



과학 교육에서의 평가 연구 동향

정수임, 신동희*

이화여자대학교

Trends of Assessment Research in Science Education

Sue-Im Chung, Dong-Hee Shin*

Ewha Womans University

ARTICLE INFO

Article history:

Received 13 July 2016

Received in revised form

30 July 2016

7 August 2016

Accepted 7 August 2016

Keywords:

assessment related to science education, research trend, evaluation, measurement, cognitive domain, affective domain

ABSTRACT

This study seeks educational implication by analyzing research papers dealing with science assessment in the most recent 30 years in Korea. The main purpose of the study is to analyze the trends in published papers on science assessment, their purpose, methodology, and key words, especially concentrating on the cognitive and affective domains. We selected 273 research articles and categorized them by research object, subject, methodology, and contents. To examine the factors that affect the research trend, we also tried to contextualize papers' theme in terms of changes in national curriculum and assessment system during the contemporary period. As a result, an overall research trend reflects changes in science curriculum and assessment events such as implementation of college scholastic ability test or performance assessment. There is an unequal distribution in various aspects of the researches, showing a superiority in cognitive domains than the affective ones. By using standardized data obtained through the national and international assessment of educational achievement in science, quantitative researches were superior to qualitative ones. Studies on cognitive domain use variously written- and performance-based tests, whereas most studies of the affective ones prefer written tests. Applied research and evaluation research are predominant comparing to basic ones, which most of the research methodology is based on statistics. Lastly, we found out that key words and subjects tend to be subdivided and detailed rather than general and comprehensive, as time goes on. Such trend will be helpful to elaborate and refine assessment tools that have been regarded as a problem.

1. 서론

과학 교육 과정의 변화는 평가 내용 뿐 아니라 형식의 변화와 밀접하게 연관된다. Orpwood(2001)는 과학을 사회적 맥락에 폭 넓게 연계시키려는 과학 교육의 새로운 목표가 최근 강조되고 있지만 평가는 아직까지 이를 따라잡지 못하고 있음을 지적하며, 평가에 대한 새로운 접근을 개발하고 실시하지 않으면 교육 과정의 개혁 취지가 무색해질 수 있음을 경고했다. 평가는 교육 과정을 이루는 요소인 동시에 교육 과정과 복합적으로 상호작용한다(Cowie, 2012). 그 과정에서 평가는 교육 과정이 목표하는 핵심을 드러내어 입증하는 한편, 이를 한층 더 발전시키는 역할을 한다(Stufflebeam *et al.*, 1971). 평가의 과정과 결과는 작게는 개인의 성취와 향상에서부터, 수업 과정과 교육 프로그램은 물론 큰 틀의 교육 과정 방향을 결정하는 근거가 될 수 있다는 점에서 과학 교육자와 정책 수립자의 관심 대상이었다. 과학 교육 과정은 수많은 변형을 겪어 왔는데, 그 변형은 내용 지식에 초점을 맞출 뿐 아니라, 개념 구조와 인지적 과정, 과학 지식을 평가하고 발전시키는 인식론적 틀, 지식이 소통되어 형태를 잡아가는 사회적 과정과 맥락에 이르기까지 변화의 목표가 확장되어 왔다(Duschl, 2008; Cowie, 2012). 계획된 교육 과정은 실행에서 그치는 것이 아니라 달성된 교육 과정이어야 한다는 점에서 변화된 교육 과정과 평가

간의 상호 작용이 중요해졌고, 그 역동성 안에서 평가의 두 가지 관점, 즉 과학 교육 과정 및 프로그램 평가와 과학 학습 평가가 상호 보완적으로 관련을 맺는다. 과학 학습 평가는 수업 평가, 교실 수준의 평가 등으로 표현되며 학생들의 성취, 교사의 수업 행위, 교수 방법의 효율성을 평가한다는 측면에서 개인적 성취 뿐 아니라 학습을 돕고 프로그램을 평가하는 교육적 평가로서의 의미를 지닌다(Kim *et al.*, 2014; Pellegrino, 2012).

정기적으로 과학 교육의 동향을 정리한 학술지들은 평가 연구를 크게 과학 교육 프로그램 평가와 교실 수준 평가로 나누어 보고하는 동시에, 인지적 혹은 정의적 영역에서의 연구 경향과 대상과 규모를 다양화한 평가 연구를 보완하면서 연구에 대한 시사점을 제시했다(Fraser, Tobin, & McRobbie, 2012; Lederman & Abell, 2014). Lawrenz & Thao(2014)는 과학 교육 평가와 과학 교육 동향이 밀접하게 연계되기 때문에 과학 교육 동향을 잘 파악할 것을 권고하고 있는데, 이들이 제시한 과학 교육의 몇 가지 방향성은 STEM(Science, Technology, Engineering, and Mathematics)의 강조, 과학 교육 결과를 매개하는 변수로서 대상의 다양화, 형식적 환경 밖에서 일어나고 평생 지속되는 학습으로서 과학 학습에 대한 인식, 노동 인력의 개발 뿐 아니라 높은 수준과 교과 특이적 상황에서의 학습, 개별적이고 자발적 학습 동기에 의한 학습, 세계화의 경향 등이다. 이에 비교되는

* 교신저자: 신동희 (donghee@ewha.ac.kr)
<http://dx.doi.org/10.14697/jkase.2016.36.4.0563>

평가에서의 경향성은 문화적 인식의 성장, 특정 분야의 전문화와 직업화, 능동적 정치 활동으로 정리했고, 교육의 책무성에 대한 관심의 증가와 평가의 관계를 함께 언급했다(Lawrenz & Thao, 2014).

Lyon(2011)은 교실 수준의 평가 연구를 접근 방법에 따라 평가 중심, 교사 중심, 학생 중심의 세 가지로 분류했다. 평가 중심 접근은 평가 도구의 타당도와 신뢰도와 같이 평가의 기술적 질과 전통적 선택형 검사와 대안적 평가 전략의 공정성에 초점을 맞추고, 교사 중심 접근은 형성적 평가와 같이 교수와 학습을 향상시키기 위한 과정(Cowie, 2012)에 관심이 있으며, 학생 중심 접근은 학생의 관점과 대안적 평가 전략으로부터 비롯되는 성취와 동기를 검토한다. 이상의 논의에서 살펴본 바와 같이 평가 연구는 평가하는 목표(성취의 측정, 학습 향상, 프로그램 평가)와 대상(학생, 수업, 프로그램, 교육과정)과 상황(교실, 대규모)에 따라 다양한 수준으로 수행되었다.

교육 평가 영역은 응용 과학의 속성을 지니기(Lawrenz & Thao, 2014) 때문에 평가를 이루는 요소와 과정에 대한 이론적 토대와 경험적 증거가 함께 뒷받침되지 않으면 그 결과를 해석하여 새로운 국면을 예상하기 어렵다. 평가는 밖으로 보이지 않는 정신적 표현과 그 과정을 담은 교육적 산출물을 측정하므로 길이나 무게를 재는 것처럼 단순하지 않다(Pellegrino, 2012). 길이나 무게는 측정값이라는 결과와 형태가 측정을 대표하지만, 평가는 ‘무엇을 측정하는지’의 범위를 분명히 확정할 수 없을 뿐 아니라 관찰된 것이 우리가 보려는 ‘무엇’ 인지를 판단해야 하는 연속적 과정이다. NRC(2001)는 평가의 개념을 ‘증거로부터 추론하는 과정’으로 설정했는데, 이러한 개념화는 평가의 목적과 대상, 원리와 방법을 이끌어 가는 역할을 한다. Pellegrino(2012)는 학생들이 알고 있는 것을 추론할 수 있도록 증거를 수집하는 과정으로서 평가를 개념화하면서 인지, 관찰, 해석이 서로 연결되는 삼각형 모형을 제안했고, 이 모형은 교실 수준의 평가 틀로 다양하게 응용되었을 뿐 아니라 표준화된 측정 도구 개발의 원리로 활용되었다(Lyon, 2011; Liu, 2012). 한편, 평가 개념은 평가를 보는 철학적 입장에 따라 달라지기 때문에 이에 따라 평가를 구성하는 요소에 대한 다양한 정의가 구현될 수 있으며 평가 모형이 구체화된다. Lawrenz & Thao(2014)는 객관주의와 주관주의, 공리주의와 다원주의적 가치 사이를 오고가면서 취하는 평가의 입장과 함께, Stufflebeam & Shinkfield(2007)의 다섯 가지 프로그램 평가 유형인 의사평가, 질문 및 방법 지향, 개선 및 책무성 중심, 사회적 의제와 옹호, 절충 모형 등으로 40년 동안 연구된 평가 모형을 분류했다.

현재의 과학 교육 평가 현황과 함께 과거의 연장선상에서 평가의 발전 단계를 살펴보는 동향 연구는 세 가지 측면에서 과학 교육 연구자들이 주목해야 할 필요가 있다. 첫째, 과거로부터 현재까지 진행된 연구들을 수직적, 수평적 차원에서 조망하면서 미래의 경향을 예측할 수 있다. 둘째, 과학 교육의 전체 흐름 안에서 연구의 위상을 확인함으로써 연구자의 입장과 관점을 공고히 할 수 있는데, 이는 연구에서

사용하는 개념을 명확히 하고 연구가 나아갈 방향에서 벗어나지 않도록 안내한다. 셋째, 많은 연구의 분야가 구체적으로 전문화된 상황에서 연구자 스스로 연구의 의미를 성찰하는 기회를 제공함으로써 전체적인 맥락에서 의미를 연계하고 보다 생산적 관점을 제시할 수 있다.

국외의 과학 교육 평가에 대한 동향 연구는 이상의 논의와 같이 학술지를 중심으로 정기적으로 정리가 되어온 반면, 국내 과학 교육 평가 연구의 동향은 이론적 배경을 이루는 문헌 연구가 개별 연구에 포함되는 형식으로 축소된 경향이 있으며 과학 교육 관련 평가에 대하여 전체적으로 종합한 연구를 찾아보기 어렵다. 특히 우리나라가 국제 학업 성취도 평가에 참여하게 되면서 인지적, 정의적 영역의 성취를 국제적으로 비교할 수 있게 된 이후로, 인지적 성취에 비해 현저히 낮은 정의적 성취(Choe et al., 2013)를 우려하는 학계의 반응이 나타났다. 이는 그 동안 진행되어 왔던 과학 교육과정의 목표와 평가를 포함한 수행 과정을 특별히 정의적 차원에서 점검해 보아야 할 필요성을 드러냈다. 따라서 본 연구에서는 과거 30년간 국내에서 수행된 과학 교육에서 평가 연구의 동향을 인지적, 정의적 영역을 중심으로 파악하고 그 결과를 근거로 과학 교육 평가와 교육 과정 개선 방향의 관점을 제시하는 한편, 향후 연구에 대한 시사점을 얻고자 했다. 이러한 목적에 따라 과거 30년간(1986년~2015년) 과학 교육에서 평가 관련 연구의 전체적 동향, 인지적 영역과 정의적 영역의 평가 연구 동향, 연구 방법론과 내용에 대한 평가 연구 동향, 과학 교육 평가 연구에서 다루는 주제의 동향 등을 알아보려고 한다.

II. 연구 방법

1. 분석 대상

국내 과학 교육에서의 평가 관련 연구에 초점을 맞추어, 1986년부터 2015년까지 30년 동안 국내 학술지에 게재된 논문 중 과학 교육 평가 관련 논문을 분석했다. 한국교육학술정보원이 제공하는 데이터베이스(RISS)를 통해 ‘평가 and 과학 교육’, ‘평가 and 물리 교육’, ‘평가 and 화학 교육’, ‘평가 and 생물 교육’, ‘평가 and 지구과학 교육’, ‘측정 and 과학 교육’, ‘측정 and 물리 교육’, ‘측정 and 화학 교육’, ‘측정 and 생물교육’, ‘측정 and 지구과학교육’ 등의 키워드를 검색하여 나타난 7,564건 중 관련 없거나 중복되거나, 교육 평가가 연구 목표가 아닌 학습 모형 및 프로그램 효과 평가, 논문이 아닌 학술 발표 초록 등을 제외했다. 이 중, KCI(Korea Citation Index) 등재지와 KCI 등재 후보지에 게재된 총 273편의 논문을 최종 분석 대상으로 선정했다(Table 1). 선정 논문 중에서 인지적 영역을 다룬 논문이 57.1%로 가장 많았고, 인지적 영역과 정의적 영역을 혼합한 경우가 20.5%, 정의적 영역이 12.8%, 영역을 특정할 수 없는 기타의 경우는 9.5%로 분석되었다. KCI 등재지와 등재 후보지 논문 비율은 각각 91.2%, 8.8%였다.

Table 1. Research papers analyzed

	논문 수 (%) (N=273)				계
	인지적	정의적	혼합	기타	
KCI 등재지	136(49.8)	35(12.8)	55(20.1)	23(8.4)	249(91.2)
KCI 등재 후보지	20(7.3)	-	1(0.4)	3(1.1)	24(8.8)
계	156(57.1)	35(12.8)	56(20.5)	26(9.5)	273(100.0)

2. 분석 영역 선정

본 연구는 과학 교육에서 평가를 다룬 논문을 분석하여 동향을 파악함으로써 과학 교육 과정 개발, 적용, 개선 등의 방향을 제시하는 한편, 향후 연구에 대한 시사점을 얻고자 했다. 특히 TIMSS(Trends in International Mathematics and Science Study)와 PISA(Programme for International Student Assessment)와 같은 국제 학업 성취도 평가에서 우리나라 학생들이 보여준 높은 학업 성취, 낮은 정의적 성취 간의 격차(Choe *et al.*, 2013)는 그 동안의 과학 교육 평가 연구에서 어떤 결과가 도출되어 왔는지 살펴보아야 할 필요성을 드러냈다. 이에 연구가 다루고 있는 내용과 인적 대상 이외에도 인지적 영역과 정의적 영역을 중심으로 연구 방법론(Sung & Si, 2015; Fraenkel, Wallen, & Hyun, 2015; Gall, Gall, & Borg, 2007; Miller, 1991)을 세 가지 관점, 즉 인식론적 방법, 연구 목적, 연구 방법 등으로 분석하고, 연구 내용 측면에서는 사용한 평가 도구와 수집 자료를 분석했다(Table 2).

연구 대상의 범주는 초등, 중등 학생 뿐 아니라 유치원생과 특수 교육 대상자, 예비 교사, 영재, 교사와 학부모를 포함하여 분석했다. 유아와 특수 교육 대상자를 대상으로 한 연구는 기존의 초등, 중등 과학 교육 연구에서 제외되는 경우가 많았지만, 유아 과학 교육은 초등 과학 교육과 연장선상에 있다는 점, 특수 교육 대상자의 과학 교육 가능성은 다양한 인지적, 정의적 상황을 지닌 학생 변인으로 확장될 수 있다는 점에 의의가 있다는 점에서 본 연구에서는 분석 대상으로 포함시켰다. 교과 범주는 학교 교육 과정에 따라 세부 전공이 드러난 경우에는 물리, 화학, 생물, 지구과학, 환경 등으로 분류했고, 초등학교나 중학교에서 다루는 일반 과학 내용은 과학으로 분류했다.

복잡하게 얽혀있는 교육 현상을 이해하기 위해서는 맥락에 맞는 다양한 연구 방법론을 사용해야 하는데(Eisner, 1997; Sung & Si, 2015), 이때 연구자들이 사용한 연구 방법은 현장에서 발견한 연구 문제의 해답을 찾아가는 해법이며 접근 방법에 해당한다. 연구 문제에 적합한 연구 방법을 사용하지 않으면 수집되는 자료의 신뢰도와 연구 내용의 타당도를 평가하는 방법이 달라질 수 있다는 점(Kim, 2012), 연구자가 연구 문제에 접근하는 최선이라고 생각하는 방법을

분석하면 향후 연구 방법에 대한 통찰과 성찰을 지닐 수 있다는 점(Kim, Kim, & Kwon, 2012), 과학 교육 평가와 관련하여 당면한 문제와 연구자와 현장의 관심을 연구 방법론이라는 도구를 통해 파악할 수 있다는 점 등에서 연구 방법론 분석의 의의가 있다. 연구 방법론은 첫째, 인식론적 접근 방법에 따라 실증주의와 후기 실증주의에 입각한 양적 연구와 질적 연구(Sung & Si, 2015; Neuman, 2011; Chae, 1996) 외에 그 둘을 통합한 통합 연구(Sung & Si, 2015; Creswell, 2008)와 기타로 분류했다. 두 번째는 연구 목적에 따라 기초 연구, 응용 연구, 실행(현장) 연구(action research), 평가 연구(Sung & Si, 2015; Miller, 1991) 등으로 분류했다. 기초 연구는 원리나 사실을 발견하거나 이론을 발전시켜 지식을 확장시키는 역할을 하고, 응용 연구는 실제 문제 상황에서 나타나는 이론적 개념의 검토, 상황의 진행과 결과를 개선하는 목적을 지닌다. 실행(현장) 연구는 특정한 학교나 학급 단위에서 일어나는 문제에 대한 해답을 얻고 실제 교육 수준을 향상시키는 데 목적이 있으며(Johnson, 2005; Fraenkel, Wallen, & Hyun, 2015), 평가 연구는 프로그램이나 정책의 효과를 분석하여 목표에 도달한 정도를 측정하는 연구다. 세 번째 연구 방법론의 분류 기준은 연구의 방법에 따라 여섯 개의 범주, 즉 문헌 연구, 기술 연구·조사 연구, 상관 연구·집단 비교 연구, 실험 연구, 개발 연구, 관찰 연구 등으로 나누었다. Chae(1996)는 연구 방법론을 분류할 때 나타나는 개념상의 혼란과 불일치성이 연구 방법론의 ‘다차원성’을 무시하고 단일한 차원으로 분류하는 것에서 비롯된다고 했다. 본 연구에서 연구 방법에 따라 나는 여섯 개의 범주도 연구자에 따라서 서로 다른 차원으로 분류되기도 했다. Fraenkel, Wallen, & Hyun(2015)은 이 연구와 유사한 실험 연구, 상관 연구, 조사 연구, 인과 비교 연구 외에 본 연구에서 분석한 선행 연구에서 나타나지 않는 문화 기술지 연구와 역사 연구를 포함했는가 하면, 전혀 다른 분류 기준으로 범주화한 실행 연구와 평가 연구를 같은 차원으로 제시했고, Sung & Si(2015)는 연구 방법에 따라 실험 연구, 조사 연구, 관찰 연구, 문화 기술지 연구와 역사 연구로 분류했다. 이처럼 연구자에 따라 기준과 범주가 다름에 주목하여 본 연구에서는 여러 연구자의 기준(Sung & Si, 2015; Fraenkel, Wallen, & Hyun, 2015; Gall, Gall, & Borg, 2007)을 참고하고 분석 대상 논문에서 실제 사용한 연구 방법으로부터 여섯 가지 연구 방법을 도출했다.

Table 2. Analysis Criteria

범주		세부 범주			
기본 자료		1. 출판 연도	2. 연구자	3. 학술지명(발행처)	4. 해당 국가 교육 과정 단계
대상		유아, 초등학생, 중학생, 고등학생, 대학생(예비 교사), 학부모, 교사, 영재 학생, 특수 교육 대상자, 특정 대상자 없음			
교과		과학, 물리, 화학, 생물, 지구과학, 환경, 기타			
연구 방법론	인식론적 접근 (자료 수집 방법)	1. 양적 연구	2. 질적 연구	3. 통합 연구	4. 기타
	연구 목적	1. 기초 연구	2. 응용 연구	3. 실행 연구	4. 평가 연구
	연구 방법	1. 문헌 연구 4. 실험 연구	2. 기술 연구·조사 연구	5. 개발 연구	3. 상관 연구·집단 비교 연구 6. 관찰 연구
영역		1. 인지적: 학업 성취도, 탐구 능력, 지식(개념), 사고 3. 혼합: 인지적 영역과 정의적 영역을 함께 다룸		2. 정의적: 흥미, 태도, 동기, 자아개념, 기대, 불안, 정서 4. 기타	
연구 내용	평가 도구	1. 지필형: 선택형, 서답형, 리커트 2. 수행형: 평정형, 산출물(포트폴리오, 보고서, 과제 등), 자기·상호 평가, 구술·면담, 관찰 3. 기타: 공학적 도구(컴퓨터 프로그램 등), 신경 심리학적 도구(뇌 영상 기법, 시선 추적 등)			
	수집 자료	설문지, 면담, (대상자가 작성한) 산출물 등, (연구자의) 관찰 기록, 문헌 자료			

연구 내용의 경우, 인지적, 정의적, 두 영역이 혼합된 영역, 두 영역 중 어느 쪽에도 속하지 않는 기타 영역 등의 네 영역으로 분류했다. Bloom(1976)이 인간 특성을 인지적, 정의적, 심동적 영역으로 분류하여 적용한 이래, 인지적 영역은 대체로 지식의 사용 및 사고방식과 관련한 특성으로 지능, 창의력, 사고력, 지식, 인지 전략과 양식, 지혜(Lee & Kim, 2013) 등을 예로 들 수 있다. 본 연구에서는 개념과 지식, 탐구 능력과 다양한 사고력 외에도 STS 관련 내용, 과학의 본성 등을 인지적 영역으로 포함시켰다. 정의적 영역은 느낌과 감정, 정서에 관련된 특성으로 태도, 동기, 자아 개념, 흥미, 기대, 불안, 귀인 등을 들 수 있는데, 이 연구에서는 사회적 특성과 관련이 있는 협동, 지도력 혹은 여러 영역이 복합적으로 나타나는 인성, 가치관, 몰입(Lee & Kim, 2013) 등도 정의적 영역에 포함시켰다. 그 밖에 두 영역이 분명하게 드러나지 않는 인식 조사나 실태 조사 등은 기타로 범주화했다. 평가 도구는 연구에서 직접 사용했거나, 활용을 염두에 두고 개발된 도구의 유형, 도구를 활용한 결과물과 관련되어 있는 유형을 범주화하여 크게 지필형, 수행형, 기타로 분류하고, 지필형은 다시 선택형과 서답형(Kim *et al.*, 2014), 리커트형으로, 수행형은 평정형, 산출물, 자기·상호 평가, 구술·면담, 관찰 등으로 분류했다. 이 밖에도 컴퓨터 프로그램이나 영상 도구 등 공학적 도구를 평가 도구로 사용하거나, 학생들이 평가나 활동하는 동안에 기능성 자기 공명 영상 장치와 같은 신경 심리학적(neuropsychological) 도구(Cluss-Ehler, 2010), 시선 추적기 등을 사용한 경우도 기타로 분류했다.

3. 분석 방법

과학 교육에서 평가 연구 분석을 위한 분석틀은 273편의 논문을 1차로 읽은 후 연구자 2인이 범주화하면서 진행했다. 연구 방법론에 대한 국내외 선행 연구들(Sung & Si, 2015; Fraenkel, Wallen, & Hyun, 2015; Neuman, 2011; Kim, Kim, & Kwon, 2012; Kim, 2012; Gall, Gall, & Borg, 2007; Miller, 1991)을 고찰하면서 1차 분석틀을 수정하여 재조정된 2차 분석틀(Table 2)에 맞추어 논문을 분석했다. 분석 대상으로 선정된 273편의 논문은 Table 2의 분석틀에 따라 세부 범주로 구분했다. 기본 자료에서 ‘4. 해당 국가 교육 과정 단계’는 논문의 출판 연도가 아닌, 연구 당시 대상 학생에 해당하는 교육 과정을 기준으로 구분했다. 예를 들어 2009개정 교육 과정이 시행되던 2013년에 발표된 ‘TIMSS 2011과 NAEA 2011에 나타난 우리나라 학생들의 과학 학업 성취 특성’(Dong, Lee, & Shin, 2013) 논문에서 대상 학생은 2011년 당시 중학교 3학년 학생이며 이들은 제7차 교육 과정에 의거한 교과서와 교육 과정의 영향을 받았으므로 제7차 교육 과정으로 분류하는 식이다. 국가 교육 과정의 정확한 적용 시기를 판단하기 어려운 예비 교사, 영재, 교사 등의 경우는 초등학교 1학년 학생을 기준으로 해당 연도에 맞는 교육 과정 단계를 적용했다. 이와 같이 국가 교육 과정을 함께 고려함으로써 교육 과정의 한 부분으로서 평가의 역할을 잘 담당하고 있는지, 교육 과정의 변화에 부응해 평가도 변화하고 있는지 등의 여부를 파악할 수 있다.

양적 연구와 질적 연구의 분류는 방법론적으로는 자료의 수집 방법을 기준으로 구분하기도 한다(Neuman, 2011). 설문지, 점검표, 평가 문항 등 다양한 측정 도구를 사용하여 통계적 분석을 사용한 경우는 양적 연구로, 관찰, 면담, 문서 자료 등을 사용하여 연구자의 의견이

개입되거나 내용 분석, 간단한 기술 통계 분석을 사용하는 경우는 질적 연구로 분류했다. 양적 연구와 질적 연구를 함께 사용하되 그 정도와 수준을 순차적, 병렬적, 등위적, 주-부 패러다임적 혼합, 다층 접근 방법을 사용한 경우(Sung & Si, 2015; Neuman, 2011)는 통합 방법으로 범주화했다. 기타는 연구 방법이 구체적으로 제시되지 않거나 기존 자료를 재구성, 정리하여 어느 쪽으로 분류하기 어려운 경우다(Kim, Kim, & Kwon, 2012).

연구 목적에 따른 기초 연구, 응용 연구, 실험 연구, 평가 연구의 범주는 Sung & Si(2015)의 네 가지 분류를 따랐으나, Miller(1991)는 실험 연구를 제외한 세 가지 영역으로 분류하기도 했다. 실제로 ‘실험 연구’는 응용 연구나 평가 연구로 중복 분류될 수도 있는데, 이 연구에서는 연구의 결과를 일반화시키는 데 있어 제한적인 특수 상황이나 대상, 교육 상황에서 일어나는 문제 해결에 보다 큰 관심이 있는 경우로 한정했다. ‘산업체 부설 특별 학급(학교) 학생의 과학 성취도 평가’(Chung, Yoh, Park, & Kim, 1992) 논문처럼 해당 학생들의 특성에 맞는 성취도 평가 문항을 개발하여 적용하는 경우 실험 연구의 취지에 부합한다고 판단한 것이 그 예다. 평가 연구는 특정 프로그램의 효과나 목표에 도달하는 정도를 측정하는 경우다. 예를 들어 ‘포트폴리오 평가가 초등학생의 과학 탐구 능력 및 과학 관련 태도에 미치는 영향’(Kim, 2002) 논문은 포트폴리오 평가의 효과를, ‘중학생과 고등학생의 과학적 소양 평가’(Chung & Choi, 2007)는 해당 학생들의 과학적 소양 정도를 평가한다는 측면에서 평가 연구로 분류했다.

연구 방법에 따른 방법론 연구는 앞에서 언급한 연구자들의 범주(Sung & Si, 2015; Fraenkel, Wallen, & Hyun, 2015; Gall, Gall, & Borg, 2007)를 참고하여 273편의 논문에서 분석된 방법을 위주로 범주화했는데, 실제 분류한 결과 역사 연구나 문화 기술지 연구의 경우는 나타나지 않아서 이를 제외하고 주로 나타나는 연구 방법을 보다 세분화했다. 특히 집단 비교 연구는 통계를 사용하는 자료의 처리 방법이 유사한 상관 연구로 합쳤고, 조사 연구와 기술 연구도 ‘현상에 대한 기술’이라는 측면(Kim & Choi, 2008)에서 같은 범주로 합쳤다. 그러나 한 편의 논문에서 사용된 연구 방법이 한 가지 이상인 경우가 대부분이며, 여섯 범주 중에는 의미를 확장하면 다른 범주를 포함할 수 있는 경우(Kim & Choi, 2008)도 있다. 따라서 실제 분류에서는 주로 사용한 연구 방법을 1차로 분류하고 부차적으로 사용한 연구 방법을 2차로 적용하여 1차 방법을 중심으로 2차 방법도 함께 정리했다(Table 8). 1차와 2차 방법이 모두 같은 경우는 한 가지 연구 방법을 적용한 경우에 해당한다.

분석 결과를 진술할 때, 학생들의 인지적 성취는 높은 데 비해 과학 학습에 대한 자신감이 부족하고 흥미나 즐거움을 인식하는 정의적 특성이 낮은 국제 비교 연구 결과(Choe *et al.*, 2013; Kim *et al.*, 2009)에 주목했다. 이에 따라 인지적 영역과 정의적 영역을 축으로 연구 방법론과 내용을 분석했다. 또한 선정 논문에서 다루고 있는 주제어를 추출하여 유사한 영역으로 범주화하는 과정을 반복하여 개념, 탐구 및 사고, 정의적 영역, 기타 과학 교육 목표 등 교육 과정에서 목표에 해당하는 영역, 교수·학습 과정과 평가에 해당하는 영역으로 유형화(Table 3)하여 5년을 주기로 한 연구의 변화 경향을 교육 과정의 변화, 평가와 관련한 제도의 변화 등과 함께 의미를 고찰했다(Table 9).

Table 3. Categories of key words from research papers analyzed

범주	세부 범주
목표	개념, 탐구 및 사고, 정의적 영역, 기타 과학 교육의 목표
교수·학습 방법	탐구 학습 및 개방적 탐구 외 9개 영역, 평가에 영향을 미치는 요인
평가	기준, 방법, 결과, 개발, 실패, 대상

III. 연구 결과 및 논의

1. 전체적 연구 동향

1986년부터 2015년까지 과거 30년간 과학 교육에서 ‘평가’ 관련 연구 동향을 KCI와 KCI 등재 후보지에 게재된 논문을 통해 고찰했으며 인지적 혹은 정의적 영역에 대한 논문의 빈도, 교과, 연구 대상에 대한 논문의 수를 10년씩 묶어 집계했다(Table 4).

1986년부터 1995년까지의 기간은 국가 교육 과정상으로 제4차, 제5차, 제6차에 해당한다(Leem & Kim, 2013). 제4차와 제5차 교육과정 중 ‘지도 및 평가 상의 유의점’이라는 항목의 하위 범주에 평가에 대한 사항을 제시했고, 제6차에는 ‘평가’라는 항목으로 독립하면서 내용을 구체화했다(MOE, 1992; MOCE, 1987; MOCE, 1981). 특히 이전 교육 과정과는 달리, 지식을 ‘기본 개념’으로 제4차의 과학에 대한 태도와 제5차의 태도를 제6차에서 ‘과학적인 태도’로 변경했다. 제4차와 제5차 교육과정에서 정의적 영역의 평가는 ‘태도’에만 국한되었으나 제6차에서는 정의적 영역의 구체적인 특성-학습 과정에서

계속 탐구하려는 의욕, 상호 협동, 증거를 존중하는 태도 등을 제시했고 평가의 절차에서 공동 평가 도구를 개발하고 기준 근거에 따라 평가할 것을 제시했다. 1994년에는 대학수학능력시험이 도입, 시행되었다. 이상의 10년간 과학 교육 관련 논문에서 평가에 대한 연구는 28편(10.3%)이 분석되었다. 인지적 영역(7.7%)을 다룬 논문은 주로 탐구 능력이나 탐구 학습 평가 관련 연구가 대부분이었고 대학수학능력시험을 대상으로 한 연구는 주로 1990년과 1991년에 시행되었던 실험 평가를 대상으로 한 논문이 포함되었다. 정의적 영역(1.5%)에서는 모두 태도를 측정하는 도구의 개발과 관련한 연구였다. 이 시기 동안 주 연구 대상이 초등학생(2.6%)과 중학생(5.5%)이었던 점에 비추어 교과는 초등학교와 중학교의 과학과(8.8%)에 제한된 경향이 있으나, 대학수학능력시험이나 탐구 능력에 대한 연구에서 고등학생을 대상으로 한 경우(2.6%)도 있었다.

1996년부터 2005년까지 조사 기간의 두 번째 10년 동안은 국가 교육과정상 제6차, 제7차 교육과정에 해당한다(Kim *et al.*, 2014). 이 기간 동안 1999년에는 수행평가가 점차 확대되었고(Park & Yoo, 1999) 국제 학업 성취도 평가인 TIMSS1995, 1999, 2003과 PISA2000,

Table 4. General trends of assessment researches in science education for recent 30 years(1986~2015)

		과학 교육 평가 관련 논문 수(%) (N=273)			계
		1986년~1995년	1996년~2005년	2006년~2015년	
영역	인지적	21(7.7)	53(19.4)	82(30.0)	156(57.1)
	정의적	4(1.5)	13(4.8)	18(6.6)	35(12.8)
	혼합	2(0.7)	29(10.6)	25(9.2)	56(20.5)
	기타	1(0.4)	13(4.8)	12(4.4)	26(9.5)
교과	과학	24(8.8)	61(22.3)	61(22.3)	146(53.5)
	물리	1(0.4)	6(2.2)	11(4.0)	18(6.6)
	화학	-	8(2.9)	8(2.9)	16(5.9)
	생물	-	9(3.3)	23(8.4)	32(11.7)
	지구과학	1(0.4)	3(1.1)	10(3.7)	14(5.1)
	환경	2(0.7)	-	1(0.4)	3(1.1)
	기타	-	21(7.7)	23(8.4)	44(16.1)
	유아	-	2(0.7)	2(0.7)	4(1.5)
대상*	초등학생	7(2.6)	25(9.2)	25(9.2)	57(20.9)
	중학생	15(5.5)	28(10.3)	47(17.2)	90(33.0)
	고등학생	7(2.6)	24(8.8)	43(15.8)	74(27.1)
	대학생(예비 교사)	-	5(1.8)	13(4.8)	18(6.6)
	교사	4(1.5)	24(8.8)	23(8.4)	51(18.7)
	학부모 외	1(0.4)	2(0.7)	2(0.7)	5(1.8)
	영재 학생	-	6(2.2)	14(5.1)	20(7.3)
	특수 교육 대상자	-	-	3(1.1)	3(1.1)
	특정 대상 없음	3(1.1)	9(3.3)	5(1.8)	17(6.2)
	계	28(10.3)	108(39.6)	137(50.2)	273(100)

* 두 가지 이상의 범주에 해당되는 경우, 중복 집계함.

2003의 결과가 발표되었다(Choe *et al.*, 2013; Kim *et al.*, 2009). 1997년부터는 대학 부설 과학영재교육원이 운영되었고 2000년에는 영재교육진흥법이 제정되었다(Park, 2012). 이전 10년에 비해 발표되는 논문 수는 학회 신설, 학회지 증편 등 다양한 이유로 전체적으로 증가하는 추세였으므로 평가에 대한 논문의 비율이 다른 영역에 비해 증가했다고 하기는 어렵지만, 평가를 다룬 논문의 수는 28편에서 108편으로 이전에 비해 증가했다. 인지적 영역(19.4%)이 정의적 영역(4.8%)에 비해 여전히 높으나 인지적 영역과 정의적 영역을 함께 다룬 혼합 논문(10.6%)의 비율은 상대적으로 높아졌다. 인지적 영역에서는 국가의 교육 정책, 교육 과정이나 평가 제도와 관련하여 수행평가와 관련한 연구, 국제 학업 성취도, 대학수학능력시험의 평가들을 응용한 3차원 평가들을 활용한 연구들이 나타났고, 정의적 영역에서는 이전에 태도에만 국한되었던 특성이 효능감, 인성, 흥미, 지적 호기심 등으로 다양하게 이루어지기 시작했다. 교과는 일반적인 과학(22.3%) 이외에 생물(3.3%), 화학(2.9%), 물리(2.2%), 지구과학(1.1%)이 모두 나타났고, 교사의 인식이나 평가 실태, 과학 교육 일반에 대한 연구와 같이 탈 교과 영역으로도 확대되었다. 연구 대상에서는 영재 학생(2.2%), 대학생(예비 교사)(1.8%), 유아(0.7%)에 대한 연구가 나타나는 한편, 교사를 대상(8.8%)으로 한 인식이나 실태 조사 연구, 교수를 위한 전문적 지식이나 능력, 교과 교육학, 능력 관련 교사 수준의 요인을 고찰하는 연구 등이 증가했다. 특히 영재교육진흥법이 제정되고 시행령이 공포되어 영재 교육 기관이 설립된 이후로 영재 학생이나 영재 교육 대상자에 대한 연구가 꾸준히 증가하는 추세다.

2006년부터 2015년까지 조사 기간의 세 번째 10년의 기간은 제7차 교육 과정 후반부와 2007개정 및 2009개정 교육 과정 시기에 해당한다(Leem & Kim, 2013). 2007개정 교육 과정에서는 지필 검사가 선다형, 서술형, 논술형 등으로 구체화되었고 자유 탐구가 신설되었으며, 평가 항목에서 정의적 영역에 대한 예시에 변화가 있었다(MOEHRD, 2007). 제6차와 제7차 교육 과정에서는 과학적 태도의 관점에서 의욕, 상호 협동, 증거를 존중하는 태도 평가를 강조했다면 2007개정, 2009개정에서는 과학에 대한 흥미와 가치 인식, 과학 학습 참여의 적극성, 협동성, 과학적으로 문제를 해결하는 태도, 창의성 등 과학적 태도 이외에 과학에 대한 태도의 평가를 구체적으로 예시했다(MOEST, 2011; MOEHRD, 2007). 2008년에는 국가 수준 학업 성취도가 전수 체제로 시행되면서 전국 해당 학년의 초·중·고등학생의 학업 성취도를 분석했으나 2013년부터는 초등학교를 제외한 중학생과 고등학생에 대해서만 실시하게 되었다(Cho, Min, & Park, 2015). 이 기간 동안 인지적 영역의 평가 연구는 이전 10년간의 19.4%에 비해 30%로 꾸준히 증가했지만, 정의적 영역(6.6%)이나 혼합 연구(9.2%)는 약간 증가하거나 오히려 감소했다. 인지적 영역에서는 지필 검사의 문항을 개발하거나 분석하는 형태의 연구가 빈번하게 이루어졌고, 학습 동기, 감성, 과제 집착력, 경험 등 이전보다 다양한 정의적 특성에 대한 측정 도구 개발 연구가 진행되었다. 생물(8.4%), 물리(4%), 지구과학(3.7%) 등의 교과에서는 점차 평가에 대한 연구가 증가했고, 교과에 관련 없이 과학 교육 전반(8.4%)에 대한 기초를 세우는 연구도 수행되었다. 중학생(17.2%)과 고등학생(15.8%)을 대상으로 한 평가 연구는 이전 10년간에 비해 각각 6.9%p, 7%p가 증가했는데, 초등학교에 비해 중·고등학교에서 평가의 정례화가 많이 이루어졌고 중·고등학생을 대상으로 한 국제적, 국가적 학업 성취도 시행이 더 많았음을

감안할 때 평가에 대한 관심도가 높게 반영된 결과다. 예비 교사를 대상으로 한 연구(4.8%)는 3%p가 증가하여 논문 수의 절대치는 낮지만 상대적인 비율이 증가했는데, 이들을 대상으로 한 연구는 실행을 위한 접근이 용이하고 대상자의 연구에 참여하려는 태도가 비교적 분명하게 드러난다. 또한 초·중·고등학생을 대상으로 한 연구로 발전시킬 때 사전적, 완충적 역할을 할 뿐 아니라, 예비 교사 스스로도 '평가'에 대한 학습을 경험할 수 있다. 영재 학생을 대상으로 한 연구(5.1%)는 각 영재 교육 기관에서 운영한 결과가 누적됨에 따라 점차 증가하고 있고, 특수 교육 대상자를 대상(1.1%)으로 과학 개념, 탐구 기능, 수업 참여나 성취도에 대한 연구도 이루어져 학습자의 지적 특성을 고려한 과학 교육 가능성을 탐색했다.

1986년 이후 30년 동안 과학 교육에서 이루어진 평가 관련 연구 동향을 정리하면, 국가 교육 과정의 변화와 함께 평가와 관련한 대내외적 사건을 반영한 경우가 많았다. 인지적 영역(57.1%)을 다룬 연구는 정의적 영역(12.8%)이나 인지와 정의적 영역을 함께 다룬 연구(20.5%)에 비해 많이 나타났다. 초반에는 일반적인 탐구 능력 측정에 관심이 많았으나 점차 교과와 연관된 개념이나 능력, 사고 등으로 확대되었고 정의적 영역에서도 과학적 태도나 과학에 대한 태도 이외에 좀 더 구체화되거나 세분화된 정의적 특성을 다루는 연구로 진행되었다. 교과는 초등학교와 중학교 수준의 과학 일반에 대한 연구(53.5%)가 가장 많았으며 교과에 관련 없이 과학 교육 일반에 대한 연구(16.1%)도 수행되었고, 생물(11.7%), 물리(6.6%), 화학(5.9%), 지구과학(5.1%) 순으로 나타났다. 연구 대상은 중학생(33.0%), 고등학생(27.1%), 초등학생(20.9%) 순으로 나타났고, 영재 학생(7.3%)의 경우 다양한 활동과 연관된 평가를 개발하거나 적용할 수 있다는 점에서 점차 그 비율이 늘어나고 있다. 한편, 최근 들어 유아(1.5%)나 특수 교육 대상자(1.1%)에 대한 평가도 발표되고 있고, 예비 교사나 대학생(6.6%)을 대상으로 한 연구도 나타나 연구 대상자가 점차 다양해지고 있다.

2. 인지적 영역과 정의적 영역의 평가 연구 동향

인지적 영역과 정의적 영역의 평가 연구를 분석하는 관점을 인식론적 관점(Table 5), 연구에서 다루는 평가 도구(Table 6), 연구 대상자(Table 7)를 중심으로 분석했다. 철학적 입장이 다른 두 연구자, 즉 객관적인 실재를 인정하고 지식을 일반화하는 데 관심이 있는 연구자와 일반화시킬 수 있는 본질은 없으며 단편적이 아니라 총체적으로 연구해야 한다고 생각하는 연구자는 문제에 대한 접근 방법이 서로 다르다(Sung & Si, 2014; Chae, 1996). 객관적 자료를 근거로 일반적 원리와 법칙을 발견하려는 연구자는 양적 연구 방법을, 특정 현상에 대하여 깊이 있고 자세한 지식을 얻으려는 연구자는 질적 연구 방법을 주로 사용한다. 어떤 연구 방법을 사용했는지의 여부는 그 연구의 기저를 이루고 있는 철학적 입장에 대한 통찰과 함께 문제를 제기하고 그에 접근하는 방법에 대한 정보를 제공한다.

Table 5에서 과학 교육의 평가와 관련된 연구들은 75.5%가 실증적인 자료를 활용한 양적 연구로 분석된 반면, 질적 연구는 6.6%, 양적인 자료와 질적인 자료를 통합하여 보완적으로 사용한 연구는 8.4%로 나타났다. 평가와 관련된 현상에 주목하여 해석하고 의미를 이해하려는 질적 연구가 양적 연구에 비해 상대적으로 적었으나 1990년대

Table 5. Research types

	영역 별 인식론적 접근 방법에 따른 연구 유형 논문 수 (%) (N=273)				계
	인지적	정의적	혼합	기타	
양적 연구	118(43.2)	28(10.3)	43(15.8)	17(6.2)	206(75.5)
질적 연구	9(3.3)	4(1.5)	1(0.4)	4(1.5)	18(6.6)
혼합 연구	12(4.4)	1(0.4)	8(2.9)	2(0.7)	23(8.4)
기타	17(6.2)	2(0.7)	4(1.5)	3(1.1)	26(9.5)
합계	156(57.1)	35(12.8)	56(20.5)	26(9.5)	273(100.0)

후반부터는 양적 연구와 질적 연구를 혼합한 연구가 점차 증가하는 추세이며(Figure 1), 이러한 혼합 연구의 증가 경향은 2000년대 초반 이후 국제적인 추세로 보고된 바 있다(Lawrenz & Thao, 2014). Shin & Choi(2014)는 양적 연구에서는 논의와 글쓰기의 효과 분석, 인식 조사 연구 등이 많았고 질적 연구에서는 활동 내용을 분석한 사례 연구가 많았음을 언급했다. 이와 비교해 평가 연구에서는 평가를 적용한 효과 분석, 인식과 실태 조사 연구 뿐 아니라 도구 개발, 상관 연구, 집단 비교 등 연구 목적과 방법, 내용이 다양해지면서 여러 가지 형태의 자료를 다루었다. 따라서 평가 연구에서 양적 연구의 비율이 상대적으로 높게 나타난다는 것은 다양한 형태의 자료라 하더라도 연구자들은 결국 양적으로 표현된 자료를 더 선호하고 양적 자료 수집이 가능한 주제에 더 관심이 많다고 해석할 수 있다. ‘평가와 측정’에 내재한 수량화 경향이 연구에 반영되었음도 가능한 해석이다.

인식론적 접근 방법에 따라 인지적, 정의적 영역의 평가를 다룬 연구의 분포를 나타낸 Table 5에서 기타 영역을 제외하고 공통적으로 나타나는 사항은 양적 연구에 대한 질적 연구의 비율이 인지적 영역은 7.6%, 정의적 영역은 14.6%, 혼합의 경우는 2.5%로 비교적 낮게 나타난다는 점이다. 양적 연구에 대한 통합 연구의 비율은 각각 10.2%, 3.9%, 18.4%이므로, 질적 자료를 사용한 경우를 모두 포함한 질적 연구와 통합 연구의 양적 연구에 대한 비율은 인지적 영역 17.8%, 정의적 영역 18.5%, 혼합 20.9%로 20% 내외에 해당한다. 따라서 인지적 혹은 정의적 영역이냐에 따라 양적 연구나 질적 연구를 선호하는 경향은 나타나지 않으나 거의 모든 영역에서 양적 연구를 4배 이상 더 많이 사용하고 있음을 알 수 있다.

인지적, 정의적 영역 모두에서 양적 연구가 편중되어 나타나는 현상에서 두 가지 교육적 시사점을 얻을 수 있다. 첫째, 과학 교육 관련 평가 연구의 주제가 양적 자료를 얻기 쉬운 주제로 제한될 가능성이 있다. 인지적 영역에 대한 평가 결과는 학교 교육 과정에서 쉽고 빈번하게 얻을 수 있는 반면, 정의적 영역의 평가는 공식적으로 잘 이루어지지 않았을 뿐 아니라 그 특성을 규명하여 수량화한 연구가 적은 편이다. 따라서 실증적 경향이 있는 연구자들의 관심을 받기 어렵다. 둘째, 평가 연구의 목적과 상황에 맞는 적절한 연구 방법을 간과할 가능성이 있다. 평가와 관련한 현상에 대하여 양적 연구는 일반적 지식을 얻고 싶을 때, 질적 연구는 깊이 있는 자세한 지식을 얻고 싶을 때 사용된다(Chae, 1996). 이러한 상황을 고려하지 않고 양적 연구를 우선하는 연구 풍토는 구체적이고 개별적인 평가 상황에 대한 이해를 더 어렵게 할 수 있다.

한편, 연구에서 사용한 평가 도구를 영역별로 분석한 결과가 Table 6에 정리되었다. 평가를 주제로 한 연구에서 평가 도구는 연구자가 연구 대상을 평가하는 도구와 연구 내용에서 사용한 평가 도구를 모두 포함한다. 예를 들어 ‘STS에 대한 고등학생들의 견해에 관한 평가 도구 개발’(Lim *et al.*, 2004)은 전자에 해당하고 ‘형성 평가의 피드백 유형이 학생들의 과학 성취와 태도, 교사-학생 상호작용에 미치는 영향’(Lee, Choi, & Nam, 2000)이라는 연구 내용에서 평가 도구는 ‘형성 평가’에서 사용된 평가 도구를 말한다. 평가 도구가 연구에 포함된 경우는 273편 중 228편에 해당하지만, 한 편의 논문에는 여러 형태의 평가 도구가 사용되거나 언급된 경우가 있었으므로 중복 집계된 경우를 모두 포함했다. 평가 도구를 조사한 이유는 평가라는 과정이 학생

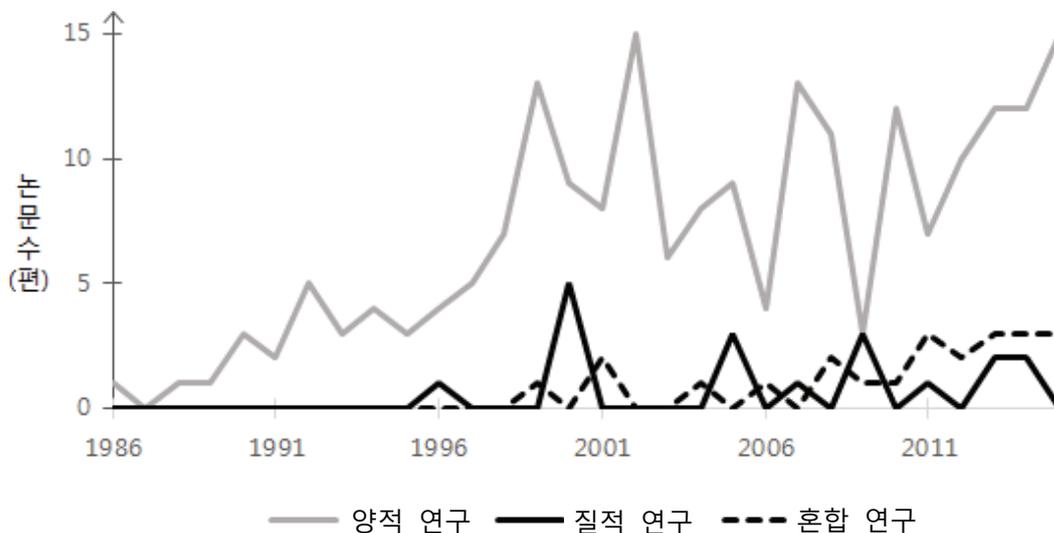


Figure 1. Number of research papers of science assessment by epistemic method

의 변화 정도를 파악하기 위하여 정보를 수집하는 활동(Kim et al., 2014)이 기본이 되며, 그 과정에는 어떤 형태로든 도구가 사용되기 때문이다. 좋은 도구는 측정하려는 대상을 정확히 파악하고, 보려는 바를 잘 드러내어 교육적 판단을 내리기 쉬운 상태로 만들어 주기 때문에 연구 자체에서 대상을 평가하는 도구나 연구 내용에서 사용한 평가 도구는 그 연구의 성패를 좌우하는 동시에 연구의 타당성과 신뢰성을 높여주는 중요한 장치다. 그러한 의미에서 본 연구에서 조사된 평가 연구에서는 도구를 개발하는 연구(37.4%)가 다른 연구에 비해 높게 나타나기도 했다(Table 6). 평가 도구 유형은 Kim et al.(2014)에서 평가를 시행하는 방법으로 제시한 구술, 지필, 보고서/과제물/포트폴리오, 관찰, 면담, 설문 등을 참고하여 228편의 논문에서 도출된 평가 도구를 귀납적으로 범주화했다.

가장 많이 활용된 유형은 지필 평가로 선택형(46.1%), 서답형(30.7%), 리커트형(26.3%)의 순으로 나타났는데, 인지적 영역에서는 선택형(32.0%)과 서답형(22.4%)을 많이 사용하고 리커트형(2.6%)이 적은 반면, 정의적 영역에서는 리커트형(10.5%)의 비율이 선택형(1.3%)이나 서답형(1.3%)보다 높게 나타났다. 일반적으로 인지적 영역에서는 객관적 평가를, 정의적 영역에서는 자기보고식 평가를 주로 사용하고 있었다. 설문이나 리커트형으로 실시되는 자기보고식 평가는 일반적 반응 경향성, 긍정 혹은 부정적 편향성, 보고의 정확성에 대한 문제점(Song, 2010)이 제기되어, 보다 다양한 평가 방법을 보완해야 할 필요가 있다. 인지적 영역에서도 기존의 객관적 평가에 대한 대안적 평가 도구가 심리 검사를 중심으로 탐색되고 있다(Holling & Preckel, 2005; Bailey & Bailey, 1974). Lee, Kim, & Jung(2006)은 자기보고식 지능 검사의 가능성을 탐색한 연구에서 목적 은폐 전략과 객관적 입증 가능성 전략을 사용하여 지능 검사의 가능성과 타당성을 모색했다. 기존에 주로 사용되는 평가 도구의 유형 외에 다른 유형의 도구를 사용할 경우 특정한 상황에서 학습자에 대한 풍부한 이해를 향상시킬 수 있다는 점에서 상황에 따라 적절히 활용할 수 있는 다양한 유형의 평가 도구가 개발될 필요가 있다. 그러한 의미에서 지필 평가 이외에 수행을 중심으로 한 평가를 조사한 결과 보고서, 포트폴리오, 과제물 등의 산출물을 통한 평가(17.1%)가 가장 많이 활용되었

으며 평정형(8.3%), 구술·면담(5.3%), 관찰(4.8%), 자기·상호 평가(3.1%) 순으로 분석이 되었다. 주목할 점은 정의적 영역을 수행형 도구로 평가한 경우는 산출물(1.3%)이나 구술·면담(0.9%)을 제외하고는 거의 이루어지지 않았고 인지적 영역에 집중되었다는 것이다. 수량화하기 어려운 학습자들의 태도, 정서와 감정, 느낌을 측정하는데 활용된 도구가 지필형인 경우가 대부분이고 수행 과정에서 정의적 영역을 평가하는 경우가 거의 전무하다는 점은 수행형 평가가 인지적 능력 뿐 아니라 정의적 능력과 심동적 능력을 모두 평가할 수 있는 총체적 접근(Kim et al., 2014)이라는 점에서 역설적이다. 정의적 영역에 대한 특성 규명과 함께 이를 정확하게 표현해서 측정할 수 있도록 다양한 유형의 도구에 대한 탐색과 개발, 활용에 대한 연구자들의 관심이 증가되어야 할 필요가 있다.

지필 평가나 수행형 평가처럼 교육 현장에서 널리 사용되고 있는 평가 방법 이외에도, 컴퓨터 프로그램을 포함한 공학적 도구를 사용하는 경우(3.1%), 뇌 영상 기법이나 시선 추적과 같은 신경 심리학적 도구(2.6%)를 사용하는 경우도 분석되었다. 심리학계에서는 뇌 영상 기법으로 인지적 판단을 내리는 뇌의 기능을 정서적 관련성과 연관하여 다양하게 접근하고 있다(Wang, Huettel, & De Bellis, 2008; Huizinga, Crone, & Jansen, 2007). 국내 과학 교육에서는 뇌 기반 과학 교육 도입의 필요성을 제기하면서 다양한 활용 사례를 모색하고 있는데 특히 교수 학습 프로그램에 적용하여 실증적으로 학생들의 변화를 평가할 수 있는 가능성을 탐색하고 있으며(Lee, Lee, & Kwon, 2010), 아직까지는 인지적 영역에 연구가 국한되어 있다.

연구 대상을 영역별로 정리한 결과가 Table 7에 정리되었다. 중학생(33.0%), 고등학생(27.1%), 초등학생(20.9%) 순으로 연구가 수행되었다. 이들에 대한 인지적, 정의적 영역의 논문 수는 인지적 영역이 정의적 영역에 비해 초·중·고등학생의 경우 4.2배~4.8배로 높은 편이나, 인지적 영역과 정의적 영역을 함께 다룬 혼합 연구의 비율은 그 차이가 다소 감소했다. 전체적으로 인지적 영역 평가가 우세한 편이지만, 정의적 영역에 대한 평가도 어느 정도 반영하고 있었다. 대상에 따른 영역별 평가 연구의 경향성이 뚜렷하게 나타나는 편은 아니지만, 영재 학생(7.3%)이나 대학생과 예비 교사(6.6%)의 경우에

Table 6. Types of assessment tool

		영역별 평가 도구 유형 논문 수(%) (N=228)				계
		인지적	정의적	혼합	기타	
지필	선택형	73(32.0)	3(1.3)	24(10.5)	5(2.2)	105(46.1)
	서답형	51(22.4)	3(1.3)	14(6.1)	2(0.9)	70(30.7)
	리커트	6(2.6)	24(10.5)	26(11.4)	4(1.8)	60(26.3)
	산출물	26(11.4)	3(1.3)	10(4.4)	-	39(17.1)
수행	평정형	8(3.5)	-	10(4.4)	1(0.4)	19(8.3)
	구술·면담	6(2.6)	2(0.9)	3(1.3)	1(0.4)	12(5.3)
	관찰	6(2.6)	-	5(2.2)	-	11(4.8)
	자기·상호 평가	1(0.4)	-	3(1.3)	3(1.3)	7(3.1)
기타	공학적 도구	5(2.2)	-	2(0.9)	-	7(3.1)
	신경 심리학적 도구	4(1.8)	1(0.4)	1(0.4)	-	6(2.6)
계 (A)		186	36	98	16	336
중복 집계된 논문 수 (B)		54	5	46	3	108
합계 (A-B)		132(57.9)	31(13.6)	52(22.8)	13(5.7)	228(100.0)

주: 두 가지 이상의 범주에 해당되는 경우, 중복 집계함

Table 7. Types of research object

	영역별 연구 대상자 유형 논문 수 (%) (N=273)				계
	인지적	정의적	혼합	기타	
유아	3(1.1)	-	1(0.4)	-	4(1.5)
초등학생	33(12.1)	8(2.9)	15(5.5)	1(0.4)	57(20.9)
중학생	57(20.9)	12(4.4)	17(6.2)	4(1.5)	90(33.0)
고등학생	47(17.2)	10(3.7)	15(5.5)	2(0.7)	74(27.1)
대학생(예비 교사)	6(2.2)	4(1.5)	5(1.8)	3(1.1)	18(6.6)
교사, 교수, 교육 전문가	24(8.8)	2(0.7)	6(2.2)	19(7.0)	51(18.7)
학부모	1(0.4)	2(0.7)	1(1.1)	1(0.4)	5(1.8)
영재 학생	10(3.7)	5(1.8)	5(1.8)	-	20(7.3)
특수 교육 대상자	3(1.1)	-	-	-	3(1.1)
특정 대상 없음	9(3.3)	3(1.1)	3(1.1)	2(0.7)	17(6.2)
계 (A)	193	46	68	32	339
중복 집계된 논문 수 (B)	37	11	12	6	66
합계 (A-B)	156(57.1)	35(12.8)	56(20.5)	26(9.5)	273(100.0)

주: 두 가지 이상의 범주에 해당되는 경우, 중복 집계함

는 혼합과 정의적 영역의 비율이 인지적 영역과 유사하게 나타나고 있었다.

3. 연구 방법론과 내용에 대한 평가 연구 동향

연구를 수행한 목적과 사용한 연구 방법론에 따라 과거 30년간 발표되었던 273편의 평가 관련 논문을 분석했다(Table 8).

Sung & Si(2014)가 제시한 기초 연구, 응용 연구, 실험 연구, 평가 연구로 구분하여 분류하면 1986년부터 2015년까지 과학 교육에서 평가를 주제 혹은 소재로 한 연구 목적은 응용 연구(76.9%), 평가 연구(15.4%), 실험 연구(5.5%), 기초 연구(2.2%) 순으로 나타났다.

실제 문제 상황에서 적용된다는 측면에서 다른 유형으로 분류한 평가 연구나 실험 연구도 넓은 의미에서는 응용 연구에 속할 수 있다는 관점에서 본다면 지금까지 국내 연구에서 ‘평가’라는 주제는 실제 맥락에서 실증적 자료에 근거하여 연구되어 왔다. 반면, 이론을 규명하여 지식을 확장시키려는 기초 연구가 상대적으로 적은 것(2.2%)은 연구자들이 평가의 기저를 이루는 기본 원리와 철학적 고찰, 새로운 이론과 모형의 형성과 같은 이론적 영역보다는 실용적, 응용적 부분에 관심을 더 기울이고 있으며, 앞으로 관련 분야의 연구 지평을 넓혀 나갈 수 있는 여지가 있음을 시사한다. 평가 연구는 ‘평가’를 소재로 한 수업이나 프로그램의 효과를 측정하는 연구와 연구 대상이 목표에 도달한 정도를 측정하는 연구 등 두 유형으로 분석되었다. 수업이나

Table 8. Analyses by research purpose and major methodology

1차	주요 연구 방법 2차	연구 목적에 따른 논문 수 (%) (N=273)					소계	총계
		기초	응용	실험	평가	기타		
개발 연구	문헌 연구	-	9(3.3)	-	-	-	9(3.3)	102(37.4)
	기술·조사 연구	-	65(23.8)	4(1.5)	-	-	69(25.3)	
	상관 연구·집단 비교 연구	-	11(4.0)	1(0.4)	3(1.1)	-	15(5.5)	
기술·조사 연구	관찰 연구	-	9(3.3)	-	-	-	9(3.3)	51(18.7)
	문헌 연구	-	27(9.9)	-	1(0.4)	-	28(10.3)	
	조사·기술 연구	-	21(7.7)	-	2(0.7)	-	23(8.4)	
상관 연구·집단 비교 연구	기술·조사 연구	-	26(9.5)	1(0.4)	12(4.4)	-	39(14.3)	43(15.8)
	집단 비교 연구·상관 연구	-	4(1.5)	-	-	-	4(1.5)	
관찰 연구	기술·조사 연구	-	16(5.9)	9(3.3)	4(1.5)	-	29(10.6)	30(11.0)
	상관 연구·집단 비교 연구	-	1(0.4)	-	-	-	1(0.4)	
실험 연구	문헌 연구	-	1(0.4)	-	-	-	1(0.4)	29(10.6)
	기술·조사 연구	-	2(0.7)	-	11(4.0)	-	13(4.8)	
	상관 연구·집단 비교 연구	-	6(2.2)	-	5(1.8)	-	11(4.0)	
문헌 연구	관찰	-	-	-	4(1.5)	-	4(1.5)	18(6.6)
	문헌 연구	6(2.2)	12(4.4)	-	-	-	18(6.6)	
	합계	6(2.2)	210(76.9)	15(5.5)	42(15.4)	-	273(100.0)	

프로그램에서 소재가 된 평가 방법은 주로 수행 평가, 포트폴리오, 형성 평가 및 진단 평가, 자기 평가 및 동료 평가 등으로 나타났다. 실행 연구는 통계 자료의 사용이나 연구 설계 측면에서 응용 연구나 평가 연구와 유사하여 구분이 어려웠지만, 연구자가 접촉하기 쉬운 집단을 대상으로 적용했으며 연구 결과의 해석이 제한될 수 있는 경우로 분류했다. 실행 연구의 경우에는 자료 수집과 분석의 측면에서 양적 연구, 질적 연구, 혼합 연구 등을 적용했는데 연구 대상의 특성을 고려하여 내용과 목적에 맞는 연구 방법을 적절하게 선택해야 할 필요가 있는 사례도 있었다.

연구 방법에 따른 방법론은 조사 대상인 273편의 논문에서 분석된 방법만으로 정리했다. 일차적으로는 선행 연구자들이 분류했던 상관 연구, 집단 비교 연구, 조사 연구, 기술 연구, 실험 연구 등으로 세분화하여 분류했고 이후 자료 처리 방법이 유사한 경우(Sung & Si, 2014), 현상의 기술이라는 연구 목적이 같은 경우(Kim & Choi, 2008) 등으로 보다 넓게 범주화했다. 한 가지 연구 방법을 주도적으로 적용한 경우는 극히 드물고 다양한 연구 방법을 사용하는 연구가 많아서 주요 방법인 1차 연구 방법을 중심으로 보충적인 2차 연구 방법을 함께 나타냈다. 평가 관련 연구 중 가장 많이 나타난 연구 방법은 개발 연구(37.4%), 기술 연구·조사 연구(18.7%), 상관 연구·집단 비교 연구(15.8%), 관찰 연구(11.0%), 실험 연구(10.6%), 문헌 연구(6.6%) 순으로 나타났다.

개발 연구는 다른 연구 방법과 차별화되는 몇 가지 특징이 있는데, 문항이나 측정 및 평가 도구를 창안하는 것을 기본으로 하여 도구의 검증과 적용을 어떻게 조합하고 어떤 단계까지 진행하느냐에 따라 함께 사용하는 2차 연구 방법이 달라졌다. 즉 ‘개발’의 과정을 무엇으로 생각하느냐에 따라 연구 내용과 설계가 달라지면서 몇 가지 유형으로 나타났다. 첫째, 가장 간단한 개발 연구는 평가 도구의 창안만을 다룬 경우이며 이 경우에는 문헌 연구를 2차 연구 방법으로 사용했다. 두 번째로 평가 도구를 창안하고 이에 대한 타당도와 신뢰도를 검증하는 연구는 2차 연구 방법으로 기술 연구·조사 연구를 진행하는 경우가 많았다. 이때 검증 과정에서는 요인 분석, 문항 반응, 신뢰도 등을 통계적 방법으로 산출하는 경우도 있었으나, 설문 조사를 통해 만족도를 조사하는 경우도 있었다. 세 번째는 평가 도구의 창안, 검증에 이어 도구를 현장에 직접 투입하여 적용하는 종합적 연구로서 이 경우에는 2차 연구 방법으로 기술 연구·조사 연구 외에 상관 연구·집단 비교 연구, 관찰 연구도 사용했다. Liu(2012)는 측정 도구를 타당화하는 전형적인 연구 방법으로 예비 검사의 수정에서부터 본 검사에 이르기까지 여러 단계로 진행되는 연구들을 보고했고, 다른 나라에 번역되거나 다문화적인 상황에서 타당도와 신뢰도를 확보하는 연구로 확장되는 경우들을 소개했다. 국내에서는 외국에서 개발된 도구를 국내의 상황에 맞추어 타당화하는 과정으로 응용된 연구들이 이에 해당한다(Ha & Lee, 2013; Jo & Ku, 2013). 개발 연구의 네 번째 유형은 도구를 창안하고 특별한 검증 과정 없이 현장 적용한 결과를 나타낸 경우다. 이 경우 사용되는 2차 연구 방법은 통계적 방법을 사용하는 연구 방법 외에 주로 관찰 연구가 많이 사용되었는데, 평가 도구 자체가 다수를 대상으로 한 통계적 방법을 사용하여 검증하거나 적용하기 어려운 경우였다.

관찰 연구(11%)가 실험 연구(10.6%)와 비슷한 비율로 나타났다는 사실에 주목할 필요가 있는데 관찰 연구는 사례에 대한 관찰과 분석

을 다루는 대표적인 질적 연구 방법이다(Kim & Choi, 2008). 특히 실험 연구 중에서도 집단을 비교하기 위하여 정량적 방법 외에 관찰(1.5%)을 사용한 경우도 있음을 고려할 때, 최근 30년 간의 평가 관련 연구에 있어서 질적 연구인 관찰 연구가 양적 연구인 실험 연구보다 많이 수행되었다고 할 수 있다. 최근 들어 질적 자료를 수집하거나 양적, 질적 자료를 함께 사용하는 경우가 꾸준히 증가하고 있어 실제 맥락에서 평가 현상을 자세히 분석하려는 경향이 나타나고 있음을 알 수 있다. 실험 연구에서는 실험 집단과 비교 집단을 설정하여 집단을 비교하거나 변인과의 상관을 보려는 연구(4%)와 함께 기술 연구·조사 연구(4.8%)도 2차 연구 방법으로 사용되었다. 이는 연구자가 실험 연구 외에 설문지 등을 활용한 조사 연구를 병행했거나, 동일 집단에 대한 실험이거나, 표본 수와 대상의 특성 상 집단 비교보다 결과의 기술에 중점을 둔 경우로 분류했다.

연구 목적과 연구 방법을 함께 비교해 보았을 때, 기초 연구의 경우 문헌 연구(2.2%)를 활용한 경우가 대부분으로 모형의 수립과 탐색, 새로운 이론을 문헌 고찰을 통해 소개했다. 국내 연구에서 평가 모형의 수립과 관련한 이론을 연구한 경우가 비교적 드문 반면, 국외의 경우 교실 수준의 평가 모형, 학습을 위한 평가 모형, 프로그램 평가 모형들을 철학적, 기술적 관점으로 소개하는 이론 연구가 활발히 진행되고 있다(Lawrenz & Thao, 2014; Pellegrino, 2012; Cowie, 2012).

응용 연구는 전체 연구의 76.9%를 차지한 만큼 다양한 연구 방법을 활용했는데 특히 기술 연구·조사 연구를 함께 수행한 개발 연구가 23.8%로 가장 빈도가 높았고, 문헌 연구를 부차적으로 사용한 기술 연구·조사 연구가 9.9%, 기술 연구·조사 연구를 가미한 상관 연구·집단 비교 연구가 9.5%로 나타나 전체적으로 기술 연구·조사 연구가 세부적으로는 많이 사용되었음을 알 수 있다. 실행 연구에서는 연구 대상에 맞는 문항이나 프로그램을 개발(1.9%)해서 적용하고 그 결과를 양적인 결과로 기술하거나 비교하는 경우도 있었지만 관찰 연구로 상황을 기술하는 경우(3.3%)가 더 많았다. 평가 연구의 경우 평가 관련 프로그램의 효과는 주로 실험 연구(7.3%)와 관찰 연구(1.5%)를 통해서 평가했고, 목표 도달 정도의 평가는 상관 연구·집단 비교 연구(4.4%)와 기술 연구·조사 연구(1.1%)로 나타났다.

분석 논문의 주제어를 유사한 영역으로 범주화하여 과학 교육의 목표, 교수·학습 과정, 평가에 해당하는 영역으로 유형화하여 12개의 주제어를 추출(Table 3)하고 관련 하위 범주별로 묶었다. 이를 5년 주기로 논문의 빈도수를 분석하여 30년 동안 빈도수가 많은 순으로 나타냈고(Table 9), 변화의 경향을 파악하기 위하여 국가 교육 과정 및 평가와 관련한 제도와 쟁점의 변화 등을 함께 고려해 의미를 고찰했다. 과학 교육의 목표에 해당하는 주제어는 탐구 및 사고(19.4%), 정의적 영역(15.8%), 기타 과학 교육의 목표(2.9%), 개념(1.8%) 순으로 나타났다. 교수·학습 과정에서는 세부적인 10항목(26.7%), 평가에 영향을 미치는 요인(5.1%) 순으로, 평가에서는 개발(49.8%), 결과(28.2%), 방법(27.8%), 대상(14.3%), 기준(13.6%), 실태(9.5%) 등의 순으로 나타났다.

평가 방법은 전체의 27.8%로 나타났는데, 가장 빈번하게 연구되었던 방법은 수행형 평가(12.8%), 포트폴리오와 보고서(4.8%), 진단 평가와 형성 평가(3.7%), 자기 평가와 동료 평가(2.6%) 순이었다. 수행형 평가에 대한 연구는 1999년 수행평가가 실시되기 시작했던 해를 전후로 1996년부터 2005년까지 10년간 8.4%가 이루어졌는데, 수행

평가의 주요 방법 중 하나인 포트폴리오와 보고서에 대한 연구가 같은 기간 동안 4.4% 이루어졌던 것을 감안한다면, 이 시기 수행 평가의 도입이 평가 제도에 있어서 하나의 전환기였으며 이에 따라 활발한 연구가 진행되었음을 알 수 있다. 수행 평가의 시행과 함께 학습의 결과보다 과정 평가에 대한 관심이 증가하면서 진단 평가와 형성 평가에 다양한 내용, 도구, 방법을 사용하여 효과를 측정하는 연구(3.7%)가 나타났는가 하면, 학습의 과정에서 자기를 평가하는 방법을 학습 전략으로 활용한 연구들(Nam *et al.*, 2004; Hwang *et al.*, 2001)과 주로 교사와 예비 교사를 대상으로 동료 평가를 활용한 연구들(2.6%)도 진행되어 방법적 측면에서 새로운 시도가 나타나기도 했다.

평가와 관련된 교수·학습 과정에서 해당 시기 동안 꾸준히 연구되고 있는 주제는 실험(6.2%)인데, 특히 제6차 교육과정과 제7차 교육과정 전환기인 1996년과 2000년 사이에 3.3% 정도 수행되었다. 논증 과정이나 논의에 대한 연구(1.1%), 컴퓨터나 웹사이트 등을 평가 도구로 활용하는 연구(2.9%)도 10년 전부터 발표되었다. 인지 발달과 개념 이해, 학습 발달 과정, 지식 상태 등 학생들의 과학적 사고와 인지 발달 과정을 추적 하는 연구들(5.5%)도 최근 점차 증가하는 경향이 있는데 학생들이 경험하는 교수·학습 과정이 학생의 실제적 발달과 일치하는지를 구체적인 과학 개념이나 이론과 연관하여 밝히려는 특징을 보인다. 이는 평가와 교수·학습 과정의 일관성 뿐 아니라 학생의 발달을 고려하여 실제적 맥락에서 교육과정을 보다 정교화 할 수 있는 증거를 제공한다는 점에서 의미가 있다. 한편, 교사를 대상으로 한 교과 교육학 지식과 전문성에 대한 평가 연구(1.8%)도 수행되었다.

과학 교육의 주요 목표인 탐구 능력에 대한 연구(13.2%)는 30년 동안 지속적으로 수행되었는데, 관찰, 측정, 문제 발견, 변인 확인, 실험 설계 등 탐구 능력의 하위 영역별로 세분화하여 진행되었다. 시간이 지날수록 탐구 능력에 대한 새로운 도구 개발보다는 기존 도구의 활용이나 연구 목적에 맞도록 수정하여 사용하는 경우가 많았다. 평가와 관련된 사고에서 나타나는 주제어는 문제 해결력(1.8%), 논리적 사고, 과학적 사고, 창의성, 시스템 사고(1.1%) 등으로 나타났지만, 논리적 사고나 과학적 사고라는 용어는 보다 구체적 용어로 대체되었다. 제7차 교육과정의 지구과학 I에서 '지구계' 내용이 도입됨에 따라(MOE, 1997b) 시스템 사고가 평가와 관련한 연구에서 최근 등장하기 시작한 반면, '창의성'을 주제로 한 평가 연구는 0.4%에 그쳤는데 이는 2007개정 교육과정에서 과학 교육의 평가 항목으로서 창의성을 명시한 것(MOEHRD, 2007)과 대비된다. 창의성을 정확히 측정하기 어렵고 현재 출판된 창의성 검사의 타당도가 비교적 낮으며(Davis, Rimm, & Siegle, 2011) 국내외에서 개발된 검사 도구가 상업용인 경우가 많다는 점(Kang *et al.*, 2008)을 고려할 때 학교 현장에서 쉽게 활용할 수 없었을 것으로 추측된다. 과학 교육에서 창의성 평가에 더 많은 관심과 연구가 필요하다.

정의적 영역에 대한 평가 연구는 태도에 대한 연구(7.3%)와 태도 이외의 영역-동기, 감성, 몰입, 자아 개념, 과학 경험, 확산성, 과제 집착력, 자기 주도성, 성취 정서, 참여, 효능감, 신념 등을 다룬 연구(8.3%)로 분류되었다. 특히 제6차 교육과정부터 평가 항목이 독립 영역으로 분리되면서 정의적 영역 평가의 구체적 예가 명문화되는데 학습 과정에서 계속 탐구하려는 의욕, 상호 협동, 증거를 존중하는 태도 등이 그것이다(MOE, 1997a). 2007개정 교육과정에서는 과학에 대한 흥미와 가치 인식, 과학 학습 참여의 적극성, 협동성, 과학적으로

문제를 해결하는 태도로 언급되었다(MOEHRD, 2007). 태도 이외의 정의적 영역 관련 평가 연구가 등장하는 시기도 이 때부터인데 평가의 목적이 다양화되며 평가 방법에 대한 다양한 정보를 제공하는 연구가 증가했고 특히 2011년 이후 정의적 영역의 평가 도구에 대한 연구와 현상 분석 및 조사 연구가 증가하는 추세다.

이 밖에도 영재 교육 대상자(8.4%), 유아(2.6%), 예비 교사(2.2%), 특수 교육 대상자(1.1%)를 대상으로 한 연구가 수행되었고, 특히 유아와 특수 교육 대상자에 대한 연구는 대상의 특성에 적합한 평가 도구와 방법을 적용해야 했다. 평가 기준에 대한 연구는 평가 기준·준거·성취 기준(5.9%), 과학 평가틀(2.9%), 교육 목표 분류틀(2.6%), 국가 교육 과정(2.2%) 순으로 나타났는데 평가 기준이나 준거를 개발하는 연구와 함께 일정 모형에 맞추어 문항을 분석함으로써 평가 문항이 목표에 비추어 얼마나 타당한지 조사하는 연구 등으로 전개되었다. 분석 기준으로는 3차원 평가틀이 1990년대 이후로 꾸준히 사용되고 있고, 최근에는 Klopfer와 Bloom의 목표 분류틀을 문항에 적용한 연구도 있었다(Choi & Paik, 2015; Lee & Jeong, 2014; Park, Bae, & Jo, 2012; Kim, Yoon, & Kwon, 2010; Seo, Kim, & Chae, 2010; Paik *et al.*, 2008; Yang *et al.*, 2008). 이들 연구의 결과가 목표에 부합하는 잘 구조화된 문항의 개발로 연계될 수 있도록 단계적 후속 연구를 제시할 필요가 있다. 평가 실태를 조사한 연구는 실태 조사(4%), 교사 인식 및 관점(3.7%), 학생 인식(1.8%) 순으로 나타났다.

한편, 평가에 영향을 미치는 요인은 배경 변인(2.6%), 구성주의·사회 문화적 접근(1.1%) 등을 주제로 연구가 되었고 학습 환경·교실 문화(0.7%)와 성취(0.7%)에 대한 주제도 나타났다. 과정과 수행을 중심으로 한 평가로의 전환은 결과보다는 상황과 과정에 중점을 두면서 학습을 촉진하고 학생을 이해하려는 구성주의적 관점을 배경으로 하고 있다. 평가 패러다임 변화를 확인하는 연구(Jo & Ku, 2013; Noh, Yoon, & Kang, 2009; Joo, 2000)와 함께 학생들의 사회 문화적 경험을 반영한 평가 도구를 소개하는 연구(Jo & Ku, 2013)도 진행되었다. 특히 국제 성취도 평가에서 대규모 설문을 통해 학생들의 정의적 영역과 배경 변인을 조사한 결과가 발표되면서 평가에 영향을 미치는 다양한 변인에 대한 연구가 지속되고 있다(Kim, Kim, & Dong, 2015; Ku, Han, & Kim, 2015; Yum & Kang, 2011; Kim & Hong, 2010; Chung, 2009; Choi & Jeong, 2008; Sohn, 2008; Park, Park, & Kim, 2001).

과학 교육의 목표인 과학 지식과 개념(1.8%)을 다루는 평가 관련 연구 이외에 기타 과학 교육의 목표로서 STS(1.1%), 과학적 소양(1.1%), 과학의 본성(0.7%)을 다룬 연구들이 있고 이들을 측정하는 평가 도구의 개발 측면에서 접근했다.

Table 9. Periodical analysis by key words

범주	논문 수 (%) (N=273)						소계	총계	
	1986년~ 1990년	1991년~ 1995년	1996년~ 2000년	2001년~ 2005년	2006년~ 2010년	2011년~ 2015년			
교육 과정	제4차 교육과정	제5차 교육과정	제6차 교육과정	제7차 교육과정	2007 개정	2009 개정			
평가 관련 특이 사항		-대학수학능력시 험(1994~) -TIMSS1995	-TIMSS1999 -PISA2000	-수행 평가 실시(1999~) -TIMSS2003 -PISA2003	-PISA2006 -TIMSS2007 -PISA2009 -자유 탐구 신설	-성취 평가제 -TIMSS2011 -TIMSS2015			
개발	문항 (반응)	개발	-	1(0.4)	1(0.4)	4(0.4)	7(2.6)	14(5.1)	
		분석	-	1(0.4)	1(0.4)	-	8(2.9)	24(8.8)	
		유형	-	1(0.4)	-	1(0.4)	1(0.4)	7(2.6)	10(5.1)
	평가 준거	인지적	-	-	2(0.7)	2(0.7)	1(0.4)	2(0.7)	7(2.6)
		정의적	-	-	1(0.4)	1(0.4)	-	2(0.7)	4(1.5)
		기타	-	-	-	1(0.4)	-	-	1(0.4)
	도구	인지적	5(1.8)	5(1.8)	13(4.8)	7(2.6)	6(2.2)	6(2.2)	42(15.4)
		정의적	-	3(1.1)	6(2.2)	2(0.7)	5(1.8)	7(2.6)	23(8.4)
		기타	-	-	-	3(1.1)	2(0.7)	-	5(1.8)
실증적 측정도구	-	-	1(0.4)	-	3(1.1)	2(0.7)	6(2.2)		
평가 결과	국제학업성취도	-	-	1(0.4)	6(2.2)	4(1.5)	8(2.9)	19(7.0)	
	선발	-	1(0.4)	4(1.5)	3(1.1)	2(0.7)	2(0.7)	12(4.4)	
	학업 성취도	-	1(0.4)	1(0.4)	5(1.8)	3(1.1)	2(0.7)	12(4.4)	
	국가 수준 학업 성취도	-	1(0.4)	-	1(0.4)	2(0.7)	6(2.2)	10(3.7)	
	통계 이론 및 기술	-	1(0.4)	-	2(0.7)	1(0.4)	5(1.8)	9(3.3)	
	채점, 척도	-	1(0.4)	-	1(0.4)	2(0.7)	3(1.1)	7(2.6)	
	대학 수학능력시험	-	3(1.1)	-	-	1(0.4)	-	4(1.5)	
	교원 임용 시험	-	-	-	1(0.4)	-	3(1.1)	4(1.5)	
	수행형 평가	-	-	9(3.3)	14(5.1)	8(2.9)	4(1.5)	35(12.8)	
평가 방법	포트폴리오, 보고서	-	1(0.4)	6(2.2)	6(2.2)	-	-	13(4.8)	
	진단 평가, 형성 평가	-	-	5(1.8)	-	1(0.4)	4(1.5)	10(3.7)	
	수업 평가	-	-	-	1(0.4)	3(1.1)	3(1.1)	7(2.6)	
	자기 평가, 동료 평가	-	-	-	3(1.1)	1(0.4)	3(1.1)	7(2.6)	
	평가 방법 간 비교	-	-	1(0.4)	2(0.7)	-	-	3(1.1)	
	과학 교육 평가 방법	-	-	1(0.4)	-	-	-	1(0.4)	
교수 학습 과정	실험	3(1.1)	2(0.7)	9(3.3)	2(0.7)	1(0.4)	-	17(6.2)	
	인지 발달·개념 이해·학습 발달 과정, 지식 상태	-	1(0.4)	2(0.7)	3(1.1)	1(0.4)	8(2.9)	15(5.5)	
	상호작용, 피드백, 개념 검사	-	-	3(1.1)	2(0.7)	1(0.4)	2(0.7)	8(2.9)	
	교육 공학 (컴퓨터, 웹)	-	-	-	-	3(1.1)	5(1.8)	8(2.9)	
	교과서/과학관	-	-	-	1(0.4)	1(0.4)	3(1.1)	5(1.8)	
	교과 교육학 지식, 교사의 전문성	-	-	-	4(1.5)	1(0.4)	-	5(1.8)	
	개념도	-	-	2(0.7)	2(0.7)	1(0.4)	-	5(1.8)	
	탐구 학습, 개방적(자유) 탐구,	2(0.7)	-	-	1(0.4)	1(0.4)	-	4(1.5)	

Table 9. Periodical analysis by key words(Continued)

범주	논문 수 (%) (N=273)						소계	총계
	1986년~ 1990년	1991년~ 1995년	1996년~ 2000년	2001년~ 2005년	2006년~ 2010년	2011년~ 2015년		
교육과정	제4차 교육과정	제5차 교육과정	제6차 교육과정	제7차 교육과정	2007 개정	2009 개정		
평가 관련 특이 사항		- 대학수학능력 시험(1994~) -TIMSS1995	-TIMSS1999 -PISA2000	-수행 평가 실시(1999~) -TIMSS2003 -PISA2003	-PISA2006 -TIMSS2007 -PISA2009 -자유 탐구 신설	-성취 평가제 -TIMSS2011 -TIMSS2015		
	논증 과정, 논의	-	-	-	2(0.7)	1(0.4)	3(1.1)	
	탐구 과정, 탐구 수준	-	-	1(0.4)	1(0.4)	1(0.4)	3(1.1)	
	탐구 능력	5(1.8)	6(2.2)	8(2.9)	5(1.8)	8(2.9)	36(13.2)	
	문제 해결력	-	-	1(0.4)	2(0.7)	-	5(1.8)	
	지식(가설) 생성력	-	-	-	1(0.4)	3(1.1)	4(1.5)	
탐구 및 사고	시스템 사고	-	-	-	-	3(1.1)	3(1.1)	53 (19.4)
	논리적/과학적사 고/창의성	1(0.4)	1(0.4)	-	-	1(0.4)	3(1.1)	
	가설 평가 능력	-	-	-	-	2(0.7)	2(0.7)	
정의적 영역	태도 이외의 정의적 영역	-	-	4(1.5)	3(1.1)	6(2.2)	23(8.3)	43 (15.8)
	태도	-	4(1.5)	4(1.5)	5(1.8)	4(1.5)	20(7.3)	
	영재 교육 대상자	-	1(0.4)	5(1.8)	4(1.5)	4(1.5)	23(8.4)	
평가 대상	유아	-	-	2(0.7)	2(0.7)	1(0.4)	7(2.6)	39 (14.3)
	예비 교사	-	-	2(0.7)	-	2(0.7)	6(2.2)	
	특수 교육 대상자	-	-	-	-	-	3(1.1)	
	평가 기준, 준거, 성취 기준	-	1(0.4)	1(0.4)	3(1.1)	4(1.5)	16(5.9)	
평가 기준	과학 평가틀	1(0.4)	-	2(0.7)	1(0.4)	1(0.4)	8(2.9)	37 (13.6)
	교육 목표 분류틀	-	-	-	-	4(1.5)	7(2.6)	
	국가 교육과정	-	-	3	2(0.7)	1(0.4)	6(2.2)	
	평가 실태	-	-	3(1.1)	4(1.5)	1(0.4)	11(4.0)	
평가 실태	교사 인식, 관점	-	-	2(0.7)	3	2(0.7)	10(3.7)	26 (9.5)
	학생 인식	-	-	-	4(1.5)	-	5(1.8)	
	배경 변인	-	-	-	1(0.4)	4(1.5)	7(2.6)	
평가에 영향을 미치는 요인	구성주의, 사회 문화적 접근	-	-	1(0.4)	-	1(0.4)	3(1.1)	14 (5.1)
	학습 환경, 교실 문화	-	-	1(0.4)	-	-	2(0.7)	
	성차	-	-	-	2(0.7)	-	2(0.7)	
기타 과학 교육의 목표	STS	-	-	-	3(1.1)	-	3(1.1)	8 (2.9)
	과학적 소양	-	-	-	-	2(0.7)	3(1.1)	
	과학의 본성	-	-	1(0.4)	1(0.4)	-	2(0.7)	
개념	과학 지식(개념)	-	-	2(0.7)	1(0.4)	1(0.4)	5(1.8)	5(1.8)
	합계 (A)	17	36	112	126	123	587	
	중복된 논문 수 (B)	8	17	57	73	64	314	
	총계 (A-B)	9(3.3)	19(7.0)	55(20.1)	53(19.4)	59(21.6)	273(100)	

주 1: 두 가지 이상의 범주에 해당되는 경우, 중복 집계함.

주 2: 교육과정 시기는 초등학교 1학년의 시작 시기를 기준으로 함.

IV. 결론 및 제언

본 연구에서는 1986년부터 2015년까지 과학 교육 현장에서 실시된 평가에 대한 논문을 분석하여 30년간의 동향을 파악했고, 그 결과를 바탕으로 과학 교육 평가와 과학 교육 과정의 개선 방향을 제시하고, 향후 연구에 대한 시사점을 얻고자 했다. 한국교육학술정보원의 데이터베이스를 통해 선정된 273편의 논문에 대하여 연구 방법별, 연구 내용별 분석을 수행했다. 연구 방법론은 자료 수집 방법과 연구 목적, 연구 방법으로 다시 범주화하고, 연구 내용은 다루는 내용이 인지적 혹은 정의적 영역인지에 따라 분류하여 자료 수집 방법과 대상, 평가 도구별 비율을 비교했다. 또한 분석 논문의 주제어 12개를 추출하여 5년 주기로 논문의 빈도수를 분석하고, 시기별 교육과정과 평가 제도의 변화와 연계하여 그 의미를 고찰했다. 본 연구의 결론은 다음과 같다.

첫째, 과학 교육에서 평가 연구의 전체적 동향은 국가 교육 과정의 변화와 함께 평가와 관련한 대내외적 변화를 반영한 경우가 많았다. 평가 연구에 반영된 평가 관련 제도 변화는 대학수학능력시험의 시행, 국가 교육 과정 문서에서 평가 항목의 독립과 내용의 구체화, 수행평가 실시, 국제 학업 성취도 평가 참여와 국가 수준 학업 성취도 실시 등이다. 연구 대상이 정의적 영역보다 주로 인지적 영역의 평가에 많이 나타났지만, 인지적 영역과 함께 정의적 영역을 함께 다루려는 경향도 점차 늘어나고 있다. 국제 학업 성취도나 국가 수준 학업 성취도에서 수집된 표준화된 자료는 그 동안 인지적 영역에 국한되었던 관심을 정의적 영역과 다양한 배경 변인과의 관계에 주목하도록 환기시키는 역할을 했다. 연구에서 다루는 대상이 점차 구체적이고 세분화된 것도 두드러진 추세다. 일반적인 탐구 능력에서 교과와 연관된 개념이나 세부 능력과 사고로 구체화되었고, 태도를 중심으로 한 정의적 영역이 효능감, 동기, 흥미, 불안, 정서 등 다양한 분야로 확장되었다.

둘째, 논문에서 다루는 내용에 대하여 인지적 영역과 정의적 영역을 중심으로 분석했을 때, 두 영역 모두 질적 연구보다는 양적 연구에 더 많이 편중되는 경향이 있었다. 평가 도구는 인지적 영역의 경우, 지필 평가와 수행형 평가 등을 다양하게 시도한 반면, 정의적 영역의 경우, 거의 지필 평가만을 실시하고 있었다. 연구 측면에서 보면, 대상에 따라 어느 영역의 연구가 많이 이루어졌는지에 대한 경향성은 두드러지지 않았지만, 공통적으로 인지적 영역이 정의적 영역이나 혼합 연구의 비율보다 높았다. 평가 연구에서 질적 연구보다 양적 연구가 우세한 경향은 연구자들이 특정 현상에 대한 깊이 있는 지식보다 객관적 자료를 근거로 한 결과의 일반화에 주력하고 있음을 나타낸다. 연구에서 활용한 평가 도구를 분석한 결과, 인지적 영역은 지필 검사 중에서 선택형과 서답형을 위주로 한 객관적 평가를, 정의적 영역은 리커트형의 자기 보고식 평가를 주로 사용했다. 학습자에 대한 총체적 평가(Kim et al., 2014)가 가능한 수행형 평가는 수량화되지 않는 학습자의 태도와 정서나 느낌을 반영할 수 있지만, 실제 연구에서는 정의적 영역보다는 인지적 영역에서 주로 사용되었다. 이는 정의적 영역의 평가에서 측정하고 평가하고자 하는 특성을 분명히 규명하지 않은 채로 정도와 수준을 양화하고 척도화하는 과정에 어려움이 있음을 시사한다. 수행형 평가는 정의적 영역만을 위해 별도로 시행되기보다는 대부분 인지적 영역과 함께 수행되기 때문에 정의적 요소를

부차적인 것으로 여길 가능성이 있었다.

셋째, 연구를 수행한 목적에 따라 평가 관련 논문을 분류하면 이론을 규명하여 지식을 확장시키려는 기초 연구보다는 실제 맥락에서 실용적, 응용적 내용을 다루는 응용 연구와 프로그램의 효과와 연구 대상의 목표 도달을 측정하는 평가 연구가 많았다. 연구 방법론에 있어서 대부분 한 가지 이상의 연구 방법을 적용했는데, 평가 및 측정 도구와 문항을 개발하는 연구가 많이 발표되었다. 이들은 도구를 고안하여 검증하고 현장에 적용하는 과정에서 그 단계와 수준을 어떻게 조합하느냐에 따라 몇 가지 유형으로 분류되었고, 그 유형에 따라 연구 방법이 다르게 적용되었다. 한 예로 도구를 개발한 이후 특별한 검증 과정 없이 현장 적용한 경우 주로 관찰 연구가 많이 사용되었는데, 다수를 대상으로 한 통계적 방법을 사용하기 어려운 평가 도구를 개발한 경우에 해당한다. 평가와 측정 관련 주제를 다루는 연구에서는 연구 대상으로부터 얻은 정량적 자료를 해석하려는 과정이 자연스럽게 양적 연구에 대한 선호와 연결되는 경향이 있었다. 미시적 평가 과정과 상황에 대하여 자세히 알아보려는 연구 목적과 내용을 지닌 연구의 경우 양적 연구를 주로 채택하고, 보조적으로 질적 연구를 병행한 경우가 있었지만, 연구 목적과 대상이 속한 상황과 관계없이 정량적 자료만 근거하여 분석하는 연구도 있었다.

넷째, 분석 논문의 주제를 유사한 범주로 묶어 연구 동향을 분석한 결과, 평가 관련 개발 연구가 가장 많이 나타났고, 평가 결과, 평가 방법, 교수 학습 과정과 관련한 평가, 탐구 및 사고, 정의적 영역, 평가 대상, 평가 기준, 평가 실태, 평가에 영향을 미치는 요인, 기타 과학 교육의 목표와 개념 순으로 나타났다. 관련 연구 주제의 빈도는 평가 관련 제도와 쟁점의 변화와 더불어 변동하는 경향이 있었는데, 수행평가 도입이었던 제7차 교육과정 전후로 수행평가와 포트폴리오 및 보고서 등 평가 방법에 대한 연구가 증가한 것이 그 예다. 지난 30년 간 꾸준히 다루어진 연구 주제는 평가 도구 개발 연구, 탐구 능력에 대한 연구인데, 최근 내용이 보다 구체적이고 세분화되어 일반적 영역에서 세부 내용과 맥락에 중점을 두는 방향으로 전개되었다. 특히 평가 도구 개발이라는 측면에서 영역의 세분화는 평가하려는 구인을 분명히 하여 보다 정교한 도구의 개발을 가능하게 하는 바탕이 될 수 있지만, 학생의 능력을 총체적으로 이해하고 평가할 수 있도록 연구 결과를 수렴해서 다시 일반화해야 하는 과제가 남아있다.

이상의 과학 교육 평가 연구의 동향을 분석한 결과로부터 도출한 시사점과 제언은 다음과 같다.

첫째, 실증적 자료를 얻기 어려웠던 정의적 영역의 평가를 정량화하고 객관화할 수 있는 토대를 마련해야 하고, 실증적 성향의 연구자의 관심과 역량을 이 부분에 집중하여 전환할 필요가 있다. 그 동안 과학 교육 관련 평가 연구의 주제는 양적 자료를 얻기 쉬운 주제로 제한된 경향이 있었고, 특히 정의적 영역을 정량화하려는 시도는 상대적으로 적은 편이었다. 평가가 교육과정의 목표를 향상시키는 역할을 담당하고, 교육과정의 구현이 정의적 목표의 실현을 통하여 완성될 수 있다면, 연구자들은 인지적 영역과 정의적 영역 간 연구의 불균형을 보고한 본 연구의 분석 결과를 주목할 필요가 있다.

둘째, 정의적 평가의 정량화를 위해서는 관찰 행동이 분명하게 드러날 수 있도록 평가 요소를 분리해서 명확히 정의하는 분석 작업을 선행해야 한다. 정의적 성취를 활동이나 평가를 수행하는 과정에서 인지적 성취에 당연히 따라올 것으로 예상되는 부차적 성취로 여기는

연구의 전제를 개선할 필요가 있다. 예를 들어 총체적 평가로서 수행형 평가는 인지적 성취 뿐 아니라 정의적, 심동적 능력 등 다양한 영역의 성취가 배양될 수 있는 인큐베이터로서의 역할을 할 수 있지만, 특정 맥락 안에서 목표와 성취를 분명히 정의하고 측정하지 않으면 내부 사정을 알기 어려운 블랙박스 되어버리기 쉽다. 구체적 맥락에서 정의적 성취 목표와 성취 행동을 분명히 규명하려는 시도는 측정 도구의 개발과 정교화로 이어지고, 이는 실제 평가 상황에서 목표와 정의적 성취의 연결을 드러내 보여줌으로써 과학 교육자로서 하아금 인간의 사고와 느낌이 상호작용하는 과정을 가늠할 수 있도록 할 것이다.

셋째, 과학 교육에서 평가 관련 연구가 수많은 맥락과 연계되어 미세한 영역으로 세분화되어 발산되더라도 이를 다시 큰 틀로 수렴하여 묶을 수 있는 이론적 기초를 정립할 필요가 있다. 구체적 내용을 다루려는 경향성은 정의적 영역의 연구에서 뚜렷하게 나타나는데 지난 30년간 정의적 영역의 대표 영역으로 여겨졌던 태도 이외에 동기를 포함한 감성, 몰입, 자아 개념, 효능감, 신념 등 정의적 영역에 대한 연구가 꾸준히 발표되었고 약 10여 년 전부터 일반 태도 연구보다 더 많이 나타났다. 정의적 영역에서 다루는 이들 개념 중 일부는 태도를 구성하는 하위 개념으로(Abd-El-Khalick *et al.*, 2015; Unfried *et al.*, 2015) 재편성될 수도 있고, 태도와는 다른 독자적 영역으로 독립될 수도 있다. Krynowsky(1988)와 Koballa(1988)는 과학 교육에서 수행된 태도 연구의 약점은 명료하지 않은 개념과 구인을 명확히 하려는 이론적 기초의 부족이라고 주장했다. 이들 주장이 30여년이 지난 현재까지 유효한 시사점을 제공하는 한 측면에는 과학 교육에서 평가를 다루는 연구가 이론적 틀을 정립하기보다는 실용적 목적을 달성하는 데 치중했던 현실이 있다.

넷째, 연구 목적과 대상의 크기, 상황에 적합한 연구를 설계하기 위해 다양한 연구 방법론이 지지될 수 있는 토대를 마련할 필요가 있으며, 연구자는 통계 분석 결과가 현상을 설명하기에 적합한지 판단할 수 있는 전문성을 지녀야 한다. 연구자는 연구의 목적이 일반적 원리의 추구인지, 현상에 대한 심도 있는 이해인지에 따라 상황에 맞는 타당하고 적절한 연구 방법을 채택할 수 있어야 한다. 동시에 양적 접근 혹은 질적 접근을 연구에 주로 사용했는지의 여부와 상관없이 목적과 맥락에 맞도록 설계된 연구를 존중하는 연구 풍토가 조성되어야 한다.

마지막으로 과학 교육 평가가 교육 과정의 개선에 기여할 수 있는 지를 지속적으로 모니터링하면서 평가하는 기관과 체계를 정비해야 한다. 평가의 목표는 성취의 측정 뿐 아니라 학습의 향상이나 프로그램의 평가에도 있다. 평가의 대상 또한 학생 외에 수업이나 프로그램 및 교육 과정으로 다양하며 평가 상황도 교실이나 대규모 수준 등으로 각기 다르다. 교육 과정과 연계된 기존의 평가 기구와 체계가 주로 학생의 성취와 대규모 수준의 평가에 편중했다면, 추후에는 다양한 평가의 목표와 대상과 상황을 아우르는 균형 있는 체제로 정비할 필요가 있다. 과학 교육에서 평가 연구는 앞에서 논의한 바와 같이 국가 교육 과정이나 평가와 관련한 대내외적 사건들과 연결되어 때로는 교육 과정에 앞서 발전 방향을 제시하고, 사후에 중점 사항을 잘 반영했는지 평가함으로써 교육 과정의 개선에 기여해왔다. 연구자는 연구의 목표, 내용과 영역, 방법론이 어느 한 면에 치중되지 않은 균형 있는 평가를 지향함으로써 교육 과정의 핵심을 완성하고 성취할 수

있도록 관심과 역량을 기울여야 한다.

국문요약

이 연구는 최근 30년 동안 과학 교육에서 평가를 다룬 국내 논문을 분석하여 동향을 파악함으로써, 과학 교육 과정의 개선 방향을 제시하고 향후 연구에 대한 시사점을 얻고자 했다. 연구의 주요 목표는 과거 30년간 국내 학술지에 게재된 과학 교육 관련 평가 연구의 동향을 인지적 영역과 정의적 영역을 중심으로 분석하고, 연구 목적과 연구 방법론, 연구의 주제어별 특징을 분석하는 것이다. 이를 위해 한국교육학술정보원이 제공하는 데이터베이스를 통해 273편의 연구 논문을 선정했고, 대상, 교과, 연구 방법론, 연구 내용 등의 범주로 분류하여 각각의 논문을 분석했다. 연구의 동향에 영향을 미치는 요인을 알아보기 위해 같은 시기에 진행된 국가 교육 과정의 변화와 평가와 관련한 대내외적 사건을 연계하여 설명을 시도했다. 분석 결과, 과학 교육에서 평가 연구의 전체적 동향은 국가 교육 과정과 평가 관련 제도의 변화를 반영한 경우가 많았다. 연구의 대상이나 주제가 정의적 영역보다 인지적 영역에 편중해서 나타나는 경향이 있으나, 국가 수준 학업 성취도나 국제 학업 성취도 평가에서 얻은 표준화된 자료를 통해 정의적 영역과 다양한 배경 변인의 관계에 주목할 필요가 있음을 인식하게 되었다. 인지적 혹은 정의적 영역에 상관없이 두 영역 모두 양적 연구가 질적 연구보다 우세하게 나타났으며, 인지적 영역은 평가 도구로서 지필 평가와 수행형 평가 등을 다양하게 사용한 반면, 정의적 영역에서는 대부분 지필 평가만을 실시하고 있었다. 평가 연구의 목적은 주로 실제 상황에서 실용적이고 응용적인 내용을 다루는 응용 연구와 평가 연구가 많았고, 연구 방법론도 양적 자료를 다루는 통계를 기반으로 한 방법을 사용했다. 분석 논문의 주제를 유사한 범주로 묶어 연구 동향을 분석한 결과, 주제가 점점 일반적인 내용에서 구체화, 세분화되는 경향이 있었다. 이러한 경향은 그 동안 과학 교육 관련 평가의 문제점으로 지적되었던 개념의 명료화와 관련하여 평가 도구를 정교화할 수 있는 기초가 될 수 있을 것이다. 연구 결과를 바탕으로 정의적 영역의 평가에서 실증적 자료를 얻기 위한 방안 마련, 다양한 맥락으로 세분화된 평가를 수렴할 수 있는 이론적 기초 정립, 양적 자료에 편중된 연구의 균형을 추구하는 연구 풍토의 개선, 평가의 다양한 측면을 반영하여 완성된 교육과정으로 개선하는 기관과 체계의 정비 등을 시사점으로 제안했다.

주제어 : 과학 교육 관련 평가 연구, 연구 동향, 인지적 영역 평가, 정의적 영역 평가, 평가 도구, 연구 방법론

References

- Abd-El-Khalick, F., Summers, R., Said, Z., Wang, S., & Culbertson, M. (2015). Development and large-scale validation of an instrument to assess Arabic-speaking students' attitudes toward science. *International Journal of Science Education*, 37(16), 2637-2663.
- Bailey, R. C., & Bailey, K. G. (1974). Self-perceptions of scholastic ability at four grade levels. *The Journal of Generic Psychology*, 124, 197-212.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Choe, S., Ku, J., Kim, J., Park, S., Oh, E., Kim, J., & Baek, H. (2013). *Strategies for improving the affective characteristics of Korean students based on the results of PISA and TIMSS*. Seoul: Korea Institute for Curriculum and Evaluation.

- Chae, S. (1996). Towards a reconstruction of the social scientific methods: a synthesis of qualitative and quantitative methods. *The Journal of Educational Research*, 34(5), 281-297.
- Cho, J., Min, K., & Park, I. (2015). Consequential validity of National Achievement in elementary schools. *The Journal of Curriculum and Evaluation*, 18(1), 135-154.
- 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.
- Choi, W., & Jeong, E. (2008). Relationship between science achievement and student-related variable in National Assessment of Educational Achievement in 2006. *Journal of the Korean Association for Science Education*, 28(8), 848-859.
- Chung, B. (2009). A comparative study on the school differences in PISA 2006 science performance: Finland, Japan, Chinese Taipei, Hong Kong, and Korea. *Korean Journal of Comparative Education*, 19(2), 1-26.
- Chung, Y., & Choi, J. (2007). An assessment of the scientific literacy of secondary school students. *Journal of the Korean Association for Science Education*, 27(1), 9-17.
- Chung, W., Yoh, S., Park, J., & Kim, E. (1992). Evaluation of scientific inquiry skills for students of the special classroom(school) managed by firm. *Journal of Science Education Kyungpook National University*, 16, 1-17.
- Cowie, B. (2012). Focusing on the classroom: Assessment for learning. In Fraser, B. J., Tobin, K. G., & McRobbie, C. J. (Eds). *Second International Handbook of Science Education*, Dordrecht; New York: Springer.
- Creswell, J. W. (2008). *Educational research: qualitative, quantitative, and mixed methods approaches*(3rd ed.). CA: Sage Publications.
- Davis, G. A., Rimm, S. B., & Siegle, D. (2011). *Education of the gifted and talented*(6th ed.). Boston: Pearson.
- Dong, H., Lee, I., & Shin, J. (2013). Characteristics of science achievement of Korean students in TIMSS 2011 and NAEA 2011. *Journal of Educational Evaluation*, 26(5), 1091-1114.
- Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, 32, 268-291.
- Eisner, E. W. (1997). The promise and perils of alternative forms of data representation. *Educational Researcher*, 26(6), 4-10.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2015). *How to design and evaluate research in education* (9th ed.). New York: McGraw-Hill.
- Fraser, Tobin, & McRobbie. (2012). *Second international handbook of science education*, Dordrecht; New York: Springer.
- Gall, M. D., Gall, J. P., & Borg. W. R. (2007). *Educational research*(8th ed.). Boston: Pearson.
- Ha, M., & Lee, J. (2013). The item response, generalizability, and structural validity for the translation of Science Motivation Questionnaire II (SMQII). *Journal of Learner-Centered Curriculum and Instruction*, 13(5), 1-18.
- Holling, H., & Preckel, F. (2005). Self-estimates of intelligence methodological approaches and gender differences. *Personality and Individual Differences*, 38, 503-517.
- Huizinga, H. M., Crone, E. A., & Jansen, B. J. (2007). Decision-making in healthy children adolescents and adults explained by the use of increasingly complex proportional reasoning rules. *Developmental Science*, 10(6), 814-825.
- Hwang, S., Kim, H., Yoo, J., & Park, S. (2001). Ninth grader's self assessment of scientific process skills in open investigation. *Journal of the Korean Association for Science Education*, 21(3), 506-515.
- Jo, S., & Ku, J. (2013). An exploration in the measurement and application of attitude towards science based on social constructivism. *Journal of the Korean Association for Science Education*, 33(2), 466-477.
- Johnson, A. P. (2005). *A short guide to action research* (2nd ed.). Boston: Pearson.
- Joo, H. (2000). A little idea of concept mapping evaluation method in construction option. *The Journal of Elementary Education*, 13(2), 221-237.
- Kang, S., Park, E., Park, I., Yang, J., Lee, Y., Hong, H., & An, J. (2008). The analysis of the relationships between tests of the creative thinking and tests of the hypotheses generating ability. *Journal of Research Institute of Curriculum Instruction*, 12(1), 253-271.
- Kim, D., & Hong, H. (2010). Measuring the effects of school science curriculum on students science outcomes based on PISA 2006 and comparing them with those of Finland. *The Journal of Curriculum Studies*, 28(1), 87-112.
- Kim, K., Kim, S., Kim, M., Kim, S. H., Kang, M., Park, H., & Jung, S. (2009). Comparative analysis of curriculum and achievement characteristics between Korea and high performing countries in PISA & TIMSS. Seoul: Korea Institute for Curriculum and Evaluation.
- Kim, H. (2012). A research analysis for an alternative paradigm in christian education with a focus on research methodology. *A Journal of Christian Education in Korea*, 29, 347-377.
- Kim, J., Kim, S., & Dong, H. (2015). International comparison of cognitive attributes using analysis on science results at TIMSS 2011 based on the cognitive diagnostic theory. *Journal of the Korean Association for Science Education*, 35(2), 267-275.
- Kim, M. (2002). The effect of portfolio assessment on the scientific investigation ability and science-related attitude of students in elementary school. *Journal of Educational Evaluation*, 15(1), 87-113.
- Kim, R., Kim, G., & Kwon, N. (2012). Trends in research design and methods: Research on elementary and secondary mathematics curriculum. *School Mathematics*, 14(3), 395-408.
- Kim, S., & Choi, T. (2008). *Research methodology in education*. Seoul: Hakjisa.
- Kim, Y., Park, Y., Park, H., Shin, D., Jung, J., & Song, S. (2014). *World of science education*. Seoul: Book's hill.
- Kim, Y., Yoon, K., & Kwon, D. (2010). Analysis of summative evaluation objectives in middle school biology based on Bloom's revised taxonomy of educational objectives. *Journal of Science Education*, 34(1), 164-174.
- Koballa, T. R. (1988). Attitude and related concepts in science education. *Science Education*, 72(2), 115-126.
- Krynowsky, B. (1988). Problems in assessing student attitude in science education: A partial solution. *Science Education*, 72(4), 575-584.
- Ku, J., Han, J., & Kim, S. (2015). Effects of educational context variables on gender difference in science achievement among top performing countries of PISA 2012. *Journal of Educational Evaluation*, 28(5), 1381-1400.
- Lawrenz, F., & Thao, M. (2014). Review of science education program evaluation. In Lederman, N. G., & Abell, S. K. (Eds). *Handbook of Research on Science Education*, volume II. New York: Routledge.
- Lederman, N. G., Abell, S. K. (2014). *Handbook of research on science education*, volume II, New York: Routledge.
- 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, H., Choi, K., & Nam, J. (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.
- Lee, Y., & Kim, E. (2013). *Human characteristics and education*. Hakjisa: Seoul.
- Lee, J., Lee, I., & Kwon, Y. (2010). The development of the brain-based analysis framework for the evaluation of teaching-learning program in science. *Journal of the Korean Association for Science Education*, 30(5), 647-667.
- Lee, S., Kim, J., & Jung, K. (2006). Potential of a self-report measure for intelligence. *The Korean Journal of Educational Psychology*, 20(4), 931-953.
- Leem, Y., Kim, Y. (2013). A historical study on the Korean Science Curriculum for the elementary and secondary schools. *Biology Education*, 41(3), 483-503.
- Lim, J., Kang, S., Kong, Y., Choi, B., Nam, J. (2004). The development of instrument to assess high school students' views on Science-Technology-Society. *Journal of the Korean Association for Science Education*, 24(6), 1143-1157.
- Liu, X. (2012). Developing measurement instruments for science education research. In Fraser, B. J., Tobin, K. G., & McRobbie, C. J. (Eds). *Second International Handbook of Science Education*, Dordrecht; New York: Springer.
- Lyon, E. (2011). Beliefs, practices, and reflection: Exploring a science teacher's classroom assessment through the assessment triangle model. *Journal of Science Teacher Education*, 22, 417-435.
- Miller, D. C. (1991). *Handbook of research design and social measurement*. Newbury Park, CA: Sage Publications.
- Ministry of Culture and Education. (1981). *Elementary education curriculum*. Notification No. 1981-442 of MOCE. Seoul: MOCE.
- Ministry of Culture and Education. (1987). *Elementary education curriculum*. Notification No. 87-9 of MOCE. Seoul: MOCE.
- Ministry of Education. (1992). *Elementary education curriculum*. Notification No. 1992-16 of MOE. Seoul: MOE.
- Ministry of Education. (1997a). *Elementary education curriculum*. Notification No. 1997-15 of MOE. Seoul: MOE.

- Ministry of Education. (1997b). High school science curriculum reference book. Notification No. 1997-15 of MOE. Seoul: MOE.
- Ministry of Education and Human Resources Development. (2007). Elementary education curriculum. Notification No. 2007-79 of MOEHRD. Seoul: MOEHRD.
- Ministry of Education Science, and Technology. (2011). Science curriculum. Notification No. 2011-361 of MOEST. Seoul: MOEST.
- Nam, J., Choi, J., Kong, Y., Moon, S., & Lee, S. (2004). The effects of feedback types in self assessment on the students's science concept understanding and science-related attitudes in the middle school science. *Journal of the Korean Association for Science Education*, 24(3), 646-658.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D. C.: National Academy Press.
- Neuman, W. L. (2011). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Needham Heights, MA: Allyn & Bacon.
- 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 the Korean Society of Elementary Science Education*, 28(3), 352-360.
- Orpwood, G. (2001). The role of assessment in science curriculum reform. *Assessment in Education*, 8(2), 135-151.
- 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 the Research Institute of Curriculum Instruction*, 12(1), 55-65.
- Park, D., Park, C., & Kim, S. (2001). The effect of school and school and student level background variables on math and science achievements in middle schools. *Journal of Educational Evaluation*, 14(1), 127-149.
- Park, H., Bea, J., & Jo, K. (2012). Analysis of instructional and evaluational objectives in Chemistry I textbooks. *Journal of the Korean Chemical Society*, 56(4), 491-499.
- Park, K. (2012). Analysis of research trends on gifted education in Korea. *Journal of Gifted/Talented Education*, 22(4), 823-840.
- Park, S., & Yoo, J. (1999). Science performance assessment. *The Korea Association of Yeolin Education*, 7(1), 247-262.
- Pellegrino, J. W. (2012). Assessment of science learning: Living in interesting times. *Journal of Research in Science Teaching*, 49(6), 831-841.
- Seo, Y., Kim, H., & Chae, H. (2010). Analysis of the end-of-chapter questions in Chemistry II according to revised Bloom's taxonomy of educational objectives. *Journal of the Korean Chemical Society*, 54(3), 329-337.
- Shin, J., & Choi, A. (2014). Trends in research studies on scientific argument and writing in Korea. *Journal of the Korean Association for Science Education*, 34(2), 107-122.
- Sohn, W. (2008). Multi-level factors influencing the affective domain of Korean students: Results from PISA 2006. *Journal of Educational Evaluation*, 21(4), 81-105.
- Song, H. (2010). Development of a self-reported executive function rating scale for the Korean high school students: a preliminary study. *The Korean Journal of Clinical Psychology*, 29(1), 109-124.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M. (1971). *Educational evaluation and decision-making in education*. Itasca, IL: Peacock Publishers Incorporated.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: Wiley.
- Sung, T., & Si, K. (2015). *Research methodology* (2nd ed). Hakjisa: Seoul.
- Unfried, A., Faber, M., Stanhope, D. S., & Wiebe, E. (2015). The development and validation of a measure of student attitudes toward science, technology, engineering, and math (S-STEM). *Journal of Psychoeducational Assessment*, 33(7), 622-639.
- Wang, L., Huettel, S., & De Bellis, M. D. (2008). Neural substrates for processing task-irrelevant sad images in adolescents. *Developmental Science*, 11(1), 23-32.
- Yang, I., Na, J., Lim, S., & Choi, H. (2008). An analysis of elementary schools's science test items by Kopfer's taxonomy of educational objectives: Focusing on the first term of the 5th grade. *Journal of Korean Elementary Science Education*, 27(3), 221-232.
- Yum, S., & Kang, D. (2011). Influence of curricular context factors within student- and school-levels on the 2006 PISA science achievements of Korean students. *Journal of the Research Institute of Curriculum Instruction*, 15(2), 281-304.