

초등학교 교사의 과학 평가 전문성에 대한 자가진단 내용 분석

강훈식 · 강석진[†]

(춘천교육대학교) · (전주교육대학교)[†]

Analysis of Elementary School Teachers' Self-diagnosis on Their Competency for Assessment in Science

Kang, Hunsik · Kang, Sukjin[†]

(Chuncheon National University of Education) · (Jeonju National University of Education)[†]

ABSTRACT

In this study, we analyzed elementary school teachers' self-diagnosis on their competency for assessment in science. A questionnaire was administered to 217 elementary school teachers and in-depth group interviews with 6 teachers were also conducted. The results revealed that the mean scores of five subcategories, 'choosing assessment methods', 'developing assessment instruments', 'administering, scoring, and grading', 'analyzing, interpreting, using, and communicating assessment results', and 'recognizing ethics in assessment', were 3.82, 3.50, 3.92, 3.76, and 4.09 on a scale of 5 points, respectively. Teachers having more teaching experience showed higher mean scores in all subcategories. However, only the teachers having more than 20 years of teaching experience were statistically better than those having less than 5 years of teaching experience in four subcategories except for 'recognizing ethics in assessment'. There were no significant differences in terms of major, whereas the mean scores of the teachers having master's degree were statistically higher than their counterpart in some subcategories. In addition, the teachers who had completed training for assessment exhibited statistically higher mean scores than their counterpart in four subcategories except for 'recognizing ethics in assessment'. Educational implications of these findings are discussed.

Key words : assessment in science, competency, self-diagnosis, elementary school teacher

I. 서 론

평가는 학생의 학습과 성취에 대한 교사의 의사 결정을 뒷받침하는 정보를 수집하고 해석하는 활동으로(McMillan *et al.*, 2002), 평가 결과는 학생들의 학습 목표 달성 정도 및 다양한 인지적·정의적 특성의 파악, 설정된 교수 목표의 타당성 및 학습 지도 방법의 유효성 점검 등을 위한 정보를 제공하여 교수·학습의 질 개선을 위한 자료로 사용된다

(Kwon *et al.*, 2012). 즉, 평가는 수업의 질 향상과 교육과정의 충실한 운영을 위해 필수적인 요소이며(Kim, 2002), 평가 없이는 효과적인 교수·학습 활동이 이루어지기 어렵다(American Federation of Teacher *et al.*, 1990). 실제로 교사의 평가는 학생들의 학습 향상에 효과적인 것으로 보고되고 있다. 예를 들어, 형성평가의 교수 효과를 조사한 연구를 종합한 Black and Wiliam(1998)의 연구에 따르면, 형성평가의 효과 크기(effect size)는 0.4~0.7로 가장

효과적인 교육 처치 중 하나였다. 따라서 교사의 평가 전문성 향상은 성공적인 교수·학습 및 교사 전문성 발달을 위한 핵심적인 요소라고 할 수 있다 (Cho & Ko, 2008; Falk, 2012; Min, 2012; Park & Oliver, 2008; Veal & MaKinster, 1999).

융합인재교육 도입과 함께 학생의 적성과 소질 개발에 중점을 둔 2009 개정 교육과정(Ministry of Education, Science and Technology, 2011)에 부합하도록 평가 제도가 개선되어야 한다는 사회적 요구에 따라, 성취평가제가 도입되어 2012학년도 중학교 1학년 입학생부터 시행되고 있다. 초등학교의 경우에는 단편적인 암기 위주의 선택형·단답형·결과중심 평가를 서술형·논술형·과정중심 평가로 개선하는 상시평가제가 2013학년도부터 본격적으로 시행되고 있다. 그러나 지역 교육청의 제도, 학교의 교육 철학, 담임교사의 수업 방향 등에 따라 평가 시기, 횟수, 방법 등이 다양한 실정이다. 성취평가제나 상시평가제는 평가를 직접 실행해야 하는 교사의 역할에도 큰 변화를 요구하고 있다. 예를 들어, 교사는 성취기준을 설정하고 학생들이 성취해야 할 지식, 기능, 태도의 특성을 설명하는 학기 단위 성취수준을 작성해야 한다. 또한, 성취기준 도달 정도를 측정할 수 있는 평가 도구를 개발하고, 지필평가 및 수행평가의 결과를 종합하여 개별 학생의 성취수준을 최종 판단하는 등의 평가 활동을 하게 된다(Kim *et al.*, 2012). 즉, 학교 현장에서 교사들이 직접 성취기준에 따른 평가 계획, 평가 도구 개발, 평가 결과 분석과 해석 및 활용 등의 과정을 거쳐야 하므로, 교사들에게 이를 위한 전문성이 요구된다. 특히 과학과의 경우 개념 이해뿐만 아니라 창의력, 탐구능력, 태도 등에 대한 평가도 중요하게 다루어져야 하므로(MOEST, 2011), 교사들의 과학 평가 전문성을 향상시키기 위한 방안을 적극적으로 모색할 필요가 있다.

그러나 지금까지 교사의 과학 평가 전문성과 관련된 연구는 매우 부족하다. 일부 진행된 연구들도 주로 특정 평가 영역이나 방법 및 평가 수준에 대한 교사의 인식(Kim & Hyun, 2005; Kim & Kim, 2002), 평가 실태(Kim *et al.*, 2000), 특정 평가 방법의 효과(Choi *et al.*, 2000) 등을 조사하는 측면에 치중되어 있고, 교사의 과학 평가 전문성 향상에 관한 연구(Kim, 2006; Min, 2012)는 매우 부족하다. 특히 여러 과목을 동시에 가르쳐야 하기 때문에 관련

전문성이 부족할 것으로 예상되는 초등학교 교사들을 대상으로 한 연구는 찾아보기 힘들다. 또한, 초등학교 교사양성 과정이나 각종 직무연수 과정에서 과학 평가 전문성 함양을 위한 교육의 기회가 충분하게 제공되고 있지 못한 실정이다. 따라서 초등학교 교사의 과학 평가 전문성 제고에 관한 보다 다각적이고 심층적인 연구가 요구된다.

이를 위해서는 우선 초등학교 교사들의 과학 평가 전문성 수준을 체계적으로 파악할 필요가 있다. 이를 통해 교사들의 과학 평가 전문성 향상 방안을 모색하는 데 유용한 정보를 얻을 수 있기 때문이다. 그러나 지금까지는 과학과 수행평가에 관한 인식이나 실태(Jang & Kim, 2002; Kim, 2005; Ko *et al.*, 2014) 및 과학 평가관(Noh *et al.*, 2009) 등과 같이 특정 측면에서 초등학교 교사의 과학 평가 전문성 수준을 조사한 연구만이 진행되었다. 즉, 초등학교 교사가 과학 평가를 수행하는 데 필요한 총체적인 전문성 수준 관련 정보는 현재로서는 매우 부족한 실정이다. 이에 이 연구에서는 초등학교 교사의 과학 평가 전문성에 대한 자가진단 내용을 체계적으로 분석하였다.

II. 연구 방법 및 절차

1. 연구 대상 및 절차

이 연구에서는 서울, 경인, 춘천 교육대학교의 초등 과학교육 관련 대학원 강좌나 교사 직무연수를 수강하고 있는 교사 217명을 대상으로 설문 조사를 실시하였으며, 연구 대상의 구체적인 특징은 Table 1과 같다. 또한, 6명의 교사들을 편의 표집하여 집단 면담을 약 60분 동안 실시하였다.

2. 검사 도구 및 방법

설문지는 한국교육과정평가원에서 개발한 ‘교사의 학생평가 전문성 진단도구(Nam *et al.*, 2006)’를 사용하였다. 이 설문지는 전국의 평가 전문가들과 교사들을 대상으로 한 타당화 작업을 통해 타당도가 확보된 상태이며, 5개 하위 영역에 대한 총 60개의 5단계 리커트 척도 문항으로 구성되어 있다. 즉, 평가 유형과 평가 방법에 대한 이해 및 선정 능력과 관련된 ‘평가 방법의 선정’ 영역 13문항, 평가 도구의 개발 및 선정 능력과 관련된 ‘평가 도구의

Table 1. Teachers' characteristics

Characteristics	Subcategories	Frequency (%)
Gender	Male	70 (32.3)
	Female	147 (67.7)
Location of school	Gangwon-do	175 (80.6)
	Seoul Metropolitan area (Seoul, Gyeonggi-do, Incheon)	38 (17.5)
	Others	3 (1.4)
	No response	1 (0.5)
Teaching experience	Less than 5 years	58 (26.7)
	5 to 10 years	45 (20.7)
	10 to 20 years	57 (26.3)
	More than 20 years	57 (26.3)
Major	Science	24 (11.1)
	Non-science	132 (60.8)
	No response	61 (28.1)
Academic degree	Bachelor	173 (79.7)
	Master	30 (13.8)
	No response	14 (6.5)
Completion of training	Yes	115 (53.0)
	No	101 (46.5)
	No response	1 (0.5)

개발' 영역 13문항, 평가의 실시와 채점 및 성적 부여 능력과 관련된 '평가 실시·채점·성적 부여' 영역 9문항, 평가 결과의 분석과 해석 및 활용 능력뿐만 아니라 평가 결과에 대한 관련자들과의 의사소통 능력과 관련된 '평가 결과의 분석·해석·활용·의사소통' 영역 18문항, 전반적인 평가 과정에서의 윤리성과 관련된 '평가의 윤리성' 영역 7문항으로 구성되어 있다. 그리고 각 영역에 대하여 평균이 4.5점 이상이면 '평가 전문성이 높음', 3.5점 이상 4.5점 미만이면 '평가 전문성이 있음', 2.5점 이상 3.5점 미만이면 '평가 전문성이 낮음', 2.5점 미만이면 '평가 전문성이 매우 낮음'으로 해석하도록 그 준거가 제시되어 있다. 이 설문지는 특별한 수정 없이 특정 상황을 설정한다면 모든 과목에서 광범위하게 사용될 수 있으며(Nam *et al.*, 2006), 실제로 일부 특정 과목에서 활용된 바 있다(Lee *et al.*, 2011).

이에 이 연구에서는 교사들이 과학 평가 상황에 한정하여 응답할 수 있도록 하기 위해, 설문지의 제목을 '교사의 과학학습평가 전문성 수준 자기진단 설문지'로 제시하였다. 또한, '이 설문은 선생님

들께서 생각하시는 본인의 과학학습평가 전문성 수준에 대한 인식을 조사하기 위한 것으로, 총 60 문항으로 구성되어 있습니다.'는 안내문을 추가하였다. 설문 직전에는 연구 참여 교사들에게 과학학습평가에 한정하여 응답해줄 것을 재차 강조 및 안내하기도 하였으며, 집단 면담에서 이를 확인할 수 있었다. 교사의 배경 변인에 따른 차이를 조사하기 위해 교사의 배경 변인을 조사하기 위한 문항도 포함시켰다. 이 연구에서 구한 신뢰도 계수(Cronbach's alpha)는 '평가 방법의 선정', '평가 도구의 개발', '평가 실시·채점·성적 부여', '평가 결과의 분석·해석·활용·의사소통', '평가의 윤리성' 영역에 대하여 각각 0.89, 0.93, 0.86, 0.94, 0.87로 높게 나타났다.

집단 면담은 설문 응답의 원인을 자세히 조사하기 위한 형태로 이루어졌다. 즉, 1개 초등 과학교육 대학원 강좌의 일부 시간을 할애하여 6명의 수강 교사들에게 설문지에 응답하도록 하였고, 응답 직후 각 문항에 대한 응답 이유에 대하여 자세하게 설명하도록 요구하였다. 답변 내용이 충분하지 않은 경우에는 반복적으로 질문하여 교사들의 생각을 가능한 충분하게 이해하기 위해 노력하였다. 이때, 교사들이 허용적인 분위기에서 자유롭게 응답할 수 있도록 하기 위해 면담은 강사와 수강생의 라포르가 충분히 형성된 강좌의 후반부에 실시하였다. 모든 면담 과정은 디지털 카메라로 촬영하였고, 녹음 내용은 전사하였다.

3. 분석 방법

문항별 및 하위 영역별로 평균과 표준편차를 산출하였고, 문항별로는 긍정 응답(4=그렇다, 5=매우그렇다)의 비율도 산출하였다. 교사의 배경 변인(교직 경력, 전공, 학력, 연수 이수 여부 등)에 따른 각 하위 영역의 평균을 비교하기 위해 일원변량분석(One-way ANOVA) 또는 독립표본 t-검증을 실시하였다. 또한 교사들과의 집단 면담 동영상 및 전사본을 반복적으로 분석하면서 의미 있다고 판단되는 내용들을 추출한 후, 추출한 내용의 타당성을 점검하고 대표적인 응답 사례를 선정하여 기술하였다.

III. 연구 결과 및 논의

1. 과학 평가 전문성에 대한 자기진단 내용 분석

초등학교 교사의 과학 평가 전문성에 대한 자가 진단 내용을 분석한 결과는 Table 2와 같다. 5점 만점 중에서 ‘평가 방법의 선정’ 영역의 평균은 3.82점으로 ‘평가 전문성이 있음’에 해당되었다. 문항별로는 3.33~4.25점의 다양한 분포를 보였으며, 특히 학생이 달성해야 할 학습 목표 확인 및 평가할 학습 목표의 진술(평균 4.23, 긍정 91.2%)과 진단·형성·총합평가의 목적과 특징의 이해(평균 4.25, 긍정 90.3%), 양적평가와 질적평가의 특징과 장·단점 이해(평균 3.94, 긍정 79.7%), 학생의 성취 수준을 파악하기에 적합한 평가 방법 선정(평균 3.95, 긍정 80.6%) 측면에 대한 인식이 매우 긍정적이었다. 즉, 많은 교사들이 대체적으로 평가 방법의 유형에 대한 이해 및 선정과 관련된 자신의 전문성 수준에 대하여 높게 인식했음을 알 수 있다. 다음 면담 사례에서 나타나듯이, 이는 많은 교사들이 교육대학 교육과정, 임용시험 준비 과정, 교사 직무면수 과정, 교직 경험 등과 같은 다양한 경로를 통해 평가 방법의 선정에 필요한 지식을 습득했기 때문으로 보인다.

임용 공부할 때 다 외워서 공부했던 거라서 뭐지는 이제 알고 있는데...

연수를 통해서도 선생님들이 많이 평가에 대해서... 수업 평가 관련된 연수가 있어요. 원격 연수도 하고 직무연수도 하고. (중략) 상시평가하면서 평가 관련 연수가 많이 늘어났어요. 서술형 평가 문항 만들고 해야 하니까.

잘 하고 있다라는 생각은 들고 있지 않아요. 그런데, 초임 별령 낮을 때보다 경력이 쌓인 지금은 훨씬 더 이런 보는 눈이나 평가에 대한 눈이나 분석하는 시각이 많이 풍부해졌다고는 생각이 들거든요.

Table 2. Independent *t*-test results for teachers' self-diagnosis on their competency for assessment in science by major

Category	M	SD
Choosing assessment methods	3.82	0.49
Developing assessment instruments	3.50	0.65
Administering, scoring, and grading	3.92	0.53
Analyzing, interpreting, using, and communicating assessment results	3.76	0.52
Recognizing ethics in assessment	4.09	0.58

그러나 기준 참조 평가와 준거 참조 평가의 특징과 장단점에 대한 이해(평균 3.33, 긍정 46.0%), 교수·학습 과정에서 학생이 겪는 어려움을 진단하기에 적합한 평가 방법 선정(평균 3.63, 긍정 59.3%), 특별한 지도가 필요한 학생을 변별하기에 적합한 평가 방법 선정(평균 3.64, 긍정 60.8%) 측면에 대한 긍정적인 인식은 상대적으로 낮았다. 이는 예비 및 현직 교사 교육과정에서 이 부분들에 대한 교육이 강조되지 않았거나 학습 기회가 적었기 때문일 수 있다.

‘평가 도구의 개발’ 영역의 평균은 3.50점으로 ‘평가 전문성이 있음’과 ‘평가 전문성이 낮음’ 경계에 해당되는 것으로 나타났다. 문항별로 분석한 결과에서도 평가의 목적과 내용에 적합한 평가 도구 선정(평균 3.84, 긍정 76.5%), 타당도 개념 이해 및 타당도 평가 실제(평균 3.66, 긍정 64.1%), 신뢰도 개념 이해 및 신뢰도 평가 실제(평균 3.65, 긍정 61.3%), 문항 난이도 개념 이해 및 문항 난이도 평가 실제(평균 3.69, 긍정 65.0%), 문항 변별도 개념 이해 및 문항 변별도 평가 실제(평균 3.60, 긍정 59.9%) 문항만이 ‘평가 전문성이 있음’에 해당되었다. 나머지 문항들의 경우에는 3.25~3.50점의 ‘평가 전문성이 낮음’에 해당하는 점수 분포를 보였다. 즉, 많은 교사들이 평가의 목적과 내용에 적합한 평가 도구를 선정하거나 신뢰도, 문항 난이도 및 변별도를 이해 및 적용하는 데에는 자신이 있으나, 오답지 매력도 등에 대한 이해를 바탕으로 평가 도구를 직접 개발하거나 채점자 신뢰도를 이해 및 적용하는 능력은 충분하지 않다고 인식하였다. 이는 교사들이 평가 문항을 직접 개발하기보다는 교사용 지도서, 각종 기관에서 제공하는 평가문항 자료집, 상업용 문제집, 웹사이트나 교사 커뮤니티에 탑재된 문항 등을 변형하여 사용하는 경향과 관련이 있을 것이다. 즉, 교사들이 새로운 평가문항을 직접 개발하는 기회나 경험 및 동기 부여, 특히 자신의 평가문항 개발 과정과 결과물에 대한 전문가나 경력 교사들의 실질적인 조언과 피드백을 받을 수 있는 기회가 부족하여 이런 결과가 나타났다고 볼 수 있다. 다음은 새로운 평가문항의 개발 기회나 경험 및 동기 부여 부족과 관련된 면담 사례들이다.

모든 선생님들한테 문항 제작의 기회가 주어지지 않아요. 대부분 있는 문제를 고쳐서 쓰거나 그렇게 해왔고.

(중략) 자신 있게라는 말이 붙으면 자신 없죠.

에듀넷 같은 경우에는 평가문항을 제작할 수 있는 소프트웨어가 있어요. 에듀넷에 들어가면 시험지를 어떠한 형태로 할 것인지하고 단원만 클릭을 해놓고 평가하게 되면, 거기에 이제 여러 가지 시험문제가 나와요. 문제 은행이 있어요. 그것만 딱 클릭하면, 제가 알아서 상 몇 개, 중 몇 개, 하 몇 개 해가지고 배치가 돼서 딱 나와요. 시험지를 만들어주는 거죠. (중략) 이거 말고도 천재교육에서 운영하는 T셀파라는 데도 있고, 강원도에는 에듀원도 있고, 문제는 많이 나와요.

동기 부여가 되게 중요한 것 같은데, 제가 열심히 몇 주간 고민해서 고치고 고치고 해서 낸 적도 있고, 이번 같은 경우에는 대충 이렇게 문제 썩썩 붙여가지고 중경어 미만 비슷하게 하고, 너무 심한 문제는 조금 고치고 냈거든요. 하루만에. 몇 주간 투자한 문제나 하루 만에 똑딱 만들어낸 문제나 투입했을 때 그 다음에 다른 게 없다는 거죠. 개인의 만족 빼고는. (중략) 학습 목표를 충실하게 이해했는가 고민하고 내는 거나, 아니면 이미 학습 목표를 이미 충분히 알고 거기에 맞는 문제를 찾아오나 비슷하더라고요.

‘평가 실시·채점·성적 부여’ 영역의 평균은 3.92점으로 ‘평가 전문성이 있음’에 해당되었으며, 특히 평가 실시나 채점과 관련된 문항들의 경우에는 4점 이상의 높은 점수를 보였다. 즉, 대부분의 교사들이 학생에게 평가 계획에 대하여 사전 공지하고 있고(평균 4.12, 긍정 82.0%), 평가 실시에 적합한 물리적·심리적 환경을 조성 및 관리하고 있으며(평균 4.00, 긍정 79.7%), 평가 과정에서 학생들을 효과적으로 지도 및 감독하고 있고(평균 4.43, 긍정 93.1%), 채점 기준을 정확하게 숙지하여 채점하고 있으며(평균 4.34, 긍정 94.0%), 학생의 답안을 점검하여 채점 기준을 수정 및 보완할 수 있고(평균 4.24, 긍정 90.3%), 학생에게 부여한 점수의 범위를 확인하여 학생의 능력을 제대로 변별하고 있는지 확인할 수 있다(평균 3.99, 긍정 81.1%)고 인식하였다. 이는 많은 교사들이 피험자로서의 경험, 교육대학 또는 교사 직무연수 과정에서의 학습 경험, 교사로서의 평가 실시 경험, 매체와 선배 교사의 조언 등을 통한 간접 경험을 바탕으로 평가 실시와 관련된 경험과 기술을 지니게 되었다고 인식한 것으로 볼 수 있다. 다음 면담 사례가 이를 잘 보여준다. 예를 들어, 첫 번째 사례에서는 교육대학에서의 학습 경험,

피험자로서의 경험, 선배 교사의 조언, 두 번째 사례에서는 매체를 통한 간접 경험, 세 번째 사례에서는 교사로서의 평가 경험 관련 내용이 포함되어 있다.

대학교 때도 배웠거든요. 잠깐. 감독하는 사람이 이리저리 움직였을 때, 특정한 학생 옆에서 있었을 때, 그 학생이 시험 치는 데 부담을 느끼고 심리적 압박을 받을 수 있다. 그러니까 그러지 마라. 자꾸 돌아다니지 마라. 자꾸 돌아다녔을 때 애들이 역시나 심리적으로 그릴 수 있다. 그런 과목이었는지, 그냥 그 분이 말씀하시다 나왔는지는 모르겠지만, 들었었구요. 그것 외에는 제가 학생 때 했던 경험으로 애들 그대로. 그 때 선생님들이 어떻게 감독을 하셨던가가 저의 모습을 완성하고 있는 것 같아요. (중략) Tea 타임 때 가면 (선배 교사에게) 제가 여쭙보면 말씀해 주시더라고요. 제가 ‘15점 맞은 애가 자꾸 그래요’라고 그랬더니 ‘맨 앞에 자리를 옮겨라’ 그런 식으로도 얘기를 해주시죠.

텔레비전 보니까 대학학교 같은 경우에는 그게 없다고 그러더라고요. 시험 감독관 없이 자율적으로 하는 시험을 하는데, 아이들한테 전부 다 개네 양심에 맡기는 거죠. 그런데 커닝을 하거나 이런 부정행위를 하는 애들이 오히려 더 없다고 그 얘기를 하더라고요. 그래서 저도 그렇게까지는 못하겠고, 애들을 못 믿겠어 가지고 방법을 쓴 게, 제가 앞에 있고 애들을 보게 되면 제 눈치를 봐요. 은연 중에. 그래서 아예 저는 맨 뒤로 가서 의자를 놓고 앉거든요. 그러면 고개를 돌리는 애들이 없어요. 제가 뒤에서 보고 있다는 생각이 드니까.

과학 같은 경우에 주관식 채점할 때에는 글자가 조금 틀려도 맞다고 하거든요. 왜냐하면 애가 개념은 아는 거니까. 물체를 쓰는데 물체를 이게 아니라 그냥 아이(H)로 쓰면, 개념은 아는 거니까, 국어도 아니고 맞게는 해주거든요. 그런 융통성은 발휘는 하거든요.

‘평가 결과의 분석·해석·활용·의사소통’ 영역의 평균은 3.76점으로 ‘평가 전문성이 있음’에 해당되었으며, 문항별로는 3.46~3.96점으로 점수가 다양하게 분포하였다. 진단평가 결과에 기초하여 학생들의 수준과 요구에 적합한 수업 계획 수립(평균 3.95, 긍정 80.2%), 형성평가 결과에 기초하여 학생들의 학습 곤란 정도와 문제점 진단 및 해결 방안 제공(평균 3.95, 긍정 77.4%), 총합평가 결과에 기초하여 학생의 성취 변화 정도 파악 및 차기 수업 계획을 위한 정보로 활용(평균 3.94, 긍정 77.4%), 학

생, 학부모, 교육 관련자들에게 평가의 목적, 결과, 해석 방법 설명(평균 3.92, 긍정 75.1%), 평가 결과에 기초하여 학생과 학부모에게 학생의 강점과 약점에 대해 설명하고 개선 방향 제시(평균 3.96, 긍정 79.3%) 등에서는 인식이 비교적 긍정적이었다. 이는 학교 차원에서 평가 결과를 분석, 해석, 활용하도록 요구하고 있을 뿐만 아니라, 담임교사제라는 초등학교의 특성상 교사가 학생들을 총체적으로 평가하고 판단할 수 있는 기회가 많기 때문이라고 해석할 수 있다. 다음 면담 내용에서 이를 확인할 수 있는데, 첫 번째 사례는 학교 차원의 요구, 두 번째 사례는 초등학교 교사의 총체적 평가 기회와 관련된 것이다.

학교 업무를 하면서 익히고, 그냥 학교에서 그냥 그렇게 하라고 하니가 그러는 거죠. 어디서 배우고 그런 거 없는 것 같아요. 특별한 연수를 한다거나 그런 건 없고, 학교 차원에서 이런 이렇게 해야 된다. 학교 나름대로의 계획이 있는 거예요. 학교 나름대로의 계획이 있기 때문에 학교 계획에 따라 업무를 하면서 익혀 나가는 거죠. 배우나가는 거죠.

선생님들이 애를 계속 보고 있는 사람이고. 초등 같은 경우에는 담임선생님이니까 애가 지금 당장 과학을 40점을 맞았다 해도, 애는 왜 40점을 맞는지를 선생님은 그 점수만을 보고 판단하지 않잖아요. 가정환경이나 현재 심리상태나 아니면 오늘의 특별한 일이 있었거나 평소 애가 했던 생활태도 그런 걸 다 종합적으로 해서 그렇게 판단하지, 애가 지금 당장 40점을 맞았다고 해서 그 40점만 놓고 보지는 않는 것 같아요.

반면, 학생들의 점수 분포나 문항 특성에 대한 기초 통계 자료 산출(평균 3.46, 긍정 50.9%), 평가 방법에 따라 동일한 학생에 대한 평가 결과가 일치하지 않을 경우 그 원인에 대한 분석 및 해석(평균 3.59, 긍정 60.4%), 학생, 평가 도구, 평가 환경, 평가자 등에 의한 오차를 고려한 평가 결과 해석(평균 3.59, 긍정 59.9%), 표준화 검사에서 보고되는 여러 가지 검사 점수에 대한 정확하고 타당한 해석(평균 3.51, 긍정 52.5%), 검사의 기준에 근거한 학생들의 상대적 위치에 대한 해석 및 설명(평균 3.62, 긍정 54.8%), 검사의 준거에 기초하여 학생들이 무엇을 알고 할 수 있는지에 대한 해석 및 설명(평균 3.65, 56.7%) 등과 같이 정량적 자료 분석 및 해석과 관련된 전문성에 대한 인식은 다른 항목에 비하

여 상대적으로 낮은 편이었다. 이는 최근 정성적인 평가가 상대적으로 강조되고 있는 초등학교 현장의 특성이 반영된 결과라 해석된다. 또한, 정량적인 분석 및 해석과 관련된 학습 기회나 적용 기회가 부족한 것도 원인이 될 수 있다. 다음 면담 사례가 이 주장을 뒷받침해준다. 즉, 첫 번째 사례에서는 초등학교의 경우 구체적인 정량적 평가 및 이에 기초한 피드백을 요구하고 있지 않음을 보여주고 있고, 두 번째 사례에서는 평가 결과에 대한 정량적인 분석 및 해석 기회가 부족함을 보여주고 있다.

초등에서는 점수화해서 내보내지는 않으니까요. 저희는 말로 써가지고 글로 써서 다 내보내니까요. 국어과의 어느 단원을 잘하고 어느 단원이 조금 부족한지. 과학과도 그렇구요. (중략) 왜 애가 이렇게 점수가 낮은지에 대해 성향을 보고서, 선생님은 그 원인을 아니까, 거기다 써 주죠. (중략) 저희는 3단계 평가냐 4단계 평가냐 이런 식으로 해요. 점수가 아니라. 예를 들어 노력 요함, 보통, 잘함, 매우 잘함 이런 식으로 4단계로 하던지.

중간고사나 기말고사 봤을 때는 그렇게 분석해서 대책까지 해가지고 보고하고 이런 거는 없고. 그냥 채점하다가 이 문제 왜 이렇게 많이 틀렸냐. 그건 어쨌든 그 문제의 문제일 수도 있고, 아니면 내가 뭔가를 잘못 가르쳐서 그런 결과가 나올 수도 있긴 하니까, 그런 거에 대해서 다시 한 번 돌아보는 과정은 있지만 보고서로 해가지고 한 적은 없어요. 그런데 진단평가 같은 경우에는, 애가 점수가 너무 낮으면 애가 왜 이렇게 점수가 낮았냐, 그 원인 분석해 가지고 그건 보고를 하는 데, 다른 거는 굳이 시험 본 다음에 그렇게 해 본 적은 없는 것 같아요.

‘평가의 윤리성’ 영역의 경우, 전체 평균은 4.09점으로 ‘평가 전문성이 있음’에 해당되었고, 문항별로는 3.80-4.34점의 점수 분포를 보였다. 즉, 대부분의 교사들이 평가 관련 지침 및 규정 준수(평균 4.05, 긍정 81.6%), 평가 도구(평균 4.28, 긍정 87.1%) 및 평가 결과(평균 4.34, 긍정 89.4%)의 보안과 유지, 합리적이고 공정한 절차에 의한 평가 실시(평균 4.31, 긍정 88.0%) 측면에서 전문성을 갖추고 있다고 인식하는 것으로 나타났다. 이보다는 약간 부족하나 평가 도구에 성별이나 사회경제적 배경에 따라 불리하게 작용할 수 있는 내용이나 방법의 선정 여부 확인(평균 3.80, 긍정 69.9%), 평가 활동에 대한 반성적 평가(평균 3.82, 긍정 71.4%) 측면에서 자신의 전문성을 긍정적으로 인식한 교사도 많은

편이었다. 다음 면담 사례에서 알 수 있듯이, 평가의 윤리성 측면에서 어긋나는 행위를 할 경우 교사 자신에게 불이익이 생긴다는 두려움이 이러한 인식의 주요한 원인으로 보인다. 또한, 학생 개인 정보와 인격의 중요성에 대한 사회적 공감대가 형성되어 있는 것도 한 가지 원인일 수 있다.

관리 지침 규정 이런 거는, 안 지키면 저희가 불이익이 있으니까 지키 수밖에 없구요. 이거는요. 하다못해 작년 거 평가지를 그대로 가지고 올해 썼다가, 그게 걸려 가지고 문제가 된 경우도 있었고. 또 이 학교에서 쓰던 평가지를 다른 학교에 가져와서 썼다가도 문제된 경우도 있기 때문에, 이런 거는 저희가 안 지키 수가 없거든요.

2. 배경 변인에 따른 분석

1) 교직 경력에 따른 분석

교직 경력에 따른 일원변량분석 결과를 Table 3에 제시하였다. 모든 하위 영역에서 교직 경력이 증가할수록 평균이 증가하였으며, ‘평가의 윤리성’ 영역을 제외한 4개 하위 영역에서 주효과가 유의미하였다($p < 0.05$). Scheffé 사후검증 결과, 4개 하위 영역에서 모두 교직 경력이 5년 미만인 교사와 20년

이상인 교사의 평균 차이만 통계적으로 유의미하였다. 이러한 결과는 교직 경력이 증가하면서 평가 방법의 선정, 평가 도구의 개발, 평가 실시·채점·성적 부여, 평가 결과의 분석·해석·활용·의사소통 능력이 자연스럽게 터득될 수 있으나, 적어도 20년 이상이 지난 후에야 의미 있는 전문성 향상이 이루어질 수 있음을 시사한다. 교사의 평가 전문성 향상은 성공적인 교수·학습 및 교사 전문성 향상을 위한 핵심적인 요소이므로(AFT *et al.*, 1990; Black & Wiliam, 1998; Cho & Ko, 2008; Kim, 2002; Kwon, 2012; Park & Oliver, 2008; Veal & MaKinster, 1999), 보다 단기간에 초등학교 교사들의 과학 평가 전문성을 향상시킬 수 있는 방안을 모색할 필요가 있다. 특히, ‘평가 도구의 개발’ 영역의 경우에는 20년 이상의 교직 경력을 지닌 교사들도 평균이 3.70점 미만 수준에 머물러 있었으므로, 이 능력을 효과적으로 향상시킬 수 있는 방안 마련이 절실히 요구된다.

한편, ‘평가의 윤리성’ 영역의 경우에는 교직 경력에 따라 유의미한 변화가 없었는데, 이는 평가 윤리 위반 시 교사 자신에게 생기는 불이익이나 학생 개인 정보와 인격의 중요성에 대해 인식하는 데에는 오랜 교직 경력이 요구되지 않음을 의미한다.

Table 3. One-way ANOVA results for teachers' self-diagnosis on their competency for assessment in science by teaching experience

Category	Less than 5 years (n=58)		5 to 10 years (n=45)		10 to 20 years (n=57)		More than 20 years (n=57)		MS	F	p
	M	SD	M	SD	M	SD	M	SD			
Choosing assessment methods	3.67	0.54	3.82	0.44	3.83	0.53	3.94	0.40	0.72	3.07	0.029
Developing assessment instruments	3.30	0.64	3.50	0.65	3.53	0.69	3.67	0.56	1.36	3.36	0.020
Administering, scoring, and grading	3.74	0.53	3.84	0.42	3.97	0.56	4.13	0.49	1.59	6.16	0.000
Analyzing, interpreting, using, and communicating assessment results	3.60	0.56	3.74	0.41	3.77	0.56	3.93	0.49	1.08	4.13	0.007
Recognizing ethics in assessment	3.98	0.60	4.02	0.57	4.12	0.65	4.24	0.45	0.71	2.17	0.092

Table 4. Independent t-test results for teachers' self-diagnosis on their competency for assessment in science by major

Category	Science (n=24)		Non-science (n=132)		t	p
	M	SD	M	SD		
Choosing assessment methods	3.69	0.54	3.84	0.47	1.37	0.173
Developing assessment instruments	3.27	0.68	3.51	0.65	1.71	0.089
Administering, scoring, and grading	3.83	0.50	3.94	0.53	0.89	0.377
Analyzing, interpreting, using, and communicating assessment results	3.58	0.54	3.79	0.50	1.91	0.058
Recognizing ethics in assessment	3.87	0.80	4.12	0.54	1.93	0.056

2) 전공에 따른 분석

전공에 따른 독립표본 *t*-검증 결과(Table 4), 모든 하위 영역에서 과학 전공과 비전공의 평균 차이가 통계적으로 유의미하지 않았다($p>0.05$). 이러한 결과는 교육대학에서의 과학 심화전공 과정이 교사의 과학 평가 전문성에 의미 있는 영향을 미치는 못함을 의미한다. 즉, 과학 심화전공 과정에서 체계적이고 실질적인 평가 전문성 교육이 충분히 이루어지고 있지 못한 현실을 반영한 결과로 볼 수 있다.

3) 학력에 따른 분석

학력에 따른 독립표본 *t*-검증 결과는 Table 5와 같다. ‘평가 방법의 선정’ 및 ‘평가의 윤리성’ 영역에서는 석사학위 취득 교사와 미취득 교사의 평균 차이가 통계적으로 유의미하지 않았으나($p>0.05$), 나머지 3개 하위 영역에서는 석사학위 취득 교사의 평균이 통계적으로 유의미하게 높았다($p<0.05$). 이러한 결과는 대학원 과정이 평가 방법의 선정 및 평가의 윤리성 측면에 미친 영향은 미미한 반면, 평가 도구의 개발과 평가 실시·채점·성적 부여 및 평가 결과의 분석·해석·활용·의사소통 측면에 긍정적인 영향을 미침을 의미한다. ‘평가 방법의

선정’ 영역의 경우 이론적인 측면이 강하므로 교육 대학 교육과정, 임용시험 준비 과정, 교직 경험 등을 통해서도 관련 지식을 상당 부분 습득할 수 있었기 때문에 해석된다. ‘평가의 윤리성’ 영역의 경우에도 평가 윤리 위반으로 인한 불이익이나 학생 개인 정보와 인격의 중요성에 대한 인식에는 추가적인 교육이 불필요하기 때문인 것으로 보인다. 반면, 나머지 3개 하위 영역에서의 결과는 이 영역들이 비교적 실천적이고 실제적인 측면이 강한 특성을 지닌 것과 관련이 있다. 즉, 비록 과학 분야는 아니더라도 대학원 과정에서 평가 도구의 개발, 평가 실시·채점·성적 부여 및 평가 결과의 분석·해석·활용·의사소통과 관련된 이론적 내용을 학습하거나, 다른 교사나 교수와 관련 내용에 대하여 논의할 수 있는 기회가 있었기 때문에 관련 전문성이 향상된 것으로 보인다.

4) 연수 이수 여부에 따른 분석

연수 이수 여부에 따른 독립표본 *t*-검증 결과는 Table 6과 같다. ‘평가의 윤리성’ 영역을 제외한 4개 하위 영역에서 모두 평가 전문성 관련 연수 이수자의 평균이 미이수자의 평균보다 통계적으로 유의

Table 5. Independent *t*-test results for teachers’ self-diagnosis on their competency for assessment in science by academic degree

Category	Bachelor (n=173)		Master ^a (n=30)		<i>t</i>	<i>p</i>
	M	SD	M	SD		
Choosing assessment methods	3.80	0.50	3.90	0.42	1.01	0.315
Developing assessment instruments	3.44	0.64	3.76	0.62	2.50	0.013
Administering, scoring, and grading	3.88	0.53	4.09	0.43	2.08	0.039
Analyzing, interpreting, using, and communicating assessment results	3.72	0.54	3.95	0.41	2.27	0.024
Recognizing ethics in assessment	4.07	0.62	4.17	0.32	1.29	0.200

^a All non-science

Table 6. Independent *t*-test results for teachers’ self-diagnosis on their competency for assessment in science by completion of training

Category	Yes (n=115)		No (n=101)		<i>t</i>	<i>p</i>
	M	SD	M	SD		
Choosing assessment methods	3.88	0.53	3.74	0.43	2.23	0.027
Developing assessment instruments	3.66	0.64	3.33	0.61	3.86	0.000
Administering, scoring, and grading	4.00	0.52	3.84	0.53	2.35	0.020
Analyzing, interpreting, using, and communicating assessment results	3.85	0.53	3.66	0.50	2.75	0.006
Recognizing ethics in assessment	4.11	0.58	4.07	0.58	0.44	0.661

미하게 높았다($p < 0.05$). 이러한 결과는 평가 전문성 관련 연수 과정에서 평가 방법의 선정, 평가 도구의 개발, 평가 실시 · 채점 · 성적 부여 및 평가 결과의 분석 · 해석 · 활용 · 의사소통과 관련된 이론적이고 실천적인 학습과 적용 및 논의 기회가 주어졌기 때문일 수 있다. 그러나 연수 이수 후에도 여전히 ‘평가 도구의 개발’ 영역의 평균이 3.70점 미만 수준에 머물러 있었으므로, 이 능력을 효과적으로 향상시킬 수 있는 연수 방안을 모색할 필요가 있다.

한편, ‘평가의 윤리성’ 영역의 경우에는 연수 이수의 영향이 미미하였는데, 이는 교직 경력 및 학력에 따른 결과와 동일한 맥락에서 이해할 수 있을 것이다. 즉, 연수 이수 여부와 관계없이 다른 교육 및 학습 경험을 통해서도 관련 전문성을 지니는 데 별 문제가 없기 때문일 가능성이 있다.

IV. 결론 및 제언

이 연구에서는 초등학교 교사의 과학 평가 전문성에 대한 자기진단 내용을 분석하였다. 연구 결과, 초등학교 교사들은 ‘평가의 윤리성’, ‘평가 실시 · 채점 · 성적 부여’, ‘평가 방법의 선정’, ‘평가 결과의 분석 · 해석 · 활용 · 의사소통’, ‘평가 도구의 개발’ 영역 순으로 자신의 전문성 수준에 대하여 높게 평가하는 것으로 나타났다. 교직 경력이 증가할수록 모든 하위 영역의 전문성 수준에 대하여 높게 평가하였으며, 특히 교직 경력이 20년 이상인 교사가 5년 미만인 교사보다 ‘평가의 윤리성’ 영역을 제외한 모든 하위 영역의 전문성 수준에 대하여 유의미하게 높게 평가하였다. 과학 전공 여부에 따라서는 모든 하위 영역의 전문성에 대한 자기진단 내용에 유의미한 차이가 없었다. 그러나 석사학위 취득 교사가 미취득 교사보다 ‘평가 도구의 개발’, ‘평가 실시 · 채점 · 성적 부여’, ‘평가 결과의 분석 · 해석 · 활용 · 의사소통’ 영역에서만 자신의 전문성을 더 높게 평가하였다. 또한, 평가 전문성 관련 연수 이수 여부에 따라서는 이수 교사가 미이수 교사보다 ‘평가의 윤리성’ 영역을 제외한 모든 하위 영역에서 자신의 전문성 수준을 더 높게 평가하였다.

지금까지는 초등학교 교사의 과학 평가 전문성 수준을 총체적이고 체계적으로 조사한 경우가 없어, 관련 정보가 매우 부족한 측면이 있었다. 이런

점에서 볼 때 이 연구의 결과들은 초등학교 교사들의 과학 평가 전문성 수준을 체계적으로 파악하고 부족한 전문성의 향상 방안을 모색하는 데 유용한 시사점을 제공할 수 있다. 즉, 많은 초등학교 교사들이 과학 평가의 윤리성, 평가 실시 · 채점 · 성적 부여, 평가 방법의 선정, 평가 결과의 분석 · 해석 · 활용 · 의사소통 측면에서는 전반적으로 전문성이 있을 가능성이 있다. 반면, 교수 · 학습과정에서 학생이 겪는 어려움을 진단하거나 특별한 지도를 필요로 하는 학생을 변별하는 데 적합한 평가 방법의 선정 및 기준 참조 평가와 준거 참조 평가에 관한 전문성이 부족할 가능성을 확인하였다. 또한, 정량적 자료 분석과 해석 및 평가 도구의 개발과 관련된 전문성은 상대적으로 부족할 가능성을 확인할 수 있었다.

교사들의 과학 평가 전문성을 향상시키기 위해서는 여러 가지 방안이 요구되는데, 대학원 과정이나 교사 직무연수 과정에서의 실천적인 학습 및 적용 경험이 관련 전문성 향상에 기여하는 것으로 나타났다. 따라서, 해당 과정을 내실화하는 것이 가장 현실적인 방안이 될 수 있다. 특히 평가 도구의 개발과 관련된 전문성은 현재의 해당 과정만으로는 기대만큼의 질적 향상이 이루어지지 않을 가능성이 있으므로, 이 부분에 대한 교육을 보다 강화할 필요가 있다. 이를 위해, 초등학교 교사들에게 해당 전문성에 대한 이론적인 학습 기회를 제공하는 것뿐만 아니라 교사들이 개별 또는 공동으로 다양한 상황에 적합한 평가 도구를 직접 개발한 후 그 과정과 결과물에 대하여 다른 교사나 전문가들의 피드백을 받아 개선하는 경험을 제공하는 방안이 효과적일 수 있다. 개발한 평가 도구를 활용하여 평가를 실시한 후, 그 평가 과정이나 결과물에 대하여 적절한 피드백을 제공한다면 그 효과는 더욱 커질 수 있다. 따라서 다양한 형태의 연수나 교사 모임, 멘토링 등을 통하여 이러한 과정을 효율적으로 진행하는 방안을 마련하기 위해 노력해야 할 것이다. 한편, 교사들이 평가 도구를 실제로 개발하는 능력이 부족하다라도 기존에 개발된 평가 도구를 선정하여 적용할 수 있는 능력은 어느 정도 갖추고 있다고 판단되므로, 우수한 평가 도구를 효과적으로 개발하여 보급하는 방안을 마련하는 일도 필요하다. 실제로 현재 여러 기관에서 평가 도구와 연수 자료들을 개발하여 보급하고 있으므로, 이 자료

들의 수준과 현장 활용성을 제고하기 위한 노력이 요구된다.

한편, 이 연구에서는 교사들의 과학 평가 전문성 수준을 설문과 면담을 통하여 조사하였는데, 교사의 인식과 실제 전문성 수준에 차이가 있을 수 있다. 따라서 인식 조사 수준을 넘어 배경 변인에 따른 교사들의 실제 과학 평가 과정 및 수행 수준을 심층적으로 조사하는 질적 연구를 진행할 필요가 있다.

참고문헌

- American Federation of Teacher, National Council on Measurement in Education, & National Education Association (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Cho, H.-H. & Ko, Y. J. (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.
- Choi, K., Lee, H. & Nam, J.-H. (2000). The effects of formative assessment with detailed feedback on students' science achievement, attitude, and interaction between teacher and students. *Journal of the Korean Association for Science Education*, 20(3), 479-490.
- 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.
- 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), 227-239.
- Kim, D.-Y. (2006). The development of a self-diagnostic program for improving science teachers' competency in student assessment - Centering on earth science contents -. Doctoral dissertation, Seoul National University, Seoul.
- Kim, H.-J., Kwack, D.-H. & 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, J. Y. (2005). Elementary teachers' understanding and practice on science performance assessment. Master's thesis, Daegu National University of Education, Daegu.
- 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), 151-171.
- Kim, S.-Y. & 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), 213-232.
- 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, Y.-G., Ko, M.-S. & Kim, M.-Y. (2011). A study of physical education teachers' specialty in evaluation. *Journal of Korean Society for the Study of Physical Education*, 16(3), 65-75.
- McMillan, J. H. (2004). Classroom assessment principle and practice for effective instruction. Boston: Allyn & Bacon.
- 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.
- Ministry of Education, Science and Technology (2011). Science education curriculum [과학과 교육과정]. Ministry of Education, Science and Technology, 2011-361.
- 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., 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, 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.
- Veal, W. R. & MaKinster, J. G. (1999). Pedagogical content knowledge taxonomies. *Electronic Journal of Science Education*, 3(4). Retrieved from <http://unr.edu/homepage/crowther/ejse/ejsev3n4.html>