란 무엇이라고 생각하는가?’라는 질문에 대해 교사들은 다음과 같이 응답하였다.

C교사: 서술형 평가는 과학적 개념을 오개념 없이 설명할 수 있는지를 평가하고, 학생들이 자신이 생각한 것을 논리적인 이유나 증거를 들어 답을 설명하는 것이라고 생각해.

J교사: 우리 주변의 자연현상, 과학적 사실, 이론에 대해 얼마나 이해하고 있는가를 주어진 문제를 통해 자신이 갖고 있는 과학과 지식을 연결 지어 풀어내는 그런 논리력을 평가하는 것이 아닐까요?

K3교사: 학습내용을 평가하는데 그 형식이 서술형이요. 기존 시험이 걸릴거시 평가니까. 실제로 알고 있는가를 풀어내는 논리력 이런 걸 서술형으로 평가할 수 있으니까요.

P1교사: 과학문제를 해결하는 과정을 평가하는 것이 과학과 서술형 평가라고 봐요.

K1교사: 단답형 지식을 측정하는 것이 아니라 과정을 평가하여 학습자의 종합적 사고력, 문제해결력, 판단력을 측정가능한 평가라고 생각해.

K4교사: 과학과의 단순한 개념을 암기하거나 단편적 지식들을 확인하는 게 아니라, 그 지식을 탐구하는 과정을 확인할 수 있는 평가방법이라고 생각해. 그러니까 애들이 서술형 평가 문제를 해결하면서 학습한 내용에 대한 총체적인 지식을 활용하도록 하게 하는 거지.

N교사: 탐구능력을 측정할 수 있는 평가 방식이라 생각해. 분산된 지식 위주의 평가보다는 종합적인 사고력을 평가하기 위해 도입된 거잖아. 더욱이 과학글쓰기가 중요해지는 요즘 서술형 평가 도입은 필요하다고 봐.

P2교사: 서술형 평가라는 건 제가 생각하기에 무엇을 아는가가 아니라 어떻게

아는가를 측정할 수 있는 문제라고 생각해요.

K2교사: 평가를 받는 사람이 정해진 답을 고르는 것이 아니라 자신이 가지고 있는 지식을 바탕으로 답하도록 해서 평가 대상자의 이해도를 점검하는 거?

초등학교 교사들은 과학과 서술형 평가가 논리력, 문제 해결력, 탐구 과정, 지식의 이해 방식 등을 평가한다고 인식하고 있었다. 이러한 인식은 서술형 평가가 피험자의 조직력, 분석력, 비판력, 종합력, 문제해결능력을 함양시킬 수 있다는(Seong, 2014) 보편적인 서술형 평가에 대한 시각과 맥을 같이 하는 것으로 볼 수 있다.

그러나 이러한 인식과는 상반되게 A광역시 소재 초등학교의 과학과 서술형 평가문항의 행동영역의 내용타당도가 낮은 것에 대한 근본적인 원인을 살펴보기 위하여 면담의 내용을 전사한 후 행동영역 내용타당도 분석 결과에 비추어 시사점을 제공하는 의미 있는 구성요소들을 도출하였다(Figure. 4). 실선은 요소간의 위계를 나타내며, 점선은 동위요소 간의 관계를 나타낸다. 초등학교 과학과 서술형 평가 내용타당도에 영향을 끼치는 것으로 판단되는 요소들은 크게 교사내적요인과 교사외적요인의 범주로 나누어 볼 수 있다.

1) 교사내적요인

교사내적요인에 포함되는 요소들은 평가문항의 내용타당도와 직접적 관련을 보인다. 그 중 이원분류표의 작성순서 문제는 과학과 서술형 평가문항의 내용타당도가 낮은 이유를 설명해준다. 서술형 평가문항 개발 절차와 관련하여 면담에 참여한 9명의 초등 교사 모두 문항을 개발한 후 이원분류표를 작성한다고 응답하였다. 또한 이원분류표를 평가 문항 개발 전에 먼저 작성하는 것이 일반적 절차임을 아는 교사는 1명뿐이었다.

J 교사: 저는 문항부터 내요. 우선 단위별로 문항비율을 정하고, 내용을 선정한 다음 문제를 내요. 문제를 보고 그걸 저 분류기준에 입각해서 어떤 수준인지를 결정하는 거죠. (중략) 그래요? 그런가? 그랬던 것 같기도 하고. 그냥 이 방식이 편한 것 같아요. 익숙하고 효율적이고요.

다른 교사들도 J교사와 유사하게 문항의 초안을 먼저 작성하고 이후 이원목적분류표를 작성한다고 말하였다. 이렇게 문항의 초안을 먼저 작성하는 이유는 대부분의 교사가 문제를 직접 개발하기보다 기존에 개발된 평가문항들을 활용하는 경우가 많기 때문이다. 그러나 문

항을 출제하고 난 후 이원분류표를 작성하게 되면 특정 인지과정을 측정하는 평가문항으로 편중되기 쉽다. 특히 서술형 문항의 경우 의도적으로 고등정신능력을 측정하도록 제작(Kim et al., 2010)되어야 하는데, 문항부터 개발하게 된다면 어떤 고등정신능력을 측정할 지에 대한 검토의 기회를 가지지 못할 수 있다. 이원분류표는 문항이 어떤 내용을 측정하는가를 보여주는 내용소와 그 내용을 어느 인지수준에서 측정하는가를 보여주는 행동소로 나누어 제시함으로써 검사 개발을 위한 전체적 틀을 제공하는데 목적이 있기 때문이다(Kim et al., 2010). 이 점에 대해 연구자가 지적하였을 때 면담 교사들은 비록 이원분류표를 미리 작성하지는 않지만, 서술형 평가의 기능을 염두에 두고 문제를 출제하고 있다고 변호하기도 하였다.

P2 교사: 부끄럽지만 선생님 얘기를 듣고 알게 되었어요. 근데 대부분의 교사가 이원목적분류표를 나중에 작성한다고 해서 그 문항을 봤을 때 전부 지식만을 측정하는 단답식의 문항을 내진 않을 거라고 생각해요. 이원분류표는 없지만 문항을 내면서 그걸 생각하면서 문제를 만드니까요.

교사들은 문항을 출제하면서 무형의 내적 기준을 통해 문항이 요구하는 인지과정의 수준을 조정한다고 말하였음에도 불구하고 문항을 출제하고 보면 특정한 인지과정에 편중되는 경우가 발생함을 인정하였다. 이는 Table 7과 Table 8을 통해서도 확인할 수 있다. 따라서 편중되었음을 인식하는 경우에는 몇몇의 문항을 수정하여 측정하고자 하는 인지과정의 수준을 조정한다고 일부 교사들은 말하였다.

K1교사: 예를 들면 문제를 내고 보니 지식이나 이해가 많게 되면 문제 형식을 바꾼다고 해야 하나? 그러니까 '이건 뭘니까?' 이런 식으로 묻지 않고 '비슷한 사례를 들어보세요.'처럼 행동용어를 바꾸어서 문제를 수정해.

P1교사: 맞아요. 쓸려요. 그래서 나중에 검토할 때 수정해요. 문제 자체를 바꾸기도 하고, 같은 내용이지만 문제 유형을 바꾸기도 하고요.

면담에 참여한 대부분의 교사들은 문항의 초안을 수정하면서, 특정 인지과정에 편중되는 것을 막기 위하여 문제를 교체하거나 문제의 유형 또는 문항지시어를 바꾸는 등의 문항 조정을 하고 있었다. 기르 고자 하거나 변화시키고자 하는 학습능력에 따라 포함되는 동사의 형태가 달라야 한다는 측면에서(Kim et al., 2010) 볼 때 K1교사와 같은 문항조정 노력은 바람직하다고 할 수 있다. 예를 들어 학습능

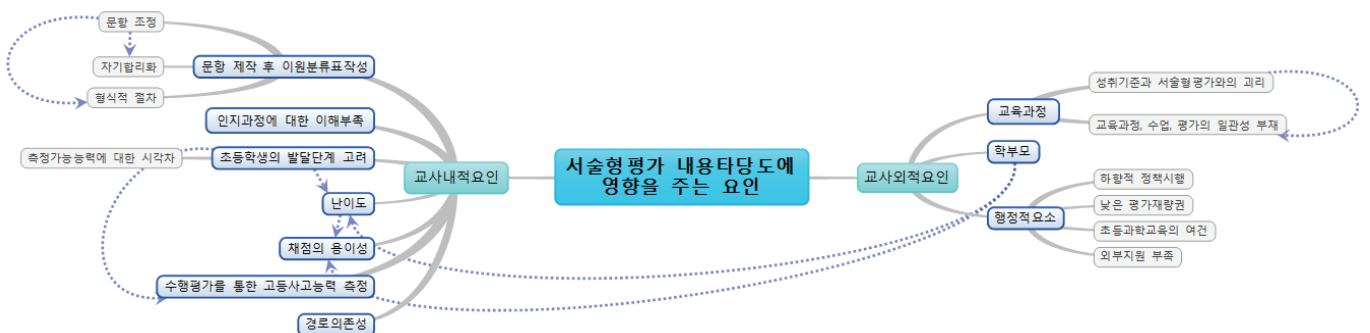


Figure 4. The factors affecting teachers' ability related to content validity of descriptive tests

력이 이해능력이라면 ‘그래프를 해석한다’로 기술되지만 적용능력이라면 ‘그래프를 그린다’로 기술되어야 하기 때문이다(Kim *et al.*, 2010). 하지만 모든 문항을 대상으로 적극적인 조정이 이루어지는 것은 아니며, K2 교사의 경우처럼 문항의 조정이 필요하다고 느낌에도 불구하고 아무런 조치 없이 넘어가는 경우도 발생함을 알 수 있었다. 따라서 전체적으로 볼 때 서술형 평가 문항에서 이원분류표의 평가목표와 평가문항의 행동영역 사이에 유의미한 차이가 발생한 것이다.

K2교사: 물론 편의성 때문인 거죠. 예를 들어 과학 같은 경우도, 녹말이나 식물 이런 부분에서 다르게 낼 수 있다는 생각이 들었지만 고치지 못했어요. 그런 생각을 아예 안 하면 아니지만... 창의성까진 아니더라도 이게 단답형이란 다른 게 뭐야. 고민이 들지만 고민하다 말았던 거죠. 교사가 자유로워지고 시간을 좀 더 쓴다면 평가의 질은 바뀔 수 있을 것 같고...

교사가 외면적으로는 이원목표분류표부터 작성하고 평가문항을 제작하지는 않지만 내면적으로 문항을 만들고 검토하는 과정에서 문항의 수준을 조정하는 절차를 거침에도 불구하고, 특정한 인지과정의 측정 문항 비중이 커지는 현상이 줄어들지 않는 이유로 면담 참여자 중 5명의 교사는 사후에 제작하는 이원목표분류표가 ‘자기합리화’의 절차를 겪기 때문이라고 설명하였다. ‘자기합리화’라는 표현은 면담에 참여한 P2교사가 언급한 용어로, 출제된 문항의 정확한 수준을 판단하지 못할 때 가능한 한 상위 수준으로 이원분류표에 표기하는 것을 의미한다.

P1교사: 이원분류표만 보면 골고루 평가하는 것처럼 보이게 하는... 교사 스스로의 위안 같은 거예요. 솔직히 이원목표분류표에서 수준은 잘 안보잖아요. 평가내용이나 답만을 보게 되니까... (중략) 실제보다 좀 더 상위 수준으로 체크할 때도 있는 거죠. ‘적용’을 묻는 것 같지만 ‘종합’이 비어 있으면 좀 이상하니까 거기에 표시하는 것 처럼요. (웃음) 누가 봐도 ‘지식’을 묻는 문항에는 그렇게는 안 해요. 어느 정도 여지가 보이면 그렇게 한다는 거예요.

P2교사: 다시 문항 내기는 귀찮으니까 이원목표분류표에 얼추 비슷하다 싶으면 거기에 표시해 버리는 경우가 다반사죠. 그리고 자기합리화를 합니다. 일단 핵심성취기준 내용을 고려해서 문항 출제를 하고 그 뒤에 행동영역 중 해당하는 걸 표시하는데 이 과정에서 한쪽으로 치우친다 싶으면 자기합리화과정을 통해서 ‘이 문항은 적용도 측정 할거야.’ 뭐 이런 식으로 간다는 거예요. 누가 봐도 아닌 거에는 체크를 안 하는데 이렇게 생각해보면 적용이라고 볼 수도 있지 않을까? 싶은 부분에는 체크를 하죠. 아주 조금이라도 여지가 있으면 체크를 할 때도 있고요.

P1교사나 P2교사의 사례처럼 교사의 ‘자기합리화’는 대부분 실제 시행된 서술형 평가에서는 낮은 수준의 인지과정을 물어보지만, 이원분류표의 인지수준은 높게 보이는 문제를 발생시켜 서술형 평가문항의 내용타당도를 낮추는 요인이 된다. ‘자기합리화’와 더불어 교사들이 이원분류표를 대하는 두드러진 태도는 이원분류표를 ‘형식적 절차’로 간주하는 경우였다. 이는 동료나 선배 교사와의 교차 검토와 같은 절차 없이 문항을 출제한 교사가 스스로 작성하여 제출하는 것

으로 완료되는 절차로 학교에서 시행되기 때문이다.

교사: 근데 사실 이원목표분류표의 내용이 사실 형식적인 면이 많죠. (중략) 학교 지침이 그렇게 출제하라고 하니 사실 표기만 그렇게 하게 되죠. K2교사: 이원목표분류표에서 빈칸으로 둘 수는 없으니까. 채워야 할 것 같으니까. 그게 없으면 모를까 (항목이) 있으니까 채워야 된다는 마음이 생기는 거죠.

‘형식적 절차’로써 이원분류표를 작성하는 경우와 ‘자기합리화’를 통한 이원분류표를 작성하는 경우는 명확히 구분이 되는 것은 아니다. 두 경우 모두 서술형 평가문항이 측정하는 인지과정보다 상위의 인지과정을 측정하는 것처럼 이원분류표를 작성한다는 결과에선 동일하다. 그러나 면담의 결과 ‘자기합리화’를 할 때보다 ‘형식적 절차’로 여기는 경우에 평가문항의 수준에 대한 교사의 고민이 적었으며, 실제 측정하는 수준과의 불일치가 더 심화되었다.

이원분류표 작성 시 발생하는 두 가지 문제 ‘자기합리화’와 ‘형식적 절차’는 서술형 평가문항의 낮은 내용타당도를 설명해주는 직접적 원인이 되지만, 깊게 들여다보면 Bloom의 인지적 영역 교육목표분류에 대한 교사의 이해 부족이 내용타당도가 높은 서술형 평가문항을 제작하는데 제약요소가 됨을 알 수 있었다.

C교사: 솔직히 제가 낸 문항이 과연 서술형 평가에 부합하는 문항들인지 확신이 서질 않아요. 그래서 그런지 이원목표분류를 작성할 때도 짚짚해요. 과연 이게 적용을 측정할 수 있는 문항인지, 과연 이게 이해를 측정하는 문항으로 맞는 수준인지 확신이 서질 않아요.

K3교사: 이 문항이 어디에 해당되는지 분류하는 게 생각보다 어려웠어요. 이해가? 적용인가? 하면서. 그건 이원목표분류를 사실 잘 모른다고 봐야죠. 그걸 배운 게 고작 학부 때 잠깐? 현장에 와서는 거의 뭐...

C교사나 K3교사가 가진 공통의 문제는 Bloom의 인지적 영역 교육목표분류에 대한 이해가 낮다는 것이다. 이처럼 교사의 평가전문성 수준이 낮은 상황에서 고등사고능력을 측정하는 서술형 평가문항이 시행되기 때문에 비록 시행된 시기는 오래 지속되었지만 근본적인 목표인 학생의 고등사고력 평가는 제대로 이루어지지 못하는 것이다. K3교사가 언급한 바와 같이 교사들이 평가기준을 설정하거나 그들의 검사문항의 질을 향상시키기 위해 필요한 분석 절차에 관해 배운 경험이 없다는 점은 매우 중요한 문제라 할 수 있다(Kwon *et al.*, 2012). 이는 앞서 지적한 바와 같이 교사가 된 이후에도 동료 및 선배 교사와 함께 문항 제작 및 검토 과정을 통해 이원목표분류표의 수준에 맞는 문항의 제작에 대한 경험이 충분히 누적되어야 고등사고력을 측정할 수 있는 문항 제작의 교사 역량이 길러지지만, 학교에서는 이러한 절차 없이 형식적으로 교사 개인이 문제를 출제하고 이원목표분류표를 작성하는 방식으로 진행되어 왔다는 점이 문제라고 할 수 있다. 그렇기 때문에 평가는 그냥 지나가는 과정으로 교사들이 인식하게 된 것이다.

C교사: 사실 평가 관련해서 제대로 고민해 본 경우가 적어요. 뭔가 평가를 지나가는 과정으로만 생각해서요. 본래 평가 후에 피드백이 있어야 하는데 보통 평가 후에 피드백 과정이 없거나 드물어서요. ‘평가가



끝이다.'라고 배우지 않았지만 끝인 것 마냥 느껴지는 경우가 많아요.

평가란 학생들에게 제공한 학습프로그램에서 교사가 의도했던 목표를 학생들이 어느 정도 달성하였으며, 그 학습프로그램의 효율성은 어떠한지, 교사의 교수방법은 학생들에게 적절한 것이었는지를 측정하고 그 결과를 교수학습 과정에 반영시켜 교수학습의 효율을 극대화시키기 위한 목적으로 행해지는 학습과정의 한 부분이다. 따라서 과학과의 학습 평가는 보다 효율적이고 바람직한 교수학습계획 수립은 물론 교수학습 과정, 교육과정 등의 개선을 위한 피드백까지 포함하여야 한다(National Research Council, 1996). 평가 결과에 대한 피드백이 이루어지지 않는다면 과학과 목표에 비추어 타당도가 높은 평가 도구였는가 하는 교사의 자기반성의 기회가 사라져 질 높은 서술형 평가도구가 제작되는데 부정적인 영향을 끼칠 것이다.

면담 과정에서 많은 교사들은 초등학생의 학습발달단계에 비추어 과연 어느 정도가 서술형 문항의 수준으로 적절한가에 대해 끊임없이 고민하고 있음을 확인할 수 있었다. 그러나 이에 대한 교사들의 생각은 차이를 보였다. 일부 교사는 초등학생들에게는 고등 사고력을 요구하는 것은 무리이기 때문에 학생들의 수준을 고려할 때 낮은 수준의 서술형 평가 문항을 만드는 것이 타당하다고 생각하였다.

**교사:** 초등수준에서는 이해 수준만 해도 괜찮다 싶기는 해요. 좀 더 높으면 적용까지. 왜냐하면 (학생들이) 연결을 짓지 못하는 경우가 많아서요. 몇 개 중 정답을 고르는 건 쉽지만 능동적으로 자신의 머리에서 관련 지식을 꺼내어 쓰는 건 그렇게 쉽진 않다고 생각해요. 게다가 학생, 학부모 눈치 때문인지... 대부분 이해, 적용 정도만 내고 있어요.

**K2교사:** 내가 말하는 어렵다는 건요. 애들이 다 풀어써야 해서 어렵다는 거예요. 애들이 나랑 얘기할 때 알고, 그게 물어가서 아는 걸 수도 있지만 꺼내주면 알 수 있는 것들인데 그거를 다 쓰라고 하니깐 너무 고등학생스럽고 대학생스러운 서술형이란 생각이 드는 거예요. (중략) 이번에 과학시험도 실험 설계하고 가설세우고 이런 게 너무 생소한 느낌? 애들도 이거 쓰느라고 시간이 너무 오래 걸렸어요.

**K4교사:** 저는 실험관찰에 있는 '생각해 볼까요' 수준으로 낸 것 같아요. (그 이상의 수준으로 내면) 애들이 못써내니까...

**P2교사:** (서술형 평가의 도입 취지를) 고려해서 출제하고자 하나 그렇게 출제했을 때 결과가 참담해서 다음번에는 좀 더 저등 사고기능을 측정하는 쪽으로 출제하는 경향이 생기는 것 같아요.

서술형 평가를 통해 고등사고력을 측정하는 것은 초등학교 학생들에게 적합하지 않다는 교사들의 생각의 타당성에 대해서는 논의가 필요하다. 그러나 일부 교사는 서술형 평가가 다른 형태의 평가와 다르게 높은 수준의 문항으로 제작이 되어야 한다고 생각하였다. 그리고 이러한 수준의 평가문항 제작이 안 되는 이유는 교사의 평가 문항 개발에 대한 이해 부족 때문이라고 생각하였다.

**K1교사:** 서술형 평가이니까 적용 이상의 과정 그러니까 적용이나 종합, 평가를 측정할 수 있어야 한다고 생각해. 내 생각은 그래. 아까 말한 그런 문제는 불륨의 이원목적분류 개념을 실제로 정확히 적용하고 확인하지 않고, 오타나 편집 상태, 정답에 오류가 있는지

이런 거를 확인하기에 급급해서 그런 거라고 생각해. 제대로 내는 경우도 많겠지만 단지 좀 길게 쓰는 주관식문항을 서술형 평가라고 인식하는 경우가 많은 것 같아.

교사들 중에는 학생들에게 요구해야 하는 인지과정의 수준과 문항의 난이도를 구분하지 못하는 경우도 있는 것으로 나타났다. 따라서 서답형 심화 문제를 서술형 문항으로 인식하여, 어려운 문제를 출제하면 그것이 고등사고과정을 묻는 문항이라고 생각하는 평가 문항 수준 판단의 오류가 발생하게 된다. 이러한 문제도 서술형 평가 문항에서 이원분류표의 평가목표와 평가문항의 행동영역 사이에 유의미한 차이를 발생시키는 원인이 된다.

**K4교사:** 우선 초등학교 아이들의 수준에 고등사고과정이 무엇인지에 대해서 이해가 되지 않아. 어디까지 물어보아야 하는 건지, 초등 수준에서 어디까지 고등사고과정이라고 해야 하는지를 모르겠다는 말이야. 그러다 보니까 대충 어려운 문제 출제하고 답을 문장으로 쓰게 하면 서술형이라고 생각하고 이게 고등사고과정을 측정하는 거라고 여기는 거지. 근데 사실 그건 서답형 심화문제이지 서술형 문항은 아닌 것 같아.

**P1교사:** 적용 이상의 문제를 출제하면 난이도가 너무 상향되는 문제도 있는 것 같아요. 그러니까 신경 써서 하다가도 결국 다시 비슷해지고. 뭐 그래요. 수준별로 골고루 평가해야 한다는 건 알고 있지만 채점의 용이성이나 난이도, 기타 여러 이유로 하기 어려우니까 (이원분류표에) 그냥 체크만 하는 거예요.

앞서 초등학생들의 발달단계를 고려하여 서술형평가를 통해 고등 사고력을 측정하기 어렵다고 생각한 교사들의 경우에도 학생들이 제대로 답을 하지 못한 이유가 학생들의 사고수준 때문인지 문항의 난이도 때문인지에 대한 판단이 필요하다. 과학 평가에서 교사들이 가장 많이 고민해야 하는 부분은 낮은 수준의 지식 습득여부를 확인하기 위한 평가에서 진일보해 어떻게 하면 높은 수준의 사고능력을 평가하고 이들에 대한 교육경험을 제공해 줄 수 있느냐 하는 것이다 (Kwon et al., 2012). 따라서 초등교사의 과학과 서술형 평가전문성은 바로 이러한 측면을 신장시키는 방향이어야 할 것이다. 초등학생들의 사고수준에 대한 교사 인식의 차이뿐만 아니라 교사들은 채점의 용이성 등 다른 요인들도 고려하여 평가 문항을 만드는 것으로 나타났다.

**P1교사:** 채점의 어려움이 있으니까. 채점의 용이함을 좀 염두 해두죠. 물론 단답형을 묻는 문제는 아니지만 부분점수를 나누기에 어느 정도 용이한 문제를 내게 되요.

**P2교사:** (웃음) 채점기준을 상세하게 정하는데 어려움을 겪으니까 서술형 문항인 척 하는 단답식 문항을 개발하기도 해요. 왜냐면 여러 가지 문제 때문에 문제를 채점하기 쉽게 내다보면 서술형으로 낼 수가 없는 거죠. (중략) 정해진 기간, 이를 테면 한 학기에 두 번, 세 달에 한번 보게 되잖아요. 그러면 세 달에 배운 걸 한 번에 평가한다는 얘기인데 그 세 달 동안 배운 걸 전부 다 서술형 평가로 한다면 문항 수가 많아질 수밖에 없고, 그러면 보는 애들이나 채점하는 사람이나 곤란해진다는 거죠.

사실상 채점의 용이성을 미리 염두하고 서술형 평가문항을 출제하

게 된다면 학생들의 다양한 응답을 제한하고 수렴적 사고를 유도하게 되어 서술형 평가가 그 기능을 제대로 하지 못할 것이라고 예상할 수 있다. 한편, 학교 현장에서 이루어지는 다양한 다른 평가들, 예를 들면 수행평가를 비롯한 수시평가를 통해 고등사고능력을 측정할 수 있으므로 서술형 평가를 포함한 지필 평가는 ‘지식’이나 ‘이해’ 정도만을 측정하는 것이 좋다는 의견도 있었다. 이는 지필평가로 고등사고능력을 요구하는 문항을 제작하는 교사 능력의 한계와도 관련이 있는 인식이라고 할 수 있다. 또한, 학교 현장에서 서술형 평가의 비율을 높여서 학생들의 고등 사고능력을 평가하고자 시도를 교사가 제대로 이해하지 못하기 때문인 것으로 볼 수 있다.

C교사: 전 서술형 평가에선 지식, 이해, 적용 이 세 수준이면 충분하다고 생각해요. 왜냐하면 현장에는 수행평가도 있으니까요. 수행평가에서 그 이상의 수준을 측정할 수 있다고 생각해요.

K2교사: 아무리 문제수가 줄어도 결국은 애들의 고차원적 사고를 이런 평가도구, 이런 형태를 가지고서 평가가 이루어져야 하는가에 대한 문제가 있어요. 평가는 그렇게 하지 않더라도 순간순간 이루어질 수 있는 거잖아요. 제가 노트검사를 엄청 꼼꼼히 하는데 그거 자체가 평가가 될 수 있는 거고. 어떤 때는 아이들과 필기하는 그 상황이 평가가 될 수 있는데 우리가 그 과도기에 있어서 그런지는 모르겠지만 정말 고차원적인 거를 얻을 수 있나? 전에 비해서 그럴 수 있다고 생각할 수 있을지 모르겠지만 교사의 입장에서는 얻어지는 게 의문스럽다는 거죠.

P2교사: 지필평가로는 적용 영역을 평가하는 건 어려워. 그러니까 수행평가로 해야지 하는 생각? 사실 수행평가가 잘 되면 범위를 묶어서 시험을 볼 필요가 없어요. 우리 학교 다닐 때 보면 시험 기간 때 비짜 공부해서 시험 치면 다 까먹잖아요. 서술형 보다 객관식이나 단답형은 그 까먹는 시간이 더 짧아진다는 점에서 평가의 효율성이 떨어지죠. (서술형 평가도구로) 그걸로 애들의 능력을 (측정) 하는 게 어느 정도 유효한 건 있겠지만 제대로 평가한다는 생각이 안 들죠. (중략) 지금 인터뷰를 하면서 느끼는 건데 서술형 평가문항의 취지를 달성하려면 지필평가보다 수행평가의 비중을 높여야 할 것 같아요. 그리고 지필평가는 단순 지식 정도를 묻는 거 정도로 해도 될 것 같아요. 절차적 지식이나 과정적 지식은 수행평가로 확인하되 그 때 그 때...

실제 상황에서 ‘수행’을 평가하는 것은 고등사고기능이 요구되는 상황을 고려하지 않는 전통적인 평가에 대한 해결책으로 평가전문가들이 제안한 것이다(Kwon *et al.*, 2012). 그러나 지필평가나 수행평가 등의 평가방법들은 과학적 사고력 및 창의적 문제해결능력 함양이라는 과학과의 교육목표를 달성하기 위해 학생들의 학습에 관한 의미 있는 정보를 교사와 학생에게 제공해주는 경로의 차이이며, 관찰 및 비형식적 질문 등과 함께 전체적 평가(holistic evaluation)를 이루어야 하는 것이다. 그러므로 수행평가가 정말로 학생들의 탐구능력을 측정하는가에 관한 타당성의 논란(Klassen, 2006; Pine *et al.*, 2006)은 차치하더라도 수행평가방식으로만 고등사고능력을 평가하겠다는 생각은 비합리적이며, 서술형 지필평가, 수행평가, 관찰 등 다양한 방법을 통해 학생들의 고등사고능력을 측정하고 평가해야 한다. National Research Council(2006)은 과학적 탐구를 통해 답할 수 있는 의문을 알아내는 것, 증거를 사용하여 묘사, 설명, 예상, 모델을 개발하는 것,

비판적이고 논리적으로 사고하여 증거와 설명을 연결하는 것, 대안적 설명과 모델을 인식하고 분석하는 것, 과학적 논증을 의사소통하고 방어하는 것이 지필평가로 평가될 수 있다고 밝히고 있다.

또한 면담에 참여한 대부분의 교사들은 현직 교사들의 ‘경로의존성’을 서술형평가가 제대로 시행되기 어려운 까닭으로 인식하고 있었다. ‘경로의존성’이란 용어는 P2교사가 면담 중에 도입한 용어인데, 법률이나 제도, 관습이나 문화 그리고 과학적 지식이나 기술에 이르기까지 어떤 것이라도 인간사회에서 한번 형성되어 버리면 그 후 외부로부터의 다양한 충격으로 형성할 때의 환경이나 여러 조건이 변경되었음에도 불구하고 그 내용이나 형태가 그대로 존속할 가능성(Arthur, 1994)을 나타내는 사회심리학적 용어이다.

P2교사: 경로의존성과 매너리즘에 빠진 교사들도 사실 많으니까요.

연구자: 경로의존성이란 것이 무엇을 말하는 거예요?

P2교사: 기존 방법에 익숙해져서 더 효과적이거나 효율적인 방법을 거부하는 거요. 세벌식 자판이 나왔음에도 결국 쿼티 자판으로 사용하고 있는 것 처럼요.

C교사: 지금의 교사들 세대가 서술형 평가에 익숙한 세대가 아니잖아요. 그 세대를 지나온 저는 서술형 평가에서 어떻게 문항을 출제하고 어떻게 채점하고 어떻게 이원목적분류표를 쓰는지 정확하지 않지 못해요.

K4교사: (인터넷에서 평가문제를 다운로드 받아 활용하는) 기존 방식에 답습하고 있는 교사들에게 갑작스럽게 서술형 평가문제를 출제하러니 어려운거야.

이러한 문제는 어느 세대에도 존재할 것이다. 따라서 교사는 끊임 없는 재교육을 통해 변화하는 교육 환경에서 효율적으로 학생들을 지도하고 평가할 수 있는 역량을 갖추어야 한다. 그렇지 않으면 교사들은 현 세대에 적응할 수 있는 인재를 기르는 것이 아니라 과거 자신이 학생이었을 때 요구되는 인재를 가르치는 오류를 반복하게 된다. 교육현장은 시대의 요구에 맞추어 변화하고 있지만 변화의 움직임에 저항하는 관성의 핵심에 교사들이 있다는 것은 큰 문제라고 할 수 있다. 특히 교육 목표의 도달을 확인하는 평가전문성 획득이라는 측면에서 볼 때 교사의 이러한 경로의존성은 반드시 해결되어야 할 과제이다.

지금까지 살펴본 바와 같이 초등학교 과학과 서술형 평가문항의 내용타당도가 낮은 까닭은 평가도구의 개발능력 등과 같은 교사내적 요인에 기인한 교사의 평가전문성 부족이 그 원인이었다. 하지만 면담을 통해 교사의 평가전문성에 영향을 주는 교사외적요인이 존재함을 알 수 있었다.

## 2) 교사외적요인

교사외적요인 중 교육과정과 관련한 요소들은 서술형 문항의 평가 목표와 직접적인 관련이 있어서 큰 영향을 주는 것으로 드러났다. 구체적으로는 서술형 평가문항이 요구할 것이라고 기대하는 인지과정의 수준과 실제 과학과 교육과정의 성취기준에서 요구하는 인지과정의 수준이 다르다는 문제가 있었다.

C교사: 여태까지 서술형 문항은 모두 교육청에서 내려온 핵심성취기준을 보고 제작했어요. 그런데 그 핵심성취기준을 보니까 정말 딱 그 단원에서 알아야 할 것들만 담고 있는 거예요. 거기에 근거하니까 적용을 넘어서는 높은 수준의 문제를 안 내도 된다고 생각한 거죠.

K2교사: 애들이 어려워하는 이유는 이거죠. 실험의 결과를 막 쫓아가서 했는데 이번 문제에서처럼 실험의 과정을 묻게 되면 여전히 저도 그렇고 애들도 그렇고 그... 학습목표에 도달하는 것이 목표가 되는 것인데 물론 그게 맞는데. 학습목표가 결과 지향적이란 말이에요. 그것만 알려고 하는 거니까...

N교사: 교육과정에서 정하는 과학교육의 수준이랑 (서술형평가에서 요구하는 수준과) 안 맞아서 괴리가 생긴다고 봐. 교육과정이 초등학교 수준에서는 단어 수준의 개념을 이해시키는 것에 초점이 맞춰져 있는데 이걸 서술형 문제로 바꾸기가 어렵다는 거야. 통합탐구기능이라고 하는 가설설정, 변인통제 이런 것이 고학년에서는 조금 이렇지게끔 교과서가 구성되었다고는 하지만 그 수준이 기초적이야.

앞서 초등학교 교사들이 초등학생들의 사고 수준으로 생각하였던 ‘적용’ 수준 이하의 낮은 인지과정 수준이 현재 과학과 교육과정의 핵심성취기준이 요구하는 인지과정 수준과 일치하는 것으로 교사들은 생각하였다. Choi & Paik (2015)의 연구 결과를 보면 2009 개정 초등학교 과학과 교육과정의 성취기준 가운데 ‘기억’과 ‘이해’의 인지 과정을 포함하고 있는 성취기준이 전체의 78.9%였다. 이러한 선행 연구결과에 비추어 본다면 위의 교사들의 인식은 타당하며, 따라서 성취기준을 바탕으로 고등사고능력을 측정하는 서술형 평가문항을 제작하는 것은 교사들에게 타당하지 않은 요구로 인식될 수 있다. 즉, 교육과정의 성취기준이 ‘지식’과 ‘이해’를 바탕으로 작성되었다는 사실은 서술형평가를 이보다 높은 수준으로 내기 어려운 문제를 야기한다.

K1교사: 사실 서술형평가가 잘 되려면 교육과정과 수업, 평가가 일관성을 가져야하지. 그래야 평가가 간단해지는 거고. 배운 내용과 별개 아니냐. 학생들의 사고과정을 중시하는 수업이 이루어졌다고 보고 서술형평가가 시행되어야 하는 건데 이게 과연 그럴까? 하는 생각이 든다는 거야.

K2교사: 이렇게 평가를 해야 한다면 담임이 그 반에서 했던 거를 중심으로... 그러면 사실 어디부터 다시 가야 하나면 교육과정부터 다시 가야 하는 거예요. 학급 안에서 이루어지는 교육과정 중심으로 평가가 가야 하는데 여기부터가 꼬이니까... 그러다 보니 저는 회의가 들더라고요. (중략) 왜냐면 왜 여기에 창의적인 게 없다고 생각했냐면 교과서에서 다 배우고 제시된 것들에 의해서 생각하는 거고 답이 정해져 있는 것들이라면 창의는 뭔가 생각하게 된다는 거죠. 평가를 비꾼다고 창의가 된다고 생각해 본 적은 없어요.

P2교사: 초등학생들의 대부분은 교사가 기대하는 수준의 답을 내지 못할 거예요. 현재까지 배워온 대로 공부했다면요.

이렇게 교육 내용의 수준과 서술형 평가에서 요구하는 수준의 불일치는 창의성과 같은 고등사고능력을 측정하는 서술형 평가의 시행을 어렵게 한다. 따라서 서술형 평가문항의 내용타당도가 낮은 까닭은 교사의 평가전문성 부족이라는 내적요인도 있겠지만, 그보다도 근본

적으로는 교육과정에서 요구하는 학생들의 성취기준과의 불일치라는 외적기준이 더 큰 문제일 수 있다. 특히 교육과정과 관련한 요소들은 교사의 평가전문성만으로 극복하기에는 한계가 있다. 따라서 평가방식의 변화를 통해 보다 높은 수준의 목표 성취를 꾀한다면 이에 걸맞은 교육과정의 변화가 우선되어야 할 것이다.

이밖에 교사외적요인 중에는 학부모의 문제도 제기되었다. 특히 평가문항을 제작할 때 학부모들의 눈치를 보게 된다는 언급을 하면서 논란이 없는 문제를 만들 수밖에 없는 교사의 입장이 많이 언급되었다.

J교사: 서술형 평가문항을 소신 있게 낼 수 있는 소신이 부재한다고 할까요? 배경 높은 서술형에서 점수를 깎으면 학부모의 눈초리들이...

N교사: 실제로 학교 현장에서는 학부모 눈치를 보며 논란이 없는 깔끔한 문항을 내기를 선호하잖아.

K2교사: 어디까지 내가 위험부담을 감수해야 하는가라는 생각이 드니까요. 2학기 때는 (서술형 평가문항을 제작할 때 교과서 외의 부분에서 발췌한) 다른 제재를 사용해보려고 생각은 하는데 그렇게 되면 진짜 어떤 위험부담이 있지요.

이는 학교 현장에서 행정을 담당하는 교장과 교감이 사회적 물의를 일으키지 않기 위해 학부모의 의견을 적극적으로 수용하는 학교 행정을 운영하는 것과 관련이 있으며, 과거와는 다른 학교 문화의 형성이라고 할 수 있다. 그러나 교육의 비전문가들인 학부모가 교사의 수업 및 평가에 의견을 내면서 교사의 전문성과 자율성이 직접적인 훼손을 받는 문제는 더욱 심각해지기 전에 처방이 필요한 부분이라고 본다. 이에 대해 J교사는 다음과 같이 말하였다.

J교사: 학부모들에게 휘둘리지 않을 수 있는 교사의 평가권 같은 게 필요해요.

한편 초등 교사들은 서술형 평가문항 제작에 영향을 미치는 교사외적 요인으로써 행정적 요소를 비중 있게 다루었으며 그 중에서도 정부가 학교로 전달하는 일방적인 정책시행에 대해 불만을 제기하고 있었다.

K4교사: 과학과 뿐 아니라 과연 서술형 평가가 잘 시행되고 있는 교과가 있을까 의문이 드네. 서술형 평가에 대한 현장의 이해도도 낮을 뿐만 아니라, 제대로 준비도 안 된 사람들에게 갑자기 ‘평가방식을 개선해야 하니 서술형으로 출제하고 채점하라고 통지하시오.’라는 일방적인 도입 때문에 서술형 평가는 그 본래 취지를 살리기 어렵다고 봐.

P1교사: 의무적으로 몇 퍼센트 이상이어야 한다고 하니까, 좀 억지로 내는 부분이 없지 않아 있어요.

P2교사: 아까도 말씀드린 대로 획일적이고 하향적인 정책 시행에서 오는 문제가 좀 있다고 생각해요. 저희 학년 예만 들어도 교육청-학교-연구부장-학년부장을 거쳐서 오는 서술형 평가방식을 접했을 때 이해도 수긍도 못하는 경우가 종종 있어요. 근데 여기에 대한 건의는 받아들여지지 않는 경우가 많으니 일단 원하는 대로 맞춰서 준비하는 과정에서 의도한 바대로 이루어지지 않는 경우가 생기는 거죠. 우리 생각이 반영이 안 되니까 무력감을 느끼는 거고, 제대로 안 되는 거죠.

학교 현장에서 이루어지는 근본적인 개혁과 변화는 이에 동참하는

교사의 전폭적인 지지와 이해를 우선해야 한다. 그렇지 않고 시행되는 정책들은 단발성의 구호적 차원에서 그치는 경우가 대부분이다. 아무리 좋은 정책과 해외 사례를 도입한다 하더라도 이를 위해 교사의 준비가 우선되어야 한다. 그러나 우리나라의 경우에는 확실적이고 지식 주입적인 단기 연수 이외에 교사의 전문성 신장 및 정책 시행 역량을 기를 수 있는 시간과 투자가 전무하다. 그렇게 되면 아무리 오래 정책이나 해외사례를 도입·시행하더라도 결국 그것은 근본적인 학교 교육의 변화를 이끌어내지 못한다. 이미 시행된 지 오래되었고, 어느 정도 학교 현장에 정착되어야 할 서술형 평가의 경우도 예외는 아니다. 서술형 평가 제작에 대한 교사들의 전문성 부족과 교육의 기회 부족은 면담을 통해 확인할 수 있었다.

C교사: 서술형평가를 도입하면서 이에 관한 연수나 매뉴얼 같은 게 적극적으로 배포되지도 이루어지지도 않았죠.

P1교사: 교육청에서 예시로 준 서술형 평가문항도 억지스러운 부분이 많더라고요. 교육청에서 배부하는 자료가 좀 좋으면 좋겠어요. 아! 이거다 하는 이웃나인이 있으면 다른 문항에 적용시켜서 출제할 수가 있으니까.

K3교사: 관련 문항이 개발되어 있으면 저희가 쓰기 쉽죠. 생으로 만들라하면 괴로운 거예요. 어렵기도 하고 시간도 걸리고 좋은 취지는 인정하지만 효율적인 방법을 찾는 거죠.

학교 단위나 교사들로부터의 자발적인 평가 개혁이 아닌 상급 행정기관으로부터의 도입이라는 점에서 서술형 평가가 형식적으로 수행될 수 있는 측면이 있음을 알 수 있었다. 특히 교육청 등에서 제시한 지침에 맞춰서 문항을 제작하는데 급급하게 된다면 서술형 평가문항의 내용타당도를 확보하기가 어려운 것은 자명한 사실이다. 이러한 인식은 면담한 교사들을 통해서도 확인할 수 있었다.

K2교사: 이 모든 게 계획서 내에서 평가해서 그런다니까요. 모든 권한을 담임한테 줘. 죽이 되던 밥이 되던. 1년에 2번을 하던 4번을 하던. 이걸 왜 거기에 맞춰서 하려니까... (중략) 차라리 학년 단위로 묶어서 교육과정을 재구성할 수 있는 기회를 주는 게 그런 시도를 해볼 수 있는 기회와 장을 마련해주는 게 도움이 될 거라고 봐요. 대신에 뭘 결과물을 내라 이리저 말고, 너무 교사들을 믿어주지 않는 것 같아요. 저 같은 사람은 내버려두면 혼자 했을 건데 이렇게 하라니까 기본 내뱉어서, 하기 싫은 사람이 되는 거예요. 창의적인 거를 하고 싶으면 교육과정을 재구성할 수 있게... 학교 프레임이 안 바뀌니까...

P2교사: 교육청에서는 학교에게 어느 정도의 자율권을 줘야 해요. 학교는 또 그만큼 교사에게 자율권을 보장해줘야 하고요. 그 보장되는 자율권 만큼 최저기준만을 제시해주면 된다고 생각해요.

교사들은 서술형 평가의 도입이 상급 행정기관으로부터 일방적으로 도입이 되었으며, 또 교사들에게 서술형 평가문항 출제와 관련하여 요구하는 것들을 부담으로 느끼고 있었다. 교사들은 이러한 문제와 관련하여 어느 정도의 재량권이 뒷받침되어야 한다고 주장하였다.

초등학교와 중등학교의 차이를 두지 않고 무조건 동일한 정책을 펴는 문제도 제기되었다.

K4교사: 특히 초등 교사들에게 문제가 많은 경우라 생각되는데, 중등의

경우 평가가 몹시 중요하기에 평가문제 출제에 엄청난 시간과 노력을 투자하지만 우리는 그렇지 않잖아. 초등학교의 특수성이 좀 반영되어야 하지 않을까? 초등의 경우 다수 교과를 혼자서 내잖아.

중등학교의 경우에는 상위 학교 진학 등의 문제로 평가가 매우 중요하고, 교사들은 각자 전공을 가지고 있으며 동일 과목의 교사들이 여럿 있어서 서로 협의하여 문제를 낼 수 있는 환경이다. 그러나 초등학교의 경우에는 한 교사가 여러 과목을 동시에 내야하며, 평가가 상대적으로 중등학교보다 중요하지 않으므로 이러한 차이를 고려하여 서술형 평가와 같은 정책도 고려되어야 한다는 것이다.

#### IV. 결론 및 제언

검사도구의 내용타당도 분석은 검사가 측정하고자 하는 내용을 측정하는지를 검정하므로 필수적인 절차이다. 그러나 초등학교 현장에서는 이러한 내용타당도의 검정 과정 없이 서술형 평가문항이 개발되고 시행되어 왔다. 따라서 이 연구에서는 A광역시의 여러 초등학교에서 실시되는 서술형 평가문항의 내용타당도를 객관적인 기준으로 분석함으로써 학교 현장에서 이루어지는 평가문항에 대해 점검하고, 보다 내용타당도가 높은 문항 개발에 노력을 기울일 필요가 있는지 알아보려고 하였다. 이를 위해 A광역시 소재 초등학교에서 시행된 5, 6학년 서술형 평가문항의 이원분류표상 평가목표의 행동영역과 실제 서술형 평가문항에서 측정될 것으로 판단되는 행동영역을 비율차 검정하였다. 그 결과 5학년 서술형 평가의 경우 ‘분석’을 제외한 모든 행동소에서 유의미한 차이를 보였다. 6학년 서술형 평가의 경우 ‘분석’, ‘평가’를 제외한 나머지 행동소에서 유의미한 차이를 보였다. 즉, 이원분류표 상에서는 고등사고능력이 측정되는 것으로 표기가 되어 있었으나 출제된 서술형 평가문항들은 저수준의 사고능력을 측정하고 있었다. 따라서 A광역시 소재 여러 초등학교에서 실시된 과학과 서술형 평가는 내용타당도가 낮다고 판단할 수 있다.

이와 같은 내용타당도 분석 결과를 심층적으로 이해하기 위해 면담을 실시하였다. 그 결과 초등학교 과학과 서술형 평가의 내용타당도에 영향을 줄 수 있는 교사내적요인과 교사외적요인을 도출하였다. 교사내적요인에는 바르지 않은 이원분류표의 작성법, 초등학교의 발달단계 고려, 난이도, 채점의 용이성, 경로의존성 등이 포함되었다. 교사들은 문항을 먼저 제작하고 난 후 이원분류표를 작성하는 잘못된 방식을 따르고 있었다. 그리고 ‘자기합리화’를 통해 이원분류표를 작성하거나 이원분류표를 ‘형식적 절차’로 간주하고 편의성을 추구하여 작성하기도 하였다. 또한 교사들은 서술형 평가형식이 초등학교생들의 발달단계에 비추어 적합한가의 문제를 고민하고 있었다. 이러한 발달단계의 적합성 문제는 난이도 하향 조절 및 채점의 용이성 추구하고 맞물려 서술형 평가문항이 고등사고능력을 측정하는 본래의 기능을 저해하였다. 그리고 새로운 평가방식을 제대로 받아들이지 않고 기존의 평가관을 고집하는 ‘경로의존성’의 문제가 면담에 참여한 교사들에 의해 제기되었다.

교사외적요인에는 교육과정 및 학부모 그리고 행정적 요소들이 관계하고 있었다. 교사들 가운데 일부는 서술형 평가가 추구하는 목적을 잘 알고 있지만 교육과정의 성취기준을 바탕으로 문제를 제작하다보면 그 목적을 달성하기 어려움을 말하였다. 이는 교육과정의 성

취기준의 행동 동사가 ‘기억’이나 ‘이해’ 등의 인지과정을 주로 지시하고 있기 때문이다. 그리고 이 때문에 교육과정과 수업 그리고 평가가 일관성을 갖기 어렵다고 말하였다. 한편 교사들은 서술형 평가문항의 경우 문제 자체나 혹은 채점 등에서 논란이 있을 수 있다는 점을 인식하고 학부모들의 눈치를 살피는 것으로도 나타났다. 더불어 서술형 평가 방식이 초등학교로 도입되는 과정이 하향적이었기 때문에 서술형 평가에 대한 충분한 소양을 갖추기에 불충분하였으며, 이에 따라 내실없는 서술형 평가의 시행을 부추긴 측면이 있다고도 말하였다. 이외에도 여러 교과 평가문항을 제작해야 하는 초등교육의 평가현실을 고려해야 한다는 의견, 교사에게 주어지는 평가재량권이 더욱 확보되어야 한다는 의견 및 서술형 평가문항 제작에 실질적 도움을 주는 연수 제공이나 문항 자료의 공유와 같은 외부지원이 필요하다는 의견이 제시되었다.

최근의 평가동향은 학습 결과를 측정하는데 목적을 두는 것 보다는 교수학습 과정으로써 학생들의 다양한 사고 과정 중심의 평가로의 전환이 강화되는 것이기 때문에 서술형 평가의 효과를 높이기 위해서는 서술형 평가에 대한 깊이 있는 이해와 학교 현장에서 평가구성, 문항제작, 채점, 결과 활용 등 일련의 평가과정에 필요한 실질적인 지식이 요구되고 있다(Assessment Reform Group, 2006). 이러한 관점에서 본 연구에서 얻은 결과와 시사점을 바탕으로 양호도 높은 서술형 평가도구를 개발하고, 평가 시행과 관련한 질 관리를 위하여 초등 과학교사의 서술형 평가전문성 역량을 제안하고자 한다(Figure 5). 아래에 제시된 초등 과학교사의 서술형 평가전문성 역량은 Kim & Juhn(2005)이 제시한 ‘학생평가전문성 기준’을 서술형 평가에 초점을 두어 수정한 것이다.

먼저 교사들은 양호도 높은 평가도구 개발 및 평가 결과의 분석, 활용을 위하여 학생들에게 요구되는 인지과정에 대한 소양을 갖추어야 한다. 서술형 평가는 고등사고능력을 측정하고 평가하는데 효과적인 만큼 교사들은 먼저 고등사고능력을 비롯한 인지과정 전반에 대한 이해가 선행되어야 하기 때문이다. 또한 인지과정에 대한 충분한 소양을 가지고 있지 못하면 서술형 평가도구의 타당도 확보에 큰 어려움을 겪을 수 있다. 다음으로 교사는 학생들의 고등사고능력 함양을 위한 수업 및 평가목표 재구성능력이 있어야 한다. 현행 교육과정의 성취기준은 ‘지식’과 ‘이해’ 수준을 요구하는 행동동사로 작성된 것이 많다(Choi & Paik, 2015). 따라서 이를 바탕으로 작성되는 수업목표 역시 ‘지식’과 ‘이해’를 요구하게 되며, 학습에 대한 성취의 기준이 되는 평가목표 역시 ‘지식’과 ‘이해’를 지향하게 될 것이다. 이러한

평가목표로는 학생들의 고등사고능력을 측정 및 평가하기가 어렵다. 따라서 교사는 교육과정의 성취기준을 바탕으로 고등사고능력 신장을 위한 수업목표를 재구성하고 이를 바탕으로 한 평가목표를 작성할 수 있는 역량이 필요하다. 평가도구의 개발능력은 평가의 목적과 내용에 적합한 평가도구를 개발하고, 평가도구의 질을 교사 스스로 점검하고 개선할 수 있는 능력을 말한다. 서술형 평가도구 개발능력에는 문항지시어의 사용 등 고등사고능력을 요구하는 평가사태의 조작 능력 등이 요구된다. 그리고 면담 속 초등 교사들은 서술형 평가에 대응하는 채점 및 성적 부여 능력이 부족하였으며, 이러한 능력의 결손은 채점이 용이한 문항 제작을 유도하여 타당도 높은 서술형 평가도구 제작을 저해하였다. 따라서 평가 계획에 부합하도록 평가를 실시하며, 정확하게 채점하고 평가의 목적에 부합하도록 성적을 부여할 수 있는가를 의미하는 평가 실시, 채점 및 성적 부여 능력이 서술형 평가전문성 역량으로 요구된다. 마지막으로 교사들이 서술형 평가의 결과를 전문성을 바탕으로 해석하는 능력이 부족하여 학부모들의 눈치를 살피며 타당도 높은 서술형 평가를 시행하지 못하는 현실 속에서 평가 결과를 타당하게 분석 및 해석하고, 평가 결과를 수업 및 학생에 대한 교육적 의사 결정에 활용하며, 학생, 학부모, 교육 관련자와 평가결과에 대해 정확하게 의사소통할 수 있는 능력 역시 초등 과학교사에게 필요한 역량이라 할 수 있다.

또한, 양호도 높은 서술형 평가도구의 제작 및 시행을 위해서는 교사 개인의 역량 강화뿐만 아니라 교사 외적인 지원 및 평가환경 조성도 함께 이루어져야 한다. 이에 다음과 같은 제언을 하고자 한다.

첫째, 학생들의 고등사고능력을 촉진하는 과학과 교육과정의 성취기준이 마련되어야 한다. 교육과정의 성취기준은 곧 교사의 수업에 반영되며, 이를 바탕으로 평가를 하게 된다. 따라서 현행의 ‘지식’ 및 ‘이해’ 위주의 사고력을 지시하는 성취기준이 과학적 사고력 및 창의적 문제해결력 신장이라는 과학과 교육목표에 부합하도록 개선되어야 할 것이다.

둘째, 실질적인 연수와 문항자료가 제공되어야 한다. 형식적이고 개론적인 연수에 그치는 것이 아니라 서술형 평가 과정에서 나타난 교사들의 평가전문성 부족을 파악하여 이를 극복할 수 있는 구체적 방안이 담긴 연수 및 문항자료가 개발되고 보급되는 것이 필요하다. 구체적으로 인지과정에 대한 이해, 수업 및 평가목표 재구성 능력, 서술형 평가도구 제작 기술 등에 관한 내용 등이 포함되어야 할 것이다.

셋째, 초등학교 평가환경이 고려되어야 한다. 철저히 담임제로 수업이 이루어지는 초등학교 교육여건 속에서 여러 과목의 서술형 평가

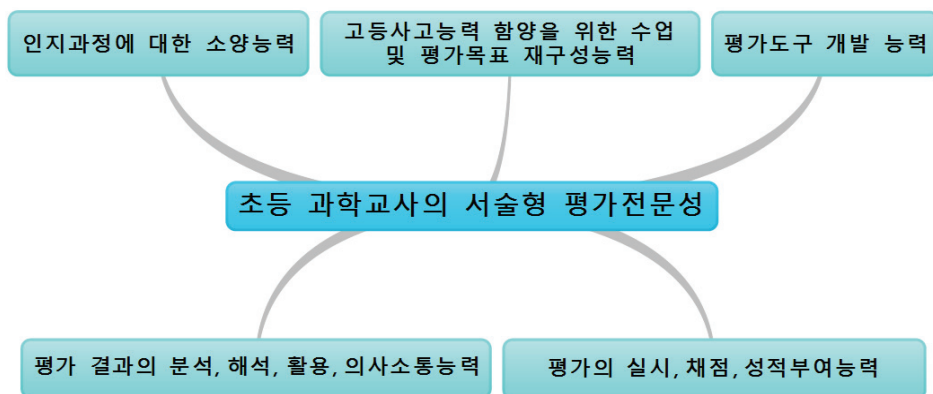


Figure 5. The abilities related to elementary school teachers' PCK of descriptive tests

문항 출제 및 채점 그리고 각종 행정적 절차 등은 초등교사에게 부담이 되는 것이 사실이다. 따라서 중등학교와는 다른 초등학교만의 특수한 평가환경에 대해 면밀히 조사하고, 이를 평가 정책에 반영하여야 할 것이다.

## 국문요약

본 연구의 목적은 초등학교에서 개발·시행된 서술형 평가 문항의 내용타당도를 분석하는데 있으며, 이 평가 문항들의 개선을 위한 기초자료를 제시하는데 연구의 의의가 있다. 이를 위하여 여러 초등학교의 서술형 평가문항을 수집하고, 이원분류표의 평가목표와 문항의 평가목표가 요구하는 행동소를 비율차 검정하였다. 분석의 결과 교사가 제작한 서술형 평가문항은 ‘지식’, ‘이해’를 주로 측정하고 있으며, 행동영역의 내용타당도가 낮음을 확인하였다. 내용타당도가 낮은 결과를 설명하기 위해 9명의 초등 교사를 대상으로 면담을 실시하였다. 면담의 결과 초등학교 과학과 서술형 평가문항의 내용타당도 확보를 저해하는 요인으로 교사내적요인과 교사외적요인을 추출하였다. 교사내적요인에는 바르지 않은 이원분류표의 작성법, 초등학생의 발달 단계 고려, 난이도, 채점의 용이성, 경로의존성 등이 포함되었다. 그리고 교사외적으로는 교육과정 및 학부모 그리고 행정적 요소 등이었다. 이상의 결과를 바탕으로 과학교사의 서술형 평가전문성을 위한 요인들을 제언하였다.

**주제어 :** 초등과학, 서술형 평가, 내용타당도, 이원분류표, 교사전문성

## References

- Arthur, W. B. (1994). *Increasing returns and path dependence in the economy*. Ann Arbor: University of Michigan Press.
- Assessment Reform Group (2006). *The role of teachers in the assessment of learning*. London, UK: Institution of Education, University of London. Retrieved from <http://www.nuffieldfoundation.org/sites/default/files/files/The-role-of-teachers-in-the-assessment-of-learning.pdf>
- Cho, A., Moon, K., You, M., & Kim, Y. (2013). A tool for assessing the middle school student's science writing and its reliability. *Biology Education*, 41(4), 675-684.
- Choi, E. (2011). *A research on teachers and students awareness and the characteristics of question in middle school science essay test*. (Master's thesis). Korea National University of Education.
- Choi, J., & Paik, S. (2015). A comparative analysis of achievement standards of the 2007 & 2009 revised elementary science curriculum with next generation science standards in US based on Bloom's revised taxonomy. *Journal of the Korean Association for Science Education*, 35(2), 277-288.
- Gullickson, A. P. (1984). Teacher perspectives of their instructional use of test. *Journal of Educational Research*, 77(4), 244-248.
- Hahn, A., Kang, H., Kwon, C., Kim, H., & Woo, J. (1997). *New elementary science teaching method*. Seoul: Education science company.
- Hong, M., & Chung, H. (2006). The development of assessment items with essay & description type test science in biology for high school students. *Journal of Science Education*, 30, 65-100.
- Kang, H., Chöng, C., & Choi, Y. (2005). An alternative exploration of Bloom's taxonomy of educational objectives : On the basis of the teachers' interview. *Secondary Education Research*, 53(1), 51-84.
- Kang, H., Gong, Y., Kwon, H., Kim, J., Bae, J., Song, M., Shin, Y., Yang, I., Yoon, H., Lee, D., Lee, M., Im, C., Im, H., Jang, S., Jeon, Y., & Chae, D. (2007). *Elementary science education theory*. Seoul: Education Science Company.
- Kim, D., & Kim, S. (2011). *Curriculum and educational evaluation*. Seoul: Hakjisa.
- Kim, E., & Han, S. (2010). The analysis of national assessment of educational achievement items using three dimensional scientific assessment framework. *Journal of the Korean Society of Earth Science Education*, 3(3), 248-256.
- Kim, E., & Kang, W. (2010). An analysis of content validity of third-grade mathematics achievement tests. *Journal of Elementary Mathematics Education in Korea*, 14(2), 177-196.
- Kim, H., & Yoo, J. (2010). An analysis on reliabilities of scoring methods and rubric ratings number for performance assessments of middle school students' science investigation activities. *Journal of the Korean Association for Science Education*, 30(2), 275-290.
- Kim, J., Boo, J., So, K., & Yang, K. (2010). *Curriculum and educational evaluation*. (4th ed.). Paju: Kyoyookgwahaksa.
- Kim, S., & Jhun, Y. (2005). For better evaluation of students in the science class (Research results, ORM 2005-51-5). Seoul: Korea Institute for Curriculum and Evaluation.
- Kim, Y., & Kim, Y. (2012). The development of a free-response test for the assessment of science process skills. *Biology Education*, 40(1), 167-177.
- Klassen, S. (2006). Contextual assessment in science education: Background, issues and policy. *Science Education*, 90(5), 820-851.
- Kwon, J., Kim, B., Choi, B., Kim, H., Paik, S., Yang, I., Kwon, Y., Cha, H., Woo, J., & Jeong, J. (2012). *Science education*. Seoul: Kyoyookgwahaksa.
- Kyungkido Education Office. (2012). *Guidebook of descriptive and essay evaluation*. Kyungkido: Kyungkido Education Office.
- Lee, D., & Jeong, E. (2014). An analysis of paper and pencil test items of life science I in high school. *Journal of Science Education*, 38(3), 670-690.
- Lee, K., & An, H. (2005). Analysis of assessment types, scoring methods and reliability of science performance assessment in middle and high school. *Journal of the Korean Association for Science Education*, 25(2), 173-183.
- Lee, M. (2013). *An analysis of the current situation regarding high school level written examinations for chemistry*. (Master's thesis). Korea National University of Education.
- McAfee, G. (2011). *Master math: AP statistics*. Boston: Course Technology, a part of Cengage Learning.
- Ministry of Educational and Science Technology. (2009). *Elementary and secondary school curriculum : General statement* (Notification No. 2009-41 of the MEST). Seoul: Ministry of Educational and Science Technology.
- Ministry of Educational and Science Technology. (2012). *Science curriculum* (Notification No. 2011-361 of the MEST). Seoul: Ministry of Educational and Science Technology.
- National Research Council. (1996). *National science education standards*. Washington D.C.: National Academy Press.
- National Research Council. (2006). *Systems for state assessment. Committee on test design for K-12 achievement*. Washington D.C.: National Academy Press.
- Paik, S., Lee, E., Kim, J., Song, Y., Kim, Y., Chung, J., & Han, J. (2008). Analysis of the content validity of the achievement evaluation items on the "water" chapter in the high school chemistry I course. *Journal of Research in Curriculum Instruction*, 12(1), 55-65.
- Paik, S., & Ryu, H. (2014). High school students' perceptions in descriptive assessment activity experiences by teacher or by peer. *Journal of the Korean Association for Science Education*, 34(6), 593-599.
- Park, H. (2012). Analysis of summative evaluation objectives in middle school science examination by Klopfer's taxonomy of educational objectives. *Journal of Science Education*, 36(2), 293-302.
- Park, I., & Kang, S. (2012). The development of assessment tools to measure scientific creative problem solving ability for middle school students. *Journal of the Korean Association for Science Education*, 32(2),

- 210-235.
- Pine, J., Aschbacher, P., Rothem, E., Jones, M., McPhee, C., Martin, C., Phelps, S., Kyle, T., & Foley, B. (2006). Fifth graders' science inquiry abilities: A comparative study of students in hands-on and textbook curricula. *Journal of Research in Science Teaching*, 43(4), 467-484.
- Shim, A. (2008). The study on the recognition of teachers and students & actual conditions about middle school science essay test. (Master's thesis). Korea National University of Education.
- Seong, T. (2014). The foundation of educational evaluation. (2nd ed.). Seoul: Hakjisa.
- Song, J. (2003). An analysis of content validity of teacher-made summative evaluation. (Master's thesis). Korea National University of Education.
- Song, K., Lee, H., & Lim, C. (2004). Development of a test of science inquiry skills for elementary school fifth and sixth graders. *Journal of the Korean Association for Science Education*, 24(6), 1245-1255.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271-286.
- Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.
- Yang, G. (2006). *Descriptive and essay assessment source book*. Seoul: Seoul Metropolitan Office of Education.
- Yang, I., Na, J., Lim, S., Lim, J., & Choi, H. (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.