

교육과정평가연구

The Journal of Curriculum and Evaluation

2021, Vol. 24, No. 1, pp. 173-201

DOI: <https://doi.org/10.29221/jce.2021.24.1.173>

국내 교사 평가 전문성 기준과 측정도구 분석

우필희 (대구남부초등학교 교사)*

송명섭 (대구교육대학교 교수)

송주범 (경북대학교 교수)

요약

학습 결과의 평가를 넘어 교수와 학습의 연결고리로서의 평가가 강조됨에 따라 평가의 질에 영향을 미치는 교사의 평가 전문성에 대한 관심이 높아지고 있다. 본 연구는 국내 교사 평가 전문성 기준과 측정도구 24편을 대상으로 하였으며, 평가 목적, 평가 상황, 평가 측면, 측정 척도의 특성으로 평가 전문성의 내용과 발달을 분석하였다. 평가 전문성 기준과 측정도구에 나타난 평가 전문성의 내용은 평가 목적과 상황이 명확하게 규정되지 않은 것이 대부분이고, 평가 과정과 관련된 지식과 기능에 대한 것이 많았다. 그러나 교육과정 개정의 영향으로 평가 전문성 측정도구에서는 형성평가를 목적으로 한 문항과 교실문화, 협력에 관한 문항의 비율이 다소 증가하였다. 대부분의 평가 전문성 측정도구는 리커트 척도의 평균으로 전문성 수준과 발달정도를 양적으로 측정하였으나 교사가 평가 전문성의 각 요소를 어떻게 발달시켜야 하는지에 대한 방향성을 제시하기에 한계가 있었다. 이 결과를 바탕으로 교사에게 필요한 평가 전문성의 내용과 발달에 대한 연구의 시사점을 제안하였다.

주제어 : 평가 전문성, 평가 전문성 내용, 평가 전문성 발달

* 제1저자 및 교신저자, woopilhee@gmail.com

I. 서 론

교사의 평가 전문성은 교사가 학습자와 관련된 교육적 결정을 하는데 필요한 역량이다. 객관적 지식을 습득하는 것을 학습으로 보는 전통주의 관점과 달리, 학습은 학습자 스스로 새로운 정보를 기존의 지식과 연결하여 개인적 의미를 생산하는 것(McMillan, 2014)으로 보는 구성주의 관점은 교수활동과 평가활동의 의미 전환을 가져왔다. 교수활동은 절대적 지식을 학습자에게 전달하는 것이 아니라 학습자의 개인적 의미 생산을 돕는 것이 되었으며, 그 결과 평가는 학습결과로서 습득한 지식을 정확하게 측정하는 것보다는 학습과정에서 겪는 학생의 어려움을 진단하고 적절한 피드백을 제공하는 데에 초점을 두고 있다. 교수와 학습의 연결고리(Wiliam, 2011)로서 평가의 강조는 이에 적합한 평가 전문성을 교사가 갖추어야 함을 의미한다(박혜영, 이명애, 이명진, 2019).

최근 교사의 평가 전문성 개념은 교사가 갖추어야 할 평가 지식이나 기능(Stiggins, 1991; Popham, 2009)으로 보던 기존의 관점을 넘어서고 있다. 표준화 검사와 같은 평가 상황은 공정한 측정을 위해 변인을 통제하는 것에 비해, 교수학습 상황은 다양한 변인들이 역동적으로 상호작용하므로 일반화된 평가 지식과 기능을 그대로 적용하기 어렵기 때문이다. 이에 평가 지식과 기능뿐만 아니라 교사의 평가 신념과 평가 경험, 평가에 대한 자신감, 교과 내용, 학습 상황 등의 변인이 상호작용하는 맥락에서 학생과 평가의 의미에 대해 의사소통할 수 있는 역량을 평가 전문성으로 바라보고 있다(성태제, 임현정, 2014; Adie, 2013; Willis, Adie, & Klenowski, 2013; Xu & Brown, 2016; Looney et al., 2018; DeLuca et al., 2019). 이러한 평가 전문성 개념의 확장으로 평가 전문성 기준과 측정방법, 평가 전문성 개발 프로그램 등의 평가 전문성 연구 방향도 바뀌었다. 즉, 이전의 평가 전문성 기준(AFT, NCME, & NEA, 1990)과 측정도구(Plake, Impara, & Fager, 1993; Mertler, 2003)의 내용은 평가 실행에 필요한 지식과 기능 중심이었으나 새로운 평가 전문성 기준(Brookhart, 2011; JCSEE, 2015)과 측정도구(DeLuca, LaPointe-McEwan, & Luhanga, 2016a)에는 학습을 위한 평가와 성취기준 평가 등이 반영되어 있다. 평가 전문성을 향상시키기 위한 프로그램도 평가 지식과 기능을 전달하기보다 자신의 평가 실행에 대한 지속적 성찰과 공동체 속 협력을 지향하고 있다(Xu & Brown, 2016).

국내의 경우, 2000년 이후 교사의 평가 전문성에 대한 논의가 지속적으로 이루어져왔다. AFT, NCME, & NEA(1990)의 평가 전문성 기준을 기반으로 개발된 학생평가 전문성 기준(이인제 외, 2004b~2004g; 김경희 외, 2006)은 교사가 개발하고 향상시켜야 할 평가 전문성의 지침이 되었다. 이후 이러한 기준들을 이론적 근거로 하여 평가 전문성 측정도구의 개발(송미영, 김경희, 2006; 이수정, 김민정, 2019), 평가 전문성 실태 조사(박정, 2013; 강훈식, 강석진, 2015), 평가 전문성 향상 프로그램의 개발(김선희, 2006; 2012)이 지속적으로 이루어졌다. 그러나 이러한 연구들은 주로 평가 도구 개발 및 수행에 초점을 둔 평가 전문성을 논의하고 있어 평가 계획 및 피드백에 대한 고려가 미비하다(박지현 외, 2018)는 한계가 있다.

최근 이러한 한계를 고려하여 성취평가제, 과정 중심 평가와 같은 정책, 학습을 위한 평가 등의

평가 동향을 반영한 평가 전문성 척도 개발(임은영, 2019)과 평가 전문성 신장을 위한 연구(박지현 외, 2018)가 수행되었다. 이와 같은 노력에도 불구하고 교육 현장에서 실행이 어려운 점, 객관성과 공정성에 대한 교사 평가 전문성의 문제(신혜진, 안소연, 김유원, 2017) 등이 제기되고 있다. 이는 비단 우리만의 문제가 아니라 국외의 연구(DeLuca & Klinger, 2010; Bennett, 2011; Xu & Brown, 2016)에서도 제기되는 문제들이다. 일부 연구(박정, 2017; DeLuca et al., 2019)에서는 평가 역할에 대한 교사의 근본적인 인식 변화를 통해 이러한 문제를 해결할 수 있다고 주장하고 있다. 이는 지금까지의 연구 방향이 교사 인식의 변화를 이끌어 내는 데 한계가 있으며 앞으로의 연구 내용과 방향에 대한 점검이 필요함을 의미한다. ‘교수-학습-평가’의 연계를 강조하는 현시점에서 교사가 실천하기 바라는 평가 전문성에 대한 이해와 지금까지 연구에서 바라보는 평가 전문성에 대한 이해가 서로 일치하는지 살펴볼 필요가 있다. 교사의 평가 전문성이 무엇이고 어떻게 발달하는지 살펴봄으로써 평가 전문성 이해의 특징을 파악할 수 있고, 이에 대한 성찰은 평가 전문성 발달을 위한 방향을 탐색하는 데에 도움이 될 것이다. 특히 평가 전문성 기준과 측정도구는 교사의 평가 전문성 수준을 측정할 뿐만 아니라 전문성의 개발 방향을 제시한다는 점에서 중요하다.

이에 본 연구에서는 국내에서 연구 보고된 교사의 평가 전문성 기준과 측정도구에 드러난 평가 전문성의 내용과 발달을 분석하고자 한다. 이를 위한 구체적인 연구 문제는 다음과 같다.

첫째, 국내에서 연구 보고된 교사 평가 전문성 기준과 측정도구의 내용은 어떠한가?

둘째, 국내에서 연구 보고된 교사 평가 전문성 측정도구에 드러난 평가 전문성의 발달은 어떠한가?

II. 이론적 배경

1. 교사 평가 전문성의 개념

평가 전문성(Assessment Literacy)이란 평가과정에 대한 교사의 이해뿐만 아니라 평가과제를 디자인하고, 학습의 질을 판단하기 위한 준거를 개발하고, 평가를 통해 수집한 정보를 해석하여 이를 근거로 피드백을 하는 교사의 능력(Hay & Penney, 2013)을 나타내는 용어이다. 이는 Stiggins(1991)에 의해 처음 사용되었으며, 연구 초점에 따라 ‘평가 지식과 기능(assessment knowledge and skills, Brookhart, 2011)’, ‘학생평가 전문성(teacher competence in the educational assessment of students, AFT, NCME, & NEA, 1990)’, ‘평가 전문성(assessment expertise, Abell & Siegel, 2011; Lyon, 2013)’, ‘평가 정체성(assessment identity, Adie, 2013; Looney et al., 2018)’ 등으로 표현되기도 한다(Willis, Adie, & Klenowski, 2013). 국내에서 평가 전문성은 교사가 평가를 시행할 경우에 필수적으로 요구되는 기초적인 지식과

기능으로서의 평가소양(한국교육평가학회, 2004)을 의미하며, ‘평가 문식성(주세형, 2011)’, ‘학생 평가 전문성(이인제 외, 2004b)’, ‘학생 평가 역량(임은영, 2019)’ 등으로 표현되기도 한다.

평가 전문성에 대한 논의는 AFT, NCME, & NEA(1990)의 교사의 학생평가 전문성 기준 연구에서 시작되었다. 평가 전문성 기준은 ‘평가 방법의 선정’, ‘평가 방법의 개발’, ‘평가의 실시, 채점 및 결과 해석’, ‘평가 결과의 활용’, ‘성적 부여’, ‘결과에 관한 의사소통’, ‘윤리적 문제 인식’의 7개 기준으로 구성되어 교사가 갖추어야 할 평가 전문성의 요소를 구체적으로 제시하고 있다. 이는 이후 TALQ(Plake, Impara, & Fager, 1993), API(Zhang & Burry-stock, 1997), ALI(Campbell, Murphy, & Holt, 2002), CALI(Mertler, 2003), the revised ALI(Mertler & Campbell, 2005)와 같은 평가 전문성 측정도구 및 평가 전문성 발달을 위한 교사교육(Mertler, 2009)의 토대가 되었다는 점에서 의의가 있다(Xu & Brown, 2016). 그러나 이상의 연구는 두 가지 측면에서 한계가 있다. 하나는 총괄 평가의 구성, 관리, 활용과 관련된 원리와 지식을 중심으로 구성되어 있어(DeLuca, LaPointe-McEwan, & Luhanga, 2016b) 21세기 평가에서 강조하는 형성 평가와 성취기준 평가의 개념, 평가 과정에서의 학생 참여 부분이 포함되어 있지 않다는 점이다(Brookhart, 2011). 또 다른 하나는 객관적인 기준과 척도로 평가 전문성의 습득 여부를 측정함으로써 탈맥락적일 뿐만 아니라 실제 평가 상황에서 평가 전문성의 여러 요소들이 통합적으로 작용한다는 점을 간과하고 있다는 점이다(Blömeke, Gustafsson, & Shavelson, 2015; DeLuca et al., 2019).

이러한 한계를 고려하여 교사와 학생 모두에게 학습 상황과 이후 교수학습과정을 안내하는 데 활용하는 형성평가 및 교사의 평가에 대한 성찰과 관련된 내용을 포함하는 평가 전문성 기준(Brookhart, 2011; JCSEE, 2015)이 개발되었다. 그러나 이 연구들은 학생 중심의 평가에 초점을 두고 있음에도 불구하고 학년, 교과, 학생 수 등의 교실 상황이 각기 다르고, 교사의 교수 경험, 신념 등이 실행에 어떤 영향을 미치는 지에 대한 고려가 부족하다.

이에 대한 대안으로 평가를 개인의 경험을 바탕으로 의미와 정체성을 구성해 나가는 문화적 행위이라고 보는 사회문화적 관점에서의 연구가 시작되고 있다(Adie, 2013; Looney et al., 2018). 즉, 교사의 평가 전문성은 미리 정의된 능력이 아니라 교사와 학생, 교육 맥락 사이의 관계에 따라 역동적인 교사학습이다. 평가에 대한 지식과 기능을 넘어 평가를 실행할 때의 자신감, 평가자로서의 교사 존재에 대한 신념까지 평가 전문성에 포함한다는 점에서 이론과 실행, 철학이 통합되어 있다(Pastore & Andrade, 2019). 평가의 방향과 사회적 요구를 반영하여 평가 전문성의 개념이 확장된 것은 다음과 같은 측면에서 의의가 있다.

첫째, ‘학습 향상’이라는 평가의 목적과 ‘교수-학습-평가’의 연계를 강조한다. 평가의 목적을 선발과 배치에 두는 평가 전문성은 공정한 측정에 관한 지식과 기능에 중점을 두고 있는 데 비해, 학습 향상을 목적으로 하는 평가 전문성은 다차원적인 평가 맥락에서 학습정보를 수집, 해석, 이용하여 교수학습의 개선을 위한 의사결정을 하는 능력에 중점을 둔다(Brookhart, 2011; JCSEE, 2015; Xu & Brown, 2016; Looney et al., 2018; Herppich et al., 2018). 학습은 학생과 환경의 상호작용을 통해 학생이 스스로 지식을 구성하는 과정이므로 학생이 처한 환경에 따라 개개인의 학습 과정은 다를 수밖에 없다. 교사가 평가 전문성을 갖춘다면 학생과의 의사소통을

바탕으로 개개인의 학습 과정을 이해하고 이후 교수학습의 방향을 결정할 수 있을 것이다.

둘째, 학생을 평가 대상이 아닌 평가 주체로 인식한다(박정, 2017; Pastore & Andrade, 2019). 이는 학생이 평가 과정에 단순히 참여하는 것을 넘어 학습 주체로서 평가에 참여하는 것을 의미한다. 학생이 학습 주체가 된다는 것은 학습 정보의 원천으로서 평가에 필요한 정보를 제공할 뿐만 아니라 동료 또는 교사에게 받은 피드백이나 자기평가 결과를 활용하여 지식을 구성해 나가고 학습 과정 및 결과를 성찰함으로 이후 학습을 결정하는 것이다. 그러므로 평가 전문성을 갖춘 교사는 학습 및 평가 과정에서 학습자에게 발언권, 선택권, 의사결정권을 부여함으로써 학생의 주체성을 향상시킬 수 있어야 한다.

셋째, 사회문화적인 평가 맥락을 고려하여 평가 전문성은 고정된 지식 기반을 습득하는 것을 넘어 교사 학습을 통해 지속적으로 성장 발전하는 것으로 인식한다. 왜냐하면 평가 맥락은 학습 상황, 교과 내용, 교사의 평가 경험과 신념, 학교 여건 등 다양한 변인들이 역동적으로 상호작용하는 과정이라서 일반화된 지식과 기능을 습득하여 이를 기계적으로 적용하는 것만으로는 교수학습이 향상되기 어렵기 때문이다. 그러므로 사회적 맥락을 고려하여 교사가 평가를 실행하고 자신의 실행과 신념에 대한 성찰과 적용을 반복하여(Timperley, 2011) 평가 전문성을 발달시키는 것이 필요하다.

2. 교사 평가 전문성 연구의 동향

교사 평가 전문성 연구는 평가 전문성의 필요성에 대한 시대적 요구와 교육 환경의 변화에 맞추어 지속적으로 이루어져 왔으며(Stiggins, 2004; Popham, 2013; Gotch & French, 2014), 연구 흐름은 크게 교사가 갖추어야 할 평가 전문성이 무엇인지 구명하는 연구, 교사의 평가 전문성을 측정하여 실태 조사 및 관련 변인을 탐색하는 연구, 평가 전문성을 어떻게 발달시킬 것인지 그 방법을 찾는 연구로 나눌 수 있다.

먼저 교사가 갖추어야 할 평가 전문성이 무엇인지 정의하고, 그 하위 영역과 요소를 구명하는 연구들이 시작되었다. 국내 연구의 시작은 2004년 한국교육과정평가원의 연구로, 학생평가 전문성을 ‘바람직한 학생 평가를 실천하기 위해 교사가 갖추어야 할 지식, 기능, 태도가 습득된 상태(이인제 외, 2004b, p.10)’로 정의내리고, 바람직한 평가를 실천하기 위해 반드시 알아야 할 지식인 내용기준 17개와 평가 상황에서 실제 할 수 있어야 할 능력인 수행기준 16개를 개발하였다. 이를 기반으로 하여 국어, 사회, 수학, 과학, 영어의 교과별 학생평가 전문성 기준이 개발되었다. 후속 연구에서는 학생평가 전문성을 ‘학생의 학습과 성취에 관한 평가 정보를 수집하고 해석하여 활용할 수 있는 능력(김수동 외, 2005, p.7)’으로 재정의 내리고, ‘평가 방법 선정’, ‘평가도구 개발’, ‘평가의 실시·채점·성적부여’, ‘평가 결과 분석·해석·활용·의사소통’, ‘평가의 윤리적 인식’의 5가지 영역의 평가 기준을 제시하였다(김경희 외, 2006). 개발된 평가 전문성 기준은 평가 목적, 평가 과정, 평가 결과에 관한 의사소통, 공정성에 대한 내용을 주로 다룬 AFT, NCME, & NEA(1990)의 평가 전문성 기준과 유사하며(박정, 2013), 평가 도구의 개발과 수행에 초점을 두고 있다. 2000년 이후의 국외 평가 전문성 기준(Brookhart, 2011; JCSEE, 2015)에서는 형성평가, 자기평가 등

학습을 위한 평가가 중요한 개념으로 부각되었으며, 이와 더불어 교사교육 및 지원에 관한 내용도 나타나기 시작하였다(DeLuca, LaPointe -McEwan, & Luhanga, 2016b). 또한 교과에 따라 평가 맥락이 달라짐을 고려하여 교과 특성을 살린 평가 전문성 기준이나 요소에 관한 연구(Abell & Siegel, 2011; Taylor, 2013)도 이루어졌다. 한편, 2012년 이후 도입된 성취평가제, 과정중심평가 등의 평가 정책은 교실평가의 내실화, 평가 계획 및 피드백, 수업과 연계한 평가 실행 능력을 갖춘 교사를 요구하고 있으며, 이러한 사회적 요구를 반영한 평가 전문성 기준이 필요하다고 본다.

국내의 평가 전문성 측정도구는 학생평가 전문성 기준(김정희 외, 2006)을 기반으로 개발되었다. 이 중 리커트 척도의 60개 문항으로 구성된 송미영, 김신영(2007)의 진단도구는 교사들의 학생평가 전문성 실태 조사 및 평가 전문성에 영향을 미치는 변인 탐색(예. 김신영, 2007; 박정, 2013; 강훈식, 강석진, 2015)에 많이 활용되었다. 학교급, 교과에 따라 기존의 도구를 수정·보완하거나(예. 하유라, 손원숙, 2013; 이수정, 김민정, 2019), 최근 강조되는 학생평가 동향을 반영한 측정 도구(박지현 외, 2018; 김유정 외, 2019; 임은영, 2019)도 개발되었다. 국외 평가 전문성 측정도구의 내용은 총괄평가, 표준화평가와 관련된 평가 과정에 관한 것이 많았으며(DeLuca, LaPointe-McEwan, & Luhanga, 2016b), 측정도구의 내용타당도 및 내적합치도 등 심리측정 증거가 약하여 평가 전문성 측정 결과를 교사평가 관점에서 해석 및 활용하는 것에 대한 문제를 제기한 연구(Gotch & French, 2014)도 있다. 평가 지식과 기능에서 평가 인식, 평가 맥락까지 확장된 평가 전문성 개념을 고려한 평가 전문성 측정도구(Brown, 2004; DeLuca, LaPointe-McEwan, & Luhanga, 2016a)는 상황에 따른 교사의 평가관과 평가 실행 수준, 전문성 개발을 위한 교사교육의 내용과 방법에 대한 요구를 측정할 수 있어 현재의 평가 전문성 수준뿐만 아니라 교사교육의 방향을 제시한다는 점에서 의의가 있다.

평가 전문성의 발달을 위해 평가 전문성 발달 과정 탐색(Lyon, 2013, Edwards, 2017), 교사교육과 다양한 매개 요인의 관계(김신영, 2007; 김신영, 송미영, 2008; 강현영 외, 2018; Koh, 2011; Lyon, 2013; Siegel & Wisseher, 2011), 맥락적 요인이 교사의 평가 전문성 발달에 미치는 영향(Xu & Liu, 2009; Adie, 2013), 학습공동체와 평가 전문성 발달과의 관계 탐색(정민수, 김연경, 부재울, 2019; Wyatt-Smith, Klenowski, & Gunn, 2010; Willis, Adie, & Klenowski, 2013)과 관련된 연구가 이루어졌다. 교사의 평가 전문성의 발달을 평가 지식의 축적이 아니라 평가 맥락에 적합한 역량이 발달하는 것으로 보는 연구(Adie, 2013)도 있었다. 평가 전문성의 효과적인 발달을 위해 참여자의 수준과 요구를 고려한 수요자 중심의 프로그램이 필요하다(강현영 외, 2018)는 연구 결과는 평가 전문성이 구체적으로 어떻게 발달되는지, 이러한 발달단계를 반영한 프로그램의 내용과 방법은 어떠해야 하는지에 대한 연구의 필요성을 제시하고 있다.

III. 연구 방법

1. 연구 대상

2019년 12월까지 국내에서 발표된 논문이나 연구 보고서 중 ‘평가 전문성’, ‘평가 문식성’, ‘평가 문해력’, ‘평가 역량’, ‘assessment literacy’, ‘assessment expertise’, ‘assessment competence’가 제목이나 주제어에 포함되고, 교사 평가 전문성 기준과 측정도구를 확인할 수 있는 110편의 논문을 선정하였다. 이들 중 연구 목적에 부합되는 논문을 선정하기 위해 과학교육전문가와의 협의를 통하여 다음과 같은 선정 기준을 정하였다. 첫째, KCI 등재 논문, 박사학위논문, 국가기관의 연구보고서로 한정한다. 단, 동일 저자의 연구 내용이 학술지나 학위논문, 국가기관의 연구보고서와 중복될 경우 학술지 논문만 선정하되, 국가기관의 연구보고서 중 일부 내용만 정리하여 학술지 논문으로 보고한 경우에는 국가기관의 연구보고서로 한다. 둘째, 평가 전문성의 일부 요소만 연구한 논문은 대상에서 제외한다. 셋째, 평가 전문성의 기준이나 평가 전문성 측정도구 개발을 목적으로 한 연구가 아닐지라도 평가 전문성의 실태나 향상 정도를 측정할 수 있는 기준이나 도구가 명시되어 평가 전문성의 내용과 발달을 분석할 수 있는 논문은 포함한다. 넷째, 기존의 개발된 평가 전문성 측정도구를 활용한 경우에는 평가 전문성의 내용과 발달을 바라보는 관점이 측정도구를 개발한 연구와 동일한 것으로 간주하여 동일한 평가 전문성 측정도구를 활용한 연구들은 최초의 평가 전문성 측정도구 개발 논문으로 한정한다. 이 기준에 따라 110편의 논문 중 학술지 논문 11편, 박사학위 논문 3편, 연구보고서 10편, 총 24편을 연구대상으로 하였다(표2,3).

2. 분석 방법

가. 분석 틀

평가 전문성을 분석하기 위해 선행연구(Looney et al., 2018; DeLuca, LaPointe-McEwan, & Luhanga, 2016b)를 바탕으로 ‘평가 목적’, ‘평가 상황’, ‘평가 측면’, ‘측정 척도’의 4가지 범주와 각 범주별 내용 요소로 이루어진 분석틀을 재구성하였다(표1).

‘평가 목적’ 범주는 Looney 외(2018)의 분석틀을 따라 ‘형성평가’, ‘총괄평가’, ‘진단평가’, ‘명시되지 않음’의 4개의 내용 요소로 구성하였다. 이와 같은 내용요소는 형성평가의 중요성을 일깨운 Black & Wiliam(1998)의 연구 이후 평가 전문성 기준의 내용 변화가 드러난 국외의 연구 결과와 비교하기 위함이다.

‘평가 상황’ 범주는 평가 도구나 평가 과정의 실행 주체에 따라 ‘표준화’, ‘교실’, ‘명시되지 않음’의 3개의 내용요소로 구성하였다. 상황이 명확하지 않을 경우 ‘교실’로 분석한 Looney 외(2018)의 분석틀과 달리, 본 연구에서는 ‘명시되지 않음’으로 구분하여 교실 평가를 강조(Shepard, 2000)하는 추세가 반영되었는지 살펴보고자 하였다.

〈표 1〉 교사의 평가 전문성 분석틀

범주		내용 요소
평가 목적		형성 평가, 총괄 평가, 진단 평가, 명시되지 않음
평가 상황		표준화, 교실, 명시되지 않음
평가 측면	평가 계획	교육과정을 고려한 평가 계획 수립, 학습자를 고려한 평가 계획 수립, 평가 목적과 활용을 고려한 평가 계획 수립, 평가 방법의 개발 및 선정, 평가 도구 개발 및 선정, 기타
	평가 실행 및 결과 해석	학습 과정 모니터링, 채점 및 성적 부여, 결과 해석, 기타
	평가 결과 관리와 이용	평가 결과의 관리, 평가 결과를 반영한 수업 개선, 학년·학교 수준의 의사결정 시 반영, 결과 보고, 기타
	평가 윤리	공정성, 윤리적·법적 책임준수, 기타
	평가 이론	신뢰도·타당도, 문항 개발, 기준·준거, 기타
	교실문화	교사의 자기 성찰, 학습목표에 대한 공유와 협상, 평가과정에 대한 공유와 협상, 자기/동료평가, 물리적 환경 조성, 문화적 환경 조성, 기타
	협력	동료교사와의 협력, 기타
측정 척도		지필평가, 리커트 척도

‘평가 측면’ 범주는 ‘평가 계획’, ‘평가 실행 및 결과 해석’, ‘평가 결과 관리와 이용’, ‘평가 윤리’, ‘평가 이론’, ‘교실 문화’, ‘협력’의 7개 소범주와 각 소범주별 2~7개의 내용 요소로 구성하였다. Looney 외(2018)와 DeLuca, LaPointe-McEwan, & Luhanga(2016b)의 분석틀을 재구성하되 일부 내용 요소는 최근 평가 동향을 반영하는지 살펴보기 위해 추가하였다. 평가에서 학습자의 역할을 어떻게 규정하는지 보기 위한 내용 요소는 ‘평가 과정에 대한 공유와 협상’, ‘학습목표에 대한 공유와 협상’, ‘자기/동료평가’(Klenowski, 2009; Xu & Brown, 2016)이다. 평가 주체성을 가진 교사가 어떻게 평가를 실행하는지 살펴보기 위한 내용요소는 ‘교사의 자기성찰’, ‘평가결과를 반영한 수업개선’, ‘동료교사와의 협력’(Willis, Adie, & Klenowski, 2013; NBPTS, 2016)이고, 평가 목적에 적합한 평가를 실시하는지 살펴보기 위한 내용 요소는 ‘평가 목적과 활용을 고려한 평가 계획 수립’이다. 평가 전문성의 기준이나 평가 전문성 측정 도구의 문항 내용이 7개의 소범주에는 속하나 적합한 내용 요소가 없을 경우에는 ‘기타’로 처리하였다.

‘측정 척도’ 범주는 측정 척도의 내용과 종류를 통해 평가 전문성 발달에 대한 연구자의 관점을 알아보기 위한 것으로 ‘지필평가’, ‘리커트 척도’의 두 가지 내용 요소로 구성하였다.

나. 평가 전문성 내용의 분석 및 데이터 처리

평가 전문성이 무엇인지 논의하기 위해서는 평가를 왜 하는지, 누가 평가를 계획하고 실행하는지에 대한 전제가 필요하다. 왜냐하면 평가 목적과 상황에 따라 요구되는 평가 전문성이 다르기 때문이다(Brookhart, 2011). 이에 본 연구에서는 평가 전문성 기준과 측정 문항의 ‘평가 목적’과 ‘평가 상황’에 따른 ‘평가측면’의 내용 요소별 빈도와 비율을 구하여 평가 전문성 내용과

교육과정 시기에 따른 특성을 비교 분석하였다.

먼저 평가 전문성이 무엇인지 분석하기 위하여 연구 대상 논문에서 평가 전문성의 기준과 측정도구를 추출하였다. 이후 평가 전문성 기준은 가장 하위 내용 요소를, 평가 전문성 측정도구는 측정 문항을 문장 단위로 나누고, 각 문장이 ‘평가 목적’, ‘평가 상황’, ‘평가 측면’의 세 가지 범주별로 각각 어떤 내용 요소에 해당하는지 분석하였다. ‘평가 측면’범주는 7개의 소범주 중 어느 내용 요소인가를 분석하였고, 한 문장에 여러 개의 내용 요소가 포함될 경우 중복 분석하였다. 예를 들면, ‘평가를 위한 물리적 환경과 심리적 환경을 적절하게 조성할 수 있다(김경희 외, 2006)’라는 평가 전문성 기준을 다음과 같이 분석하였다. 평가 목적이 명확하게 명시되지 않았으므로 ‘평가 목적’범주에서는 ‘명시되지 않음’으로, ‘평가 상황’ 범주에서는 표준화검사인지 교실 평가인지 명확하게 구분되지 않으므로 ‘명시되지 않음’으로 처리하였다. ‘평가 측면’ 범주에서는 한 문장에 ‘교실문화’ 소범주의 ‘물리적 환경 조성’과 ‘문화적 환경 조성’이라는 2개의 내용요소가 포함되므로 중복 처리하였다.

각 연구자가 전체 분석 대상 중 6편, 139개의 기준과 문항을 1차 분석하였고, 분석자간 일치도는 74.5%였다. 1차 분석 결과를 협의 검토하여 분석틀의 의미와 기준을 명료화한 후, 1인의 연구자가 분석 대상 전부를 2차 분석하였다. 그 결과를 3인의 연구자가 검토하여 합의된 결과를 도출하였다. 최종 분석 결과는 평가 전문성의 범주별 내용요소별로 빈도(%)로 나타내었다.

교사의 평가 활동은 교육과정 개정 방향과 평가 정책에 따라 영향을 받는데, 이러한 경향이 교사 평가 전문성 연구에 영향을 미치는지 살펴보기 위하여 교육과정 시기별로 분석하였다. 교육과정 시기는 교육과정 고시 시점을 기준으로 세 개로 나누었다. 최초 분석 대상의 발행연도인 2004년부터 2009개정 교육과정이 고시되기 전인 2008년까지, 2009 개정 교육과정이 고시된 2009년부터 2014년까지, 2015개정 교육과정이 고시된 2015년부터 2019년까지이다. 교육과정 개정이 주기적으로 이루어지지 않아 교육과정 시기별 빈도를 직접 비교하여 해석할 수 없으므로 각 교육과정 시기별 빈도를 백분율로 환산한 값을 비교하였다.

다. 평가 전문성 발달의 분석 및 데이터 처리

평가 전문성이 발달한다는 것은 시간의 변화에 따라 각 전문성 요소의 실행 빈도 증가와 같은 양적 발달뿐만 아니라 지식의 이해수준이나 실행의 수준과 같은 질적 발달을 모두 포함한다. 평가 전문성 측정도구의 측정 척도를 분석하면 전문성의 현재 수준을 측정하기 위해 수집하는 정보의 종류와 방법, 결과 해석 방법, 평가 전문성 발달 여부와 발달 정도의 판단기준을 파악할 수 있다. 이를 통해 평가 전문성 발달에 대한 도구 개발자의 관점을 분석할 수 있다. 이에 본 연구에서는 평가 전문성 발달이 드러난 평가 전문성 측정도구의 측정 척도를 추출하였다. 단, 국내 평가 전문성 기준은 교사들이 알고 행해야 할 능력의 목록으로 평가 전문성의 수준을 어떻게 측정할 수 있는지, 평가 전문성은 어떤 단계를 거치면서 발달하는지에 대한 내용이 없기 때문에 분석에서 제외하였다. 측정 척도의 종류에 따라 ‘지필평가’와 ‘리커트 척도’로 나누어서 빈도를 분석하였다. ‘지필평가’는 문항의 형식, 점수 산출 방법에 대한 내용을, ‘리커트 척도’는 단계의 내용, 점수 산출방법에 대한

내용을 분석하였다. 이를 바탕으로 어떻게 평가 전문성의 발달 여부와 정도를 판단하는지 논의하였다.

IV. 연구결과 및 논의

분석 대상으로 선정된 국내 평가 전문성 기준은 총 9개로, 2004년부터 2007년까지 개발되었다(표2). 한국교육과정평가원의 ‘교사의 학생평가 전문성 신장 모형과 기준개발 연구(이인제 외, 2004b)’는 학습 결과 확인과 서열 매기기의 기능을 넘어 학생의 학습 향상을 돕는 평가의 본래적 기능이 회복될 때 교육적으로 의미 있는 평가가 이루어질 수 있다는 전제 하에 문헌 연구, 설문 및 면담 조사, 워크숍을 통해 2개 범주와 33개 기준으로 구성된 학생평가 전문성 기준을 개발하였다. 개발된 평가 전문성 기준은 AFT, NCME, & NEA(1990)의 평가 전문성 기준과 상당히 유사하고, 타당도와 신뢰도 높은 평가의 실행과 관련된 내용이 주를 이루고 있다. 이는 ‘학습을 위한 평가’를 지향한다는 개발 방향과 달리 행동주의적 관점에 근거하여 평가를 인식한 결과로 보인다(박정, 2018). 이후 2004년의 평가 전문성 기준을 기반으로 일반적인 교사의 평가 전문성 기준(김경희 외, 2006)과 교과별 평가 전문성 기준(이인제 외, 2004c~2004g; 김동영, 2006; 최돈형 외, 2007)이 개발되었다. 개발된 기준은 평가 전문성 측정 도구(송미영, 김경희, 2007)의 이론적 배경이 될 뿐만 아니라 교사교육의 내용과 방법을 결정하는 역할을 하였다(김선희, 2012).

분석 대상으로 선정된 국내 평가 전문성 측정도구는 총 16개로, 2004년 이후 지속적으로 개발되었다(표3). ‘교사의 학생평가 전문성 신장연구(II)(김수동 외, 2005)’의 측정도구는 교사의 학생 평가 전문성 기준(이인제 외, 2004b)을 기반으로 개발되었다. 학교급이나 교과와 상관없이 모든 교사에게 적용 가능한 측정도구는 9개이며, 초등 교사를 위한 도구는 1개이고, 교과별 특성을 반영한 도구는 6개이다.

교사의 평가 전문성 내용과 발달은 어떠한지 ‘평가 목적’, ‘평가 상황’, ‘평가 측면’, ‘측정 척도’의 네 가지 범주와 교육과정 시기에 따라 분석한 결과 및 논의는 다음과 같다.

1. 국내 교사 평가 전문성 기준과 측정도구의 내용

가. 평가 전문성 기준의 내용

평가 전문성 기준 연구는 총 9편으로 모두 2007개정 교육과정 이전에 이루어진 것이다(표2). 평가 전문성 기준의 내용 변화를 살펴보기 위해 각 연구에서 개발한 평가 전문성 기준의 ‘평가 목적’과 ‘평가 상황’에 따라 ‘평가 측면’의 전문성 내용 요소를 분석하였다(표4).

〈표 2〉 연도별 평가 전문성 기준의 특성

연도	연구자	연구 논문명	기준의 특성
2004	이인제 외	교사의 학생 평가 전문성 신장 모형과 기준개발연구	2개 범주, 33개 기준
2004	이인제 외	국어과 교사의 학생 평가 전문성 신장 모형과 기준	4개 범주, 50개 기준
2004	이인제 외	사회과 교사의 학생 평가 전문성 신장 모형과 기준	2개 범주, 101개 기준
2004	이인제 외	과학과 교사의 학생 평가 전문성 신장 모형과 기준	2개 범주, 87개 기준
2004	이인제 외	수학과 교사의 학생 평가 전문성 신장 모형과 기준	2개 범주, 49개 기준
2004	이인제 외	영어과 교사의 학생 평가 전문성 신장 모형과 기준	2개 범주, 70개 기준
2006	김경희 외	교사의 학생평가 전문성 기준 개발	5개 기준, 55개 지식과 기능
2006	김동영	과학교사의 학생 평가 전문성 신장을 위한 자기 진단 프로그램 개발 연구: 지구과학 내용을 중심으로	5개 범주, 34개 기준
2007	최돈형 외	지속가능한 교육으로서의 환경교육 담당교사의 학생 평가 전문성 신장 모형 및 기준 개발 연구	7개 기준, 56개 내용

〈표 3〉 연도별 평가 전문성 측정도구의 특성

연도	연구자	연구 논문명	측정 척도의 특성	
			문항 수	척도의 종류
2004	이인제 외	교사의 학생 평가 실태 조사 및 전문성 신장에 대한 요구분석	4	3단 리커트 척도 ^a
2005	김수동 외	교사의 학생평가 전문성 신장 연구(II)	90	5단 리커트 척도 ^a
2006	김동영*	과학교사의 학생 평가 전문성 신장을 위한 자기 진단 프로그램 개발 연구 : 지구과학 내용을 중심으로	33	지필평가
2007	송미영김경희	교사의 학생평가 전문성 진단도구 개발	60	5단 리커트 척도 ^a
2012	박영민	쓰기 평가 지식 측정을 위한 검사 도구 개발 연구	22	지필평가
2013	하유라손원숙	초등교사의 개인 내적특성과 평가실제와의 관계	24	5단 리커트 척도 ^a
2015	김신영	교실 내 학생평가의 운영 실태와 개선 방향	9	3단 리커트 척도 ^b
2015	김석우 외	중등교사의 평가 전문성 제고 방안 : 서술형 평가 및 수행평가를 중심으로	53	5단 리커트 척도 ^a
2015	오세영	국어교사의 학생평가 전문성 연구: 평가자 협의에 따른 전략 차이를 중심으로	20	5단 리커트 척도 ^a
2015	최숙기	Rasch 모형을 활용한 국어 교사 평가 전문성 척도 개발 및 타당화 연구	10	5단 리커트 척도 ^a
2016	박은아서민희	성취평가제에서의 교사 역량 지표 개발 및 타당화	28	5/7단 리커트 척도 ^a
2017	김나영	중등학교 영어 교사의 학생평가 전문성 평가도구 개발 및 타당화 연구	25	5단 리커트 척도 ^a
2018	박지현 외	과정 중심 평가 내실화를 위한 교사의 평가 전문성 신장 방안 연구	60	5단 리커트 척도 ^a
2019	김유정 외	교사의 과정 중심 평가 역량(T-PEC) 진단도구 개발 및 타당화	20	5단 리커트 척도 ^a
2019	이수정김민정	전문교과 교사의 평가 전문성 진단도구 개발 연구	20	5단 리커트 척도 ^a
2019	임은영	교사의 학생평가 역량 연구	18	5단 리커트 척도 ^a

* 평가 전문성 기준과 측정도구를 모두 개발한 연구로 중복 분석

a: 3/5/7단계(전혀 그렇지 않다~ 매우 그렇다)

b: 3단계(잘 이해하지 못하며 이에 대한 도움이 필요하다, 이해는 하지만 실행방법을 잘 모른다, 이해하고 그 실행방법도 잘 알고 있다).

〈표 4〉 평가 목적과 평가 상황에 따른 평가측면의 평가 전문성 기준 수

평가측면	평가 목적 평가 상황	형성평가			총괄평가			진단평가			명시되지 않음			소계	계 (%)
		표준화	교실	명시되지 않음	표준화	교실	명시되지 않음	표준화	교실	명시되지 않음	표준화	교실	명시되지 않음		
평가 계획	교육과정을 고려한 평가계획 수립	3	1				1				40	109		154	
	학습자를 고려한 평가계획 수립						1				7	7		15	
	평가 목적과 활용을 고려한 평가계획 수립	2	1		1	1	1	1			5	13		25	289 (53.9)
	평가방법 개발 및 선정	6			2			2			17	12		39	
	평가도구 개발 및 선정	2									1	29	20	52	
	기타				1								4	5	
평가 실행 및 결과 해석	학습과정 모니터링	4	1					3			4			12	
	채점 및 성적부여					1					13	10		24	81 (15.1)
	결과해석							1		1	5	26		33	
	기타							1			3	8		12	
평가 결과 관리와 이용	평가결과의 관리			1							2			3	
	평가결과를 반영한 수업개선	9	8					1		1	2	6		27	60 (11.2)
	학년·학교수준의 의사결정 시 반영														
	결과보고	1	4				1				2	5		13	
	기타	1					1			1	7	7		17	
평가 윤리	공정성										5	9		14	
	윤리적·법적 책임준수										1	11		12	28 (5.2)
	기타											2		2	
평가 이론	신뢰도·타당도											1		1	
	문항 개발			3							1	4		8	24 (4.5)
	기준·준거														
	기타											15		15	
교실 문화	교사의 자기성찰										6	5		11	
	학습목표에 대한 공유와 협상	1												1	
	평가과정에 대한 공유와 협상	1									5	5		11	44 (8.2)
	자기/동료평가	1									3			4	
	물리적 환경조성										1	6		7	
	문화적 환경조성	2									3	5		10	
	기타														
협력	동료교사와의 협력										4	1		5	10 (1.9)
	기타			1								3		4	
소계		33	20		1	5	5	4	5		4	165	294	536	
계 (%)															(100)

평가 목적이 명시되지 않은 평가 전문성 기준이 대부분(463개, 86.4%)이었고, 평가 목적이 명시된 것은 형성평가(53개, 9.9%), 총괄평가(11개, 2.0%), 진단평가(9개, 1.7%) 순이었다. 평가는 평가 목적을 구체화하는 것에서 시작하고 그에 따라 평가 방법 및 절차가 달라질 수 있으므로(McMillan, 2014) 평가 목적이 명시되지 않은 기준은 구체적인 평가 실행 맥락을 보여주지 못하므로 평가의 질이 얼마나 향상되었는지 측정하고 안내하는 기준으로서의 역할(Wyatt-Smith et al., 2017)을 하기 어렵다. 교수학습평가의 연계와 함께 형성평가의 역할이 강조되고 있는 시점임을 고려하면 평가 목적을 형성평가로 명시한 평가 전문성 기준이 많아질 필요가 있다.

평가 목적에 관계없이 평가 상황이 명시되지 않은 기준은 324개(60.4%)로 가장 많았다. 이와 달리 평가 목적과 상황에 따라 필요한 평가 전문성을 구분하고 있는 국외 평가 전문성 기준(Brookhart, 2011)은 평가 상황별 평가 전문성의 내용에 대한 논의가 필요함을 보여준다. 학습 성취 결과의 보고를 목적으로 국가학업성취도평가와 같은 표준화평가를 실시할 때 필요한 평가 전문성과 교실 평가를 실시할 때 필요한 평가 전문성이 서로 다르기 때문이다.

평가 측면에서는 평가 목적과 평가 상황이 명시되지 않은 것이 많아(294개, 54.9%) 구체적인 평가 실행의 기준으로 한계가 있었다. 평가 측면별 기준 수를 살펴보면 ‘평가 계획(289개, 53.9%)’, ‘평가 실행 및 결과해석(81개, 15.1%)’, ‘평가결과 관리와 이용(60개, 11.2%)’과 같이 평가 과정과 관련된 기준이 대부분이며, ‘교실문화(44개, 8.2%)’, ‘평가윤리(28개, 5.2%)’, ‘평가이론(24개, 4.5%)’, ‘협력(10개, 1.9%)’은 상대적으로 적은 편이다. 이처럼 평가 과정과 관련된 기준이 많은 것은 개발된 기준들의 공통적인 이론적 배경인 AFT, NCME, & NEA(1990)의 기준 중 71.4%가 평가 과정과 관련된 기준이기 때문이며, 이러한 경향은 2000년 이전에 개발된 미국, 캐나다, 유럽 등 6개 지역의 평가 전문성 기준에서도 나타났다(DeLuca, LaPointe-McEwan, & Luhanga, 2016b). 그러나 1998년 형성평가의 효과에 대한 연구(Black & Wiliam, 1998) 이후 급증한 평가 연구들은 교사 중심의 평가가 아닌 학습자 중심의 평가, 학습을 위한 평가를 지향하고 있으며, 이는 2000년 이후 개발된 국외 평가 전문성 기준에 반영되어 있다(DeLuca, LaPointe-McEwan, & Luhanga, 2016b). 이와 같은 국제적 평가 동향은 국내 평가 전문성 기준 연구에도 영향을 미쳐 ‘학생의 자기/동료평가’, ‘평가과정에 대한 공유와 협상’ 등의 내용 요소에 반영된 것으로 보인다. 그러나 ‘학생의 자기평가와 학생 상호간의 평가방법을 교수·학습 과정에 적절하게 사용할 수 있다(김경희 외, 2006)’, ‘평가에서 사용될 평가의 기준을 명확하게 하고 학습자에게 전달할 수 있어야 한다(이인제, 2004c)’ 등과 같은 일부 기준은 학습을 위한 평가 개념을 온전히 반영한다고 보기 어렵다. 학생이 평가에 주체적으로 참여하기보다 교사가 전달해 준 평가 방법과 준거에 따라 평가에 참여하는 것은 교사의 지시를 준수하는 수준을 벗어나기 힘들기 때문이다. 교사에 의한 준거 공유 및 피드백은 도구주의(criteria compliance)에 빠질 수 있다는 연구결과(Torrance, 2007)와 학생들이 학습 과정을 조정하는데 평가 정보를 사용하거나 학습하는 방법을 학습하도록 하는 것이 학습을 위한 평가라는 연구결과(James & Mansell, 2009)는 평가 전략이나 방법 차원을 넘어(박정, 2019) 학생 주체성을 어떻게 확보해 줄 수 있을 것인지에 대한 논의가 필요함을 보여준다. 한편, 평가 주체성을 가진 교사의 모습은 ‘교사의 자기 성찰’, ‘동료교사와의 협력’ 등과 관련된 기준에 반영된 것으로 보인다. 그러나 ‘평가 계획에서부터 도구 개발·시행·해석·활용·의사소통에 이르는 평가 활동의 전 과정을 반성적으로 고찰하여 적절성을 판단할 수 있어야 한다(김경희 외, 2006)’와 같은 기준은 평가의 적절성을 어떻게 판단하는지, 성찰의 결과를 어떻게

활용해야 하는지에 대한 구체적인 내용이 없어 지향점이 불분명하다고 할 수 있다. 적절성의 판단 기준을 교사 계획의 실행 여부에 두는 것과 학습과정의 이해 및 교수학습 개선정도에 두는 것은 서로 다른 평가 목적을 지향하기 때문이다.

9개의 평가 전문성 기준 중 7개는 한국교육과정평가원의 학생평가 전문성 신장을 위한 3년 연구의 결과이며, 나머지 2개는 한국교육과정평가원의 평가 전문성 기준과 국외 평가 전문성 기준을 기반으로 개발된 것이어서 평가 전문성의 내용은 서로 밀접하게 관련되어 있었다. 그럼에도 불구하고 교과별 평가 전문성 기준의 평가 측면에서는 다소 다른 양상이 나타났다. 일반적인 교사의 평가 전문성 기준(이인제 외, 2004b)에서는 ‘협력’이나 ‘수업 중 학습과정에 대한 모니터링’과 관련된 기준이 없었으나, 국어(이인제 외, 2004c), 수학(이인제 외, 2004f), 영어(이인제 외, 2004g) 교과에 대한 평가 전문성 기준에는 ‘협력’과 관련된 기준이 2~5개씩 있었으며, ‘수업 중 학습과정에 대한 모니터링’과 관련된 기준 12개 중 8개는 영어 교과의 평가 전문성 기준(이인제 외, 2004g)이었다. 이는 교과 특성에 따라 강조되는 평가 맥락과 이에 필요한 평가 전문성이 다를 수 있음을 보여준다.

나. 평가 전문성 측정 도구의 내용

평가 전문성 측정도구를 개발 또는 적용한 연구는 총 16편이며, 2015개정 교육과정 이후에 10편의 연구가 이루어졌다(표3). 평가 전문성 측정도구의 내용 변화를 살펴보기 위해 ‘평가 목적’과 ‘평가 상황’에 따라 ‘평가 측면’의 전문성 내용 요소를 분석한 결과는 <표 5>와 같다.

평가 전문성 측정 문항에 평가 목적이 명시되지 않은 경우가 392개(78.7%)로 가장 많았으며, 형성평가를 목적으로 하는 문항이 69개(13.9%), 총괄평가 31개(6.2%), 진단평가 6개(1.2%) 순이었다. 평가 전문성 기준의 분석 결과에 비해 평가 목적이 명시된 경우의 비율이 다소 증가(7.7%)하였지만 여전히 명시되지 않은 비율이 높다. 이는 평가 목적이 명시되지 않은 평가 전문성 기준에 의거하여 평가 전문성 측정도구를 개발하였기 때문에 평가 목적이 명시되지 않은 측정 문항의 비율이 높을 수밖에 없는 것으로 판단된다. 평가 전문성 측정 문항에 평가 목적이 명시되지 않을 경우 개발자의 의도와 상관없이 문항의 응답자에 따라 문항의 해석이 달라질 수 있으며 평가 전문성 측정도구의 타당도와 신뢰도를 떨어뜨리는 요인이 되기도 한다.

평가 목적에 따른 평가 상황별 평가 전문성 측정 문항의 수를 비교하면, 평가 목적이 명시되지 않은 교실 평가와 관련된 문항이 가장 많았다(235개, 47.2%). 이는 평가 목적과 평가 상황이 모두 명시되지 않은 경우가 가장 많았던 평가 전문성 기준 분석 결과와는 달리 교실이라는 상황이 더 구체적으로 진술되었다는 점에서 의의가 있다.

평가 목적과 평가 상황에 따른 평가 측면별 평가 전문성 측정 문항 수를 비교하면, 평가 목적이 명시되지 않았지만 교실에서 실행된 평가와 관련된 문항이 235개(47.1%), 평가 목적과 평가 상황이 모두 명시되지 않은 문항이 148개(29.7%)로 대체로 평가 목적이나 평가 상황이 명시되지 않은 것이 많았다. 국내 평가 전문성 기준의 분석결과에 비해 평가 상황이 교실로 구체화된 문항이 늘어났다. 형성평가를 목적으로 교실에서 실행된 평가와 관련된 문항의 수가 48개(9.6%)로 평가 전문성 기준의 분석결과에 비해 다소 증가(약 3%)하였으며, 이 중 35개는 김신영(2015), 박지현 외(2018), 김유정 외(2019)의 연구에서 개발된 평가 전문성 측정 문항이다. 비록 3개의 연구에 불과하지만 평가 전문성 측정도구의 내용이 교수와 학습을 연계하는 평가를 강조하는 세계적인 평가의 방향(Birenbaum et al., 2015)과 일치한다는 점에서 의미가 있다고 본다.

〈표 5〉 평가 목적과 평가 상황에 따른 평가측면의 평가 전문성 측정 문항의 수

평가측면	평가 목적 평가 상황	형성평가			총괄평가			진단평가			명시되지 않음			소 계	계 (%)
		표 준 화	교 실	명 시 되 지 않 음	표 준 화	교 실	명 시 되 지 않 음	표 준 화	교 실	명 시 되 지 않 음	표 준 화	교 실	명 시 되 지 않 음		
평 가 계 획	교육과정을 고려한 평가계획 수립		6					1			37	10		54	
	학습자를 고려한 평가계획 수립						1	2	1		6	4		14	
	평가 목적과 활용을 고려한 평가계획 수립		1								2	1		4	164 (33.0)
	평가방법 개발 및 선정			2			1					12	4	19	
	평가도구 개발 및 선정		1			1					3	47	15	67	
	기타											5	1	6	
평 가 실 행 및 결 과 해 석	학습과정 모니터링		8	1		1					5			15	
	채점 및 성적부여 결과해석		1				4				15	12		32	102 (20.5)
	기타			2	3		4				2	17	13	41	
	기타		1	2								3	8	14	
평 가 결 과 관 리 와 이 용	평가결과의 관리		1								1	2	0	4	
	평가결과를 반영한 수업개선	1	14	8				2				7	3	35	87 (17.5)
	학년·학교수준의 의사결정 시 반영				2							2		4	
	결과보고		2	4		1	6				1	7	9	30	
	기타		1	1	1		4				1	4	2	14	
평 가 윤 리	공정성						1					7	12	20	
	윤리적·법적 책임준수												21	21	47 (9.4)
	기타											4	2	6	
평 가 이 론	신뢰도·타당도											3	4	7	
	문항 개발					1					1	4	3	9	40 (8.0)
	기준·준거											1	7	8	
	기타		6									8	2	16	
교 실 문 화	교사의 자기성찰		2									2	1	5	
	학습목표에 대한 공유와 협상											2		2	
	평가과정에 대한 공유와 협상											6	5	11	44 (8.8)
	자기/동료평가		2									6		8	
	물리적 환경조성											1	3	4	
	문화적 환경조성		2									2	3	7	
	기타											7		7	
협 력	동료교사와의 협력											10	2	12	14
	기타											1	1	2	(2.8)
소계		1	48	20	6	4	21	5	1	9	235	148		498	
계 (%)			69(13.9)			31(6.2)		6(1.2)		392(78.7)				(100)	

평가 측면의 소범주별 측정 문항 수를 평가 전문성 기준의 분석 결과와 비교해보면, 평가 계획(164개, 33.0%)과 관련된 문항의 비율은 20%정도 낮고, 평가 실행 및 결과해석(102개, 20.5%), 평가결과 관리와 이용(87개, 17.5%)과 관련된 문항의 비율은 각각 5%정도 높았으나 전체적으로는 평가 과정과 관련된 문항의 비율은 다소 낮았다(71%). 이는 내용 영역과 수행 영역으로 구분하여 개발된 평가 전문성 기준과 달리 수행 영역을 중심으로 개발된 평가 전문성 측정도구의 특성에서 비롯된 것으로 생각된다. 평가윤리(47개, 9.4%), 평가이론(40개, 8.0%)의 비율은 평가 전문성 기준과 비교하여 다소 높았으며, 교실문화(44개, 8.8%)와 협력(14개, 2.8%)의 비율은 비슷하였다.

교육과정의 개정과 평가 정책의 변화가 교사의 평가 전문성 연구에 미치는 영향을 살펴보기 위해 교육과정 시기에 따라 평가 목적, 평가 상황, 평가 측면의 평가 전문성 측정 문항의 수를 분석하였다(표 6).

교육과정 시기에 따라서는 형성평가를 목적으로 한 문항이 2007개정 교육과정 시기까지는 8.0%로 평가 전문성 기준의 분석결과(9.9%)와 비슷하였으나, 2015개정 교육과정 시기에서는 20.2%로 약 2배 증가하였다. 이는 ‘학습의 과정을 중시하는 평가를 강화하여 학생이 자신의 학습을 성찰하도록 하고, 평가 결과를 활용하여 교수학습의 질을 개선(교육부, 2015)’하고자 하는 2015개정 교육과정의 방향이 반영된 결과로 볼 수 있다.

평가 상황을 교육과정 시기별로 보면, 교실 평가와 관련된 측정문항은 2007개정 교육과정 시기(33.5%)에 비해 2015개정 교육과정 시기에서는 약 2배(76.8%) 증가하였다. 이는 교실 평가의 비중이 외부평가보다 높았던 Looney 외(2018)의 연구결과와 비슷하다. 이러한 결과는 교사 역할에 대한 인식이 외부 전문가가 개발한 평가를 교실에서 실시하는 시행자에서 평가 맥락에 적합한 평가를 개발하는 평가자로 바뀌어가고 있음을 의미한다(Xu & Brown, 2016).

〈표 6〉 교육과정 시기에 따른 평가 목적, 평가 상황, 평가 측면의 평가 전문성 측정 기준과 문항 수

범주	구분	교육과정시기	기준	측정도구		
			~2007개정 (n=536)	~2007개정 (n=188)	2009개정 (n=47)	2015개정 (n=263)
평가 목적		형성평가	53 (9.9)	15 (8.0)	1 (2.1)	53 (20.2)
		총괄평가	11 (2.1)	22 (11.7)	2 (4.3)	7 (2.6)
		진단평가	9 (1.7)	5 (2.7)	0 (0.0)	1 (0.4)
		명시되지 않음	463 (86.4)	146 (77.7)	44 (93.6)	202 (76.8)
평가 상황		표준화	5 (0.9)	13 (6.9)	1 (2.2)	2 (0.8)
		교실	207 (38.6)	63 (33.5)	27 (57.4)	202 (76.8)
		명시되지 않음	324 (60.4)	112 (59.6)	19 (40.4)	59 (22.4)
		교육과정을 고려한 평가계획 수립	154 (28.7)	11 (5.9)	2 (4.3)	41 (15.5)
평 가 계 획		학습자를 고려한 평가계획 수립	15 (2.8)	7 (3.7)	1 (2.1)	6 (2.3)
		평가 목적과 활용을 고려한 평가계획 수립	25 (4.7)	2 (1.1)	0 (0.0)	2 (0.8)
		평가방법 개발 및 선정	39 (7.3)	7 (3.7)	4 (8.5)	8 (3.0)
		평가도구 개발 및 선정	52 (9.7)	31 (16.5)	7 (14.9)	29 (11.0)
		기타	5 (0.9)	1 (0.5)	0 (0.0)	5 (1.9)
		평가 실행 및 결과 해석				
평 가 측 면		학습과정 모니터링	12 (2.2)	0 (0.0)	0 (0.0)	15 (5.7)
		채점 및 성적부여	24 (4.5)	15 (8.0)	9 (19.1)	8 (3.0)
		결과해석	33 (6.2)	22 (11.6)	2 (4.3)	17 (6.5)
		기타	12 (2.2)	4 (2.1)	0 (0.0)	10 (3.8)
		평가결과의 관리	3 (0.6)	1 (0.5)	0 (0.0)	3 (1.1)
		평가결과를 반영한 수업개선	27 (5.0)	9 (4.8)	3 (6.4)	23 (8.8)
		학년·학교수준의 의사결정 시 반영	0 (0.0)	2 (1.1)	0 (0.0)	2 (0.8)
		결과보고	13 (2.4)	13 (6.9)	1 (2.1)	16 (6.1)
		기타	17 (3.2)	7 (3.7)	1 (2.1)	6 (2.3)
		평가 윤리				
		공정성	14 (2.6)	7 (3.7)	2 (4.3)	11 (4.1)
		윤리적·법적 책임준수	12 (2.2)	13 (6.9)	2 (4.3)	6 (2.3)
		기타	2 (0.4)	2 (1.1)	2 (4.3)	2 (0.8)
		신뢰도·타당도	1 (0.2)	4 (2.1)	3 (6.4)	0 (0.0)
		문항 개발	8 (1.5)	8 (4.3)	1 (2.1)	0 (0.0)
		기준·준거	0 (0.0)	7 (3.7)	1 (2.1)	0 (0.0)
		기타	15 (2.8)	2 (1.1)	4 (8.5)	10 (3.8)
교 실 문 화		교사의 자기성찰	11 (2.1)	1 (0.5)	0 (0.0)	4 (1.5)
		학습목표에 대한 공유와 협상	1 (0.2)	0 (0.0)	0 (0.0)	2 (0.8)
		평가과정에 대한 공유와 협상	11 (2.1)	4 (2.1)	0 (0.0)	7 (2.7)
		자기/동료평가	4 (0.7)	3 (1.6)	0 (0.0)	5 (1.9)
		물리적 환경조성	7 (1.3)	2 (1.1)	1 (2.1)	1 (0.4)
		문화적 환경조성	10 (1.9)	2 (1.1)	1 (2.1)	4 (1.5)
		기타	0 (0.0)	0 (0.0)	0 (0.0)	7 (2.7)
		동료교사와의 협력	5 (0.9)	1 (0.5)	0 (0.0)	11 (4.1)
		기타	4 (0.7)	0 (0.0)	0 (0.0)	2 (0.8)

평가 측면 중 ‘교실문화’와 ‘협력’에 관한 측정문항의 수가 2007개정 교육과정 시기까지는 13문항(6.9%)이었으나 2015개정 교육과정 시기에서는 43문항(16.3%)으로 약 10% 증가하였는데, 이는 평가 맥락 관점에서는 매우 중요하다고 할 수 있다. 두 소범주는 학습을 위한 평가를 실현하기 위한 전제조건일 뿐만 아니라 평가 맥락에서 학생 주체성(student agency)과 교사 주체성(teacher agency)의 의미를 해석하는 기반이자 교실안과 밖의 협력을 의미하기 때문이다. 먼저, 학생 주체성 관점에서 보면, 교육과정 개정에 따라 학생 주체성이 조금씩 드러나기 시작한 내용 요소도 있었고 관점의 변화가 없는 내용 요소도 있었다. 예를 들어 ‘자기/동료평가’에 속하는 문항인 ‘교수학습과정에서 학생 자신에 의한 평가나 학생 상호 간에 의한 평가가 필요한 상황을 판단하고 적용할 수 있다(송미영, 김경희, 2007)’와 ‘학생들이 그들의 현재 수준에 도달해야 할 성취목표가 무엇이며 다음에 해야 할 학습이 무엇인지를 기술할 수 있게 한다(김신영, 2015)’의 내용을 비교해 보자. 전자는 자기평가와 동료평가를 평가 방법으로 생각하는 것에 비해, 후자는 학생이 주체성을 갖고 자신의 향상과 발달을 위해 평가에 참여하는 과정(Leahy et al., 2005; 박정, 2019)으로 바라보고 있다. 이와 같은 인식의 변화는 기준이나 준거와 같은 외적 근거를 넘어 학습자 자신에 근거한 평가로 확장되어 감을 보여준다(박정, 2019). 이러한 경향은 ‘문화적 환경 조성’으로 분류된 문항에서도 찾을 수 있다. 2007개정 교육과정 시기 이전의 측정문항에서는 ‘학생이 심리적으로 안정된 상태에서 평가가 시행될 수 있도록 격려하고 관리할 수 있다(김수동 외, 2005)’와 같이 학생을 평가 대상으로 바라보고 있어 학생 주체성에 대한 인식을 찾아보기 힘들었다. 이와 달리 2015개정 교육과정 시기의 측정문항에서는 ‘나는 학생이 자기주도적으로 참여할 수 있는 수업을 구성한다(김유정 외, 2019)’와 같이 학생의 참여권에 대해 고려하고 있었다. 그러나 학생의 자기주도적 참여를 어느 선까지, 어떻게 촉진하는지에 대한 구체적인 논의가 없어 평가 전문성 측정 문항으로서의 초점이 분명치 않다. 한편 ‘평가과정에 대한 공유와 협상’에 속하는 문항인 ‘평가를 실시하기 전에 학생에게 채점 기준에 대해 설명할 수 있다(김수동 외, 2005)’와 ‘학습목표에 대한 공유와 협상’에 속하는 문항인 ‘학생들에게 그들이 이해할 수 있는 용어로 성취목표에 대해 항상 알려주고 있다(김신영, 2015)’와 같은 문항에서는 학생의 목소리를 듣기보다는 교사가 정한 내용을 학생과 공유하는 수준에 그쳐 교육과정이 몇 차례 개정되어 왔으나 학생 주체성에 대한 관점 변화는 없는 것으로 보인다.

다음으로 교사 주체성 관점에서 보면, ‘교사의 자기성찰’과 ‘평가결과를 반영한 수업개선’과 관련된 문항의 비율은 교육과정이 개정됨에 따라 점차 증가했다. 이는 외부 전문가가 정해 놓은 평가 지식과 기능을 교사가 수동적으로 습득하는 차원을 넘어서기 시작했음을 의미한다. 교사가 평가의 결과를 자신의 교수활동에 반영하고 그 효과를 검증하는 과정을 지속적으로 성찰한다는 것은 교사 스스로 주체성을 가지고 평가 전문성을 향상시키고 있음을 보여준다. 또한, 2007개정 교육과정 시기까지는 1문항이었던 ‘동료교사와의 협력’과 관련된 문항이 2015개정 교육과정 시기에는 11문항으로 대폭 증가하였다. ‘나는 학년(교과) 협의회 등을 통해 평가 도구 및 방법의 적절성을 점검·보완한다(박지현 외, 2018)’와 같이 동료교사와의 협력을 강조하고 있다. 그러나 단순히 업무로서의 협력 수준을 넘어 자신의 평가 과정을 드러내고 동료와 함께 이를 평가하는 과정을 통해 전문성을 성장시키는 것이 효과적이라는 연구결과(Poskitt, 2014)에 비추어보면 동료교사와의 협력 수준에 대한 검토가 필요하다고 본다.

2. 국내 교사 평가 전문성 측정도구에 드러난 평가 전문성 발달

교사의 평가 전문성 수준을 어떻게 측정하고, 그 발달 여부와 정도를 어떻게 판단하는지 살펴보기 위하여 16개의 평가 전문성 측정도구의 측정척도를 분석하였으며(표3), 평가 전문성 발달에 대한 내용이 포함되지 않은 평가 전문성 기준은 분석에서 제외하였다.

14개의 평가 전문성 측정도구는 평가 실행의 능숙도나 이해와 실행 수준에 따라 3~7단계의 발달 단계로 구성된 리커트 척도를 활용하였으며, 외부전문가나 교사 자신이 평가한 응답의 총점이나 평균으로 평가 전문성 수준을 측정하였다. 2개의 평가 전문성 측정도구는 지필평가를 활용 하였으며, 총점에 따라 평가 전문성의 수준을 측정하였다. 국외에서도 리커트 척도(API, Zhang & Burry-stock, 1997; TAFL-Q, Pat-El et al., 2013)와 지필평가(TALQ, Plake et al., 1993; ALI, Campbell et al., 2002; Mertler & Campbel, 2005; CALI, Mertler, 2003)를 활용하여 평가 전문성을 측정하고 있다. 이와 같은 측정방법은 비교적 쉽게 평가 전문성을 측정할 수 있다는 장점이 있지만 평가 전문성 발달에 대한 관점을 제시하기에는 몇 가지 한계가 있다.

첫째, 총점이나 평균의 차이를 통해 평가 전문성의 발달여부와 상대적인 발달 정도를 파악할 수 있으나 발달 수준별 구체적인 질적 정보를 제공하지 않는다.

지필평가를 활용한 평가 전문성 측정도구는 문항의 정답 여부에 따라 평가 전문성의 유무를 판단하고, 총점에 따라 전문성의 수준을 판단하며, 총점의 향상 여부에 따라 평가 전문성의 발달 여부를 판단한다. 즉, 정답을 맞추면 해당 문항과 관련하여 이전에는 없었던 평가 전문성이 발달하였고, 총점이 향상되면 평가 전문성이 발달하였다고 판단함으로써 평가 전문성의 양적 변화가 쉽게 드러난다. 그러나 평가 전문성 측정 결과 동일한 점수를 획득한 교사의 평가 전문성 수준이 같다고 해석할 경우, 두 교사의 평가 전문성이 어떤 질적 차이가 있으며, 이후 평가 전문성 발달을 위해서 교사가 어떠한 노력을 해야 하는지 구체적인 방향성을 제시하기 어렵다. 김동영(2006)의 연구에서는 측정 점수에 따라 우수, 보통, 미흡으로 평가 전문성 등급을 평정할 수 있는 기준을 제시하고 각 등급별 평가 전문성을 향상시킬 수 있는 방향을 제시하였다는 점에서 의의가 있으나 교사의 맥락을 고려하지 않은 일반화된 내용에 그쳤다.

리커트 척도를 활용한 평가 전문성 측정도구는 발달 단계의 내용을 반영한 리커트 척도로 평가 전문성의 수준을 측정한다. 평가 전문성 측정도구 중 13개의 측정도구는 평가 전문성의 수준을 능숙도에 따라 ‘매우 잘함(매우 그렇다)’, ‘잘함(그렇다)’, ‘못함(그렇지 않다)’의 3단계 또는 여기에 ‘보통임’, ‘매우 부족함(매우 그렇지 않다)’을 추가하여 5/7단계로 나누었다(표3). 이와 같이 능숙도에 따라 평가 전문성의 발달 단계를 측정하면 리커트 척도 평균값의 향상 정도에 따라 평가 전문성의 양적 발달과 질적 발달 정도를 판단할 수 있다. 김수동 외(2005), 송미영과 김신영(2007)의 연구에서는 평균점에 따라 ‘평가 전문성이 높다’, ‘평가 전문성이 있다’, ‘평가 전문성이 낮다’, ‘평가 전문성이 매우 낮다’로 진단할 수 있는 기준선을 제시하고 있어 문항별, 영역별, 전체 영역에서의 평가 전문성 수준을 판단할 수 있다. 그러나 리커트 척도는 각 단계별 질적 판단 준거가 명확하지 않아서 측정자의 주관에 따라 평가 전문성 수준이 측정된다는 특성이 있어 측정자와 측정대상 간, 또는 측정자 간의 데이터 해석 일치를 보장할 수 없다. 이는 교사가

자기평가로 활용할 경우에도 측정 시기에 따라 단계의 해석이 달라질 수 있다는 점에서 동일하다. 또, 평가 전문성의 내용 요소별 발달 단계가 모두 똑같아서 평가 전문성 요소별 발달에 대한 구체적인 방향성을 제시하기에 제한적이다. 예를 들어, 평가 도구를 개발하거나 선정할 때 필요한 평가 전문성의 발달 단계는 평가과정을 학생들과 공유 또는 협상할 때 필요한 평가 전문성의 발달 단계와 다르다. 이를 극복하고자 평가 전문성 요소별 행동의 질과 횟수 등으로 능숙도에 대한 세부 평가 기준을 마련하여 타당도와 신뢰도를 확보한 측정도구(김나영, 2017)도 있었다.

한편, 리커트 척도의 내용을 평가 전문성의 이해와 실행 수준에 따라 ‘잘 이해하지 못하며 이에 대한 도움이 필요하다’, ‘이해는 하지만 실행방법을 잘 모른다’, ‘이해하고 그 실행방법도 잘 알고 있다’와 같이 3단계로 측정한 도구(김신영, 2015)도 있었다(표3). 이는 앞서 논의한 능숙도에 따른 리커트 척도와 두 가지 측면에서 다르다. 하나는 평가 전문성 발달 단계 간 구분의 명료성 정도이다. 능숙도에 따른 리커트 척도는 ‘부족함’과 ‘잘함’에 대한 기준이 명확하지 않아 측정자나 측정시기에 따라 측정 결과가 다르게 나올 수 있는 반면, 이해와 실행 수준에 따른 리커트 척도는 각 단계별 기준이 비교적 명확한 편이다. 다른 하나는 리커트 척도의 내용에 평가 전문성의 질적 발달 과정을 함께 논의하는지 여부이다. 능숙도에 따른 리커트 척도는 ‘매우 부족함’에서 ‘매우 잘함’까지 평가 전문성의 어떤 질적 성장이 이루어졌는지 구체적으로 드러나지 않는 반면, 이해와 실행 수준에 따른 리커트 척도는 단지 이해만 하는 수준에서 이해를 바탕으로 실행하는 수준까지 평가 전문성의 질적 성장을 보여준다. 그럼에도 불구하고 앞서 논의한 능숙도에 따른 리커트 척도와 같이 평가 전문성의 내용 요소별 발달 단계가 모두 동일하였다.

둘째, 평가 전문성 각 요소들의 발달이 평가 전문성의 발달로 연결된다고 봄으로써 다양한 변인이 상호작용하는 평가 맥락의 특성을 고려하지 못했다. 각 변인들에 대한 통제가 이루어지는 표준화 검사와 달리 교실 평가는 학생, 학습내용, 교사의 평가관 등 다양한 변인들이 상호작용한다. 특히 학교급에 따라 다른 학생의 특성과 교과별 특성은 평가 맥락에 큰 영향을 미친다. 그러므로 교사가 평가 맥락에 대한 이해 없이 일반화된 평가 전문성을 적용하는 것은 학생의 학습 성취뿐만 아니라 교사 자신의 평가 전문성 발달에도 효과적이지 않다. 이는 평가 전문성 발달에 대한 논의에는 평가 맥락에 대한 고려도 포함되어야 함을 의미한다.

이러한 한계 중 평가 전문성의 발달 단계에 대한 대안은 문제해결능력의 발달 정도에 따라 전문성 발달 단계를 구분하는 연구(Dreyfus & Dreyfus, 2005), 전문성 요소별 발달 단계가 제시된 국외의 교사 전문성 기준(AITSL, 2011), 교사 학습에 관한 연구(Cochran-Smith & Lytle, 1999)에서 찾을 수 있다. 예를 들어 의사결정의 질적 수준에 따라 ‘평가 지식과 기능을 이해하였으나 실행하지 못하는 수준’, ‘단순히 지식과 기능을 적용하는 수준’, ‘지속적인 성찰을 통해 실행을 개선하는 연구수준’, ‘리더십을 가지고 자신의 실행을 공유하는 수준’으로 세분화할 수 있다.

평가 맥락에 대한 대안은 DeLuca, LaPonite-McEwan, Luhanga(2016a)의 평가 전문성 측정도구 개발 연구에서 찾을 수 있다. 이 측정도구는 리커트 척도뿐만 아니라 우선순위를 활용하여 5가지 평가 상황에서 교사가 우선시 하는 평가 목적, 평가 과정, 공정성, 측정이론은 무엇인지 측정한다. 즉, 평가 전문성을 측정할 때 특정 평가 맥락에서 필요한 평가 지식과 기능, 실행뿐만 아니라 교사의 평가관까지 함께 고려하고 있다. 이는 평가 전문성 발달이란 특정 지식과 기능을

습득하는 것을 넘어 사회 문화적인 평가 맥락에서 목적에 적합한 의사결정역량이 향상되는 것임을 보여준다는 점에서 의의가 있다. 비록 구성주의 학습관을 가지거나 구성주의적 평가에 대해 알고 있는 교사일지라도 실제 평가 상황에서는 객관적인 평가 방법을 고수하기도 하므로(Shepard, 2000; Thomas et al., 2011) 평가 전문성 발달을 위해 교사의 평가관, 신념 등도 함께 고려할 필요가 있다.

V. 결론 및 제언

본 연구에서는 국내 교사 평가 전문성 기준과 측정도구 24편에 드러난 평가 전문성의 내용과 발달을 분석하였으며, 그 결과 다음과 같은 결론을 도출하였다.

첫째, 평가 전문성 기준과 측정 도구에 드러난 평가 전문성의 내용은 대부분 평가 과정과 관련된 지식과 기능으로 평가 목적과 평가 상황이 명확히 규정되지 않은 경우가 많았다. 평가 목적과 상황에 따라 필요한 평가 전문성이 다르므로, 구체적인 평가 맥락이 드러나지 않는 기준과 측정도구는 교사의 평가 전문성을 측정하고 안내하는 이정표의 역할을 하기에는 제한적이다.

둘째, 교육과정 개정의 영향으로 ‘형성평가’를 목적으로 한 평가 전문성 측정 문항과 ‘교실문화’와 ‘협력’에 관한 문항의 비율이 다소 증가하였다. 이러한 결과는 교사 역할을 평가 지식과 기능을 바탕으로 공정한 평가를 시행하는 자에서 학습향상을 목적으로 교수학습과 연계된 평가를 개발하고 실행하는 자로 바라보고 있음을 보여준다. 또한 학생을 평가 대상이 아닌 평가 주체로 인식하기 시작하였음을 의미한다. 그러나 ‘자기/동료평가’와 같은 일부 범주의 내용 요소에서는 학습을 위한 평가의 본질을 추구하기보다 형식적인 평가 전략으로 활용하는 것에 그쳤다.

셋째, 평가 전문성 수준은 지필평가의 총점이나 리커트 척도의 평균으로 측정하고, ‘매우 부족함’에서 ‘매우 잘함’까지 능숙도에 따라 발달 단계를 제시한 도구가 가장 많았다. 이와 같은 도구는 총점이나 평균의 향상정도로 평가 전문성의 양적 발달을 드러낸다는 장점이 있다. 그러나 리커트 척도의 발달단계는 문항의 응답자에 따라 판단 기준이 모호할 뿐만 아니라 발달단계마다 어떠한 차이가 있는지 구체적인 준거가 없어 질적 발달에 대한 방향성을 제시하기에 부족하다.

수업에서 교사와 학생의 역할 변화와 더불어 평가의 역할에 대한 인식의 변화는 이에 적합한 평가 전문성을 갖춘 교사를 필요로 한다. 이러한 교사를 양성하기 위해서 먼저 교사와 교사교육자에게 평가 전문성을 개발하기 위해 어떻게 해야 할지 구체적인 목표와 경로를 제공할 수 있는(Xu & Brown, 2016) 평가 전문성 기준이 개발되어야 할 것이다. 다음으로 평가 전문성 기준에 따라 평가 전문성의 발달 단계별 구체적인 질적 준거가 제시된 측정도구가 개발된다면 교사의 현재 평가 전문성 수준을 정확히 진단할 뿐만 아니라 전문성을 개발하기 위한 학습의 초점과 목표를 교사 스스로 세울 수 있도록 안내할 수 있을 것이다.

참고문헌

- 강현영, 고은성, 이동환, 이화영, 탁병주, 조진우, 김선희(2018). 수학과 과정중심평가 교사 연수 프로그램 개발 연구. **수학교육학연구**, 28(3), 321-343.
- 강훈식, 강석진(2015). 초등학교 교사의 과학 평가 전문성에 대한 자가진단 내용 분석. **초등과학교육**, 34(2), 153-163.
- 교육부(2015). 초·중학교 교육과정 총론. 교육부 고시 제2015-80호.
- 김경희, 김신영, 김성숙, 지은림, 반재천, 김수동(2006). 교사의 학생평가 전문성 기준 개발. **교육평가연구**, 19(2), 89-112.
- 김나영(2017). 중등학교 영어 교사의 학생평가 전문성 평가도구 개발 및 타당화 연구, 박사학위 논문, 서울대학교.
- 김동영(2006). 과학교사의 학생 평가 전문성 신장을 위한 자기 진단 프로그램 개발 연구 : 지구과학 내용을 중심으로. 박사학위 논문, 서울대학교.
- 김석우, 박상욱, 김윤용, 장재혁(2015). **중등교사의 평가 전문성 제고방안: 서술형 평가 및 수행평가를 중심으로**(수탁연구 CR2015-24). 한국교육개발원.
- 김선희(2006). 학생평가 전문성을 갖춘 수학교사 양성을 위한 수학학습평가 강좌의 교육 내용과 방법에 대한 제안. **학교수학**, 8(3), 301-326.
- 김선희(2012). 실습과 반성을 통한 수학 예비교사의 평가 전문성 신장. **수학교육학연구**, 22(2), 277-292.
- 김수동, 이의갑, 김경희, 김선희, 박은아, 신명선, 김수진, 박가나, 서수현, 전영석(2005). **교사의 학생평가 전문성 신장 연구(II)(연구보고 RRE 2005-3)**. 한국교육과정평가원.
- 김신영(2007). 교사의 학생평가전문성과 중등교사 양성과정. **교육평가연구**, 20(1), 1-16.
- 김신영(2015). 교실내 학생평가의 운영실태와 개선 방향. **교육과정평가연구**, 18(3), 257-282.
- 김신영, 송미영(2008). 예비교사의 학생평가전문성과 교육실습의 효과. **교육평가연구**, 21(1), 79-97.
- 김유정, 이경건, 장원형, 홍훈기(2019). 교사의 과정 중심 평가 역량(T-PEC) 진단도구 개발 및 타당화. **한국교원교육연구**, 36(3), 99-121.
- 박영민(2012). 쓰기 평가 지식 측정을 위한 검사 도구 개발 연구. **청람어문교육**, 46, 29-46.
- 박은아, 서민희(2016). 성취평가제에서의 교사 역량 지표 개발 및 타당화. **교육평가연구**, 29(3), 433-456.
- 박정(2013). 초등교사의 학생평가 전문성에 대한 자기평가 분석. **한국교원교육연구**, 30(3), 357-376.

- 박정(2017). 수업에서 학생평가 의미 탐색. **교육평가연구**, 30(3), 397-413.
- 박정(2018). 형성평가와 평가의 객관성. **교육평가연구**, 31(3), 483-499.
- 박정(2019). 과정중심평가를 위한 학생 자기평가 의미 탐색. **교육평가연구**, 32(3), 421-440.
- 박지현, 진경애, 김수진, 이상아(2018). **과정 중심 평가 내실화를 위한 교사의 평가 전문성 신장 방안 연구(연구보고 RRE 2018-5)**. 한국교육과정평가원.
- 박혜영, 이명애, 이명진(2019). 우리나라 미래 초·중등학교 교육평가 방향 탐색. **교육과정평가연구**, 22(3), 147-171.
- 성태제, 임현정(2014). 형성평가의 재인식에 따른 교사와 학교교육의 변화를 위한 제언. **교육평가연구**, 27(3), 597-615.
- 송미영, 김경희(2007). 교사의 학생평가 전문성 진단도구 개발. **교육과정평가연구**, 10(1), 47-74.
- 신혜진, 안소연, 김유원(2017). 과정 중심 평가 활용의 정책적 분석. **교육과정평가연구**, 20(2), 135-162.
- 오세영(2015). 국어교사의 학생글 평가 전문성 연구 :평가자 협의에 따른 전략 차이를 중심으로. 박사학위 논문, 고려대학교.
- 이수정, 김민정(2019). 전문교과 교사의 평가 전문성 진단도구 개발 연구. **직업교육연구**, 38(2), 51-65.
- 이인제, 김범기, 이범홍, 박정, 진재관, 김옥남, 서수현, 김신영, 강순희, 권재술, 김성원, 백성혜, 신동희, 이효녕, 조희형, 차희영(2004e). **과학과 교사의 학생평가 전문성 신장 모형과 기준(연구보고 RRE-2004-5-5)**. 한국교육과정평가원.
- 이인제, 남경희, 이범홍, 박정, 진재관, 김옥남, 서수현, 김신영, 서태열, 최상훈, 강선주, 심광택, 이간용, 강기원, 오연주(2004d). **사회과 교사의 학생평가 전문성 신장 모형과 기준(연구보고 RRE-2004-5-4)**. 한국교육과정평가원.
- 이인제, 박영목, 이범홍, 박정, 진재관, 김신영, 김옥남, 서수현, 김신영, 김수현, 김종철, 이성영, 이충우, 천경록(2004c). **국어과 교사의 학생평가 전문성 신장 모형과 기준(연구보고 RRE-2004-5-3)**. 한국교육과정평가원.
- 이인제, 신현용, 이범홍, 박정, 진재관, 김옥남, 서수현, 김신영, 고상숙, 김인수, 신인선, 이강섭, 김병수, 이증권, 최영기, 한인기(2004f). **수학과 교사의 학생평가 전문성 신장 모형과 기준(연구보고 RRE-2004-5-6)**. 한국교육과정평가원.
- 이인제, 전병만, 이범홍, 박정, 진재관, 김옥남, 서수현, 김신영, 박준언, 안병규, 오준일, 유제명, 이소영, 김신혜(2004g). **영어과 교사의 학생평가 전문성 신장 모형과 기준(연구보고 RRE-2004-5-7)**. 한국교육과정평가원.
- 이인제, 이범홍, 박정, 진재관, 김옥남, 서수현, 김신영(2004a). **교사의 학생평가 실태 조사 및 전문성**

- 신장에 대한 요구 분석(연구보고 RRE 2004-5-1). 한국교육과정평가원.
- 이인제, 이범홍, 박정, 진재관, 김옥남, 서수현, 김신영(2004b). 교사의 학생평가 전문성 신장 모형과 기준(연구보고 RRE-2004-5-2). 한국교육과정평가원.
- 임은영(2019). 교사의 학생평가 역량 연구. *교육과정평가연구*, 22(1), 101-124.
- 정민수, 김연경, 부재울(2019). 수학과 과정중심평가에 대한 교사학습공동체의 동료코칭 실행연구. *한국교원교육연구*, 36(1), 19-49.
- 주세형(2011). 평가 문식성 신장을 위한 국어과 교사 교육. *문법 교육*, 15, 31-50.
- 최돈형, 진옥화, 이향미, 손연아, 이성희, 조성화(2007). 지속가능한 교육으로서의 환경교육 담당교사의 학생 평가 전문성 신장 모형 및 기준 개발 연구. *환경교육*, 20(1), 1-18.
- 최숙기(2015). Rasch 모형을 활용한 국어 교사 평가 전문성 척도 개발 및 타당화 연구. *청람어문교육*, 56, 313-340.
- 하유라, 손원숙(2013). 초등교사의 개인 내적특성과 평가실제와의 관계. *초등교육연구*, 26(3), 215-236.
- 한국교육평가학회(2004). 교육평가 용어사전. 서울:원미사.

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In *The professional knowledge base of science teaching* (pp. 205-221). Springer, Dordrecht.

Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20, 91-106.

American Federation of Teachers, 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.

Australian Institute for Teaching and School Leadership [AITSL]. (2011). Australian professional standards for teachers: Graduate teachers. Retrieved from <http://www.aitsl.edu.au/australianprofessional-standards-for-teachers/standards/list?c=graduate>

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.

Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K.,

- Timperley, H., Volante, L. & Wyatt-Smith, CM. (2015). International trends in the implementation of assessment for learning : Implications for policy and practice. *Policy Futures in Education*, 13(1), 117-140.
- Black, P. J. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift für Psychologie*, 223(1), 3-13.
- Brookhart, S. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30, 3-12.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education*, 11(3), 301-318.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002). Psychometric analysis of an assessment literacy instrument: Applicability to pre-service teachers. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Cochran-Smith, M., & Lytle, S. L. (1999). Relationship of knowledge and practice: Teacher learning in communities. In A. Iran-Nejad, & C. D. Pearson (eds.), *Review of Research in Education* (Vol. 24, pp. 249-306). Washington, DC: American Educational Research Association.
- DeLuca C., Coombs A., MacGregor S., & Rasooli A. (2019). Toward a Differential and Situated View of Assessment Literacy: Studying Teachers' Responses to Classroom Assessment Scenarios. *Frontiers in Education*, 4, 94.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educ. Assess. Eval. Account.* 21, 248-266.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Teacher assessment literacy: a review of international standards and measures. *Educ. Assess. Eval. Account.* 28, 251-272.
- Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral vision: Expertise in real world contexts.

- Organization studies*, 26(5), 779-792.
- Edwards, F. (2017). A rubric to track the development of secondary pre service and novice teachers' summative assessment literacy. *Assessment in Education: Principles, Policy & Practice*, 24:2, 205-227.
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33, 14-18.
- Hay, P., & Penney, D. (2013). Assessment in physical education: A sociocultural perspective. London: Routledge.
- Herppich, S., Praetorius, A., Forster, N., Glogger-Frey, I., Karst, K., Leutner, D., et al. (2018). Teachers' assessment competence: integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*. 76, 181-193.
- James, M., & Mansell, W. (2009). Assessment in schools—fit for purpose? London: ESRC/TLRP.
- Joint Committee for Standards on Educational Evaluation. (2015). Classroom assessment standards: Practices for PK-12 teachers.
- Klenowski, V. (2009). Australian Indigenous students: Addressing equity issues in assessment. *Teaching Education*, 20(1), 77-93.
- Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255-276.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). classroom assessment: Minute-by-minute and day-by-day. *Educational leadership*, 63(3), 18-24.
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442-467.
- Lyon, E. G. (2013). Conceptualizing and exemplifying science teachers' assessment expertise. *International Journal of Science Education*, 35(7), 1208-1229.
- McMillan, J.H.(2014). Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction. 6th Edition. Boston:Pearson.
- Mertler, C. A. (2003). Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference? In Annual meeting of the Mid-Western Educational Research Association, Columbus, OH.

- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101-113.
- Mertler, C. A., & Campbell, C. (2005). Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the Assessment Literacy Inventory. In Annual meeting of the American Educational Research Association, Montreal, QC.
- National Board for Professional Teaching Standards. (2016). What teachers should know and be able to do.
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128-138.
- Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2013). Validation of assessment for learning questionnaires for teachers and students. *British Journal of Educational Psychology*, 83(1), 98-113.
- Plake, B., Impara, J., & Fager, J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-39.
- Popham, W. J. (2009). Assessment literacy for teachers: faddish or fundamental?. *Theory Into Practice*, 48, 4-11.
- Popham, W. J. (2013). Classroom assessment: What teachers need to know (7th ed.). Boston, MA: Pearson.
- Poskitt, J. (2014) Transforming professional learning and practice in assessment for learning, *The Curriculum Journal*, 25(4), 542-566.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers assessment literacy. *Journal of Science Teacher Education*, 22, 371-391.
- Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86, 22-27.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412.
- Thomas, L., Deaudelin, C., Desjardins, J., & Dezutter, O. (2011). Elementary teachers' formative evaluation practices in an era of curricular reform in Quebec, Canada. *Assessment in Education: Principles, Policy & Practice*, 18(4), 381-398.

- Timperley, H. (2011). A Background Paper to Inform the Development of a National Professional Development Framework for Teachers and School Leaders. AITSL, Melbourne, Australia.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education*, 14(3), 281-294.
- William, D. (2011). Embedded formative assessment. Solution Tree Press.
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241-256.
- Wyatt-Smith, C., Alexander, C., Fishburn, D., & McMahon, P. (2017). Standards of practice to standards of evidence: developing assessment capable teachers. *Assessment in Education: Principles, Policy & Practice*, 24(2), 250-270.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59-75.
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493-513.
- Zhang, Z., & Burry-stock, J. A. (1997). Assessment Practices Inventory: A multivariate analysis of teachers' perceived assessment competency. In Annual meeting of the American Educational Research Association, Chicago, IL.

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

ABSTRACT

Analysis of Standards and Instruments of Teacher Assessment Literacy in Korea

Pil-hee Woo

Teacher, Daegu Nambu Elementary School

Myung-seub Song

Professor, Daegu National University of Education

Ju-beom Song

Professor, Kyungpook National University

Recently, interests to teachers' assessment literacy, which influence to quality of the assessment, have been increased as assessment has been emphasized as the bridge between teaching and learning. 24 standards and instruments of teacher assessment literacy in Korea were analyzed by a framework of four identified classifications: purpose of assessment; contexts of assessment; aspect of assessment; scale of assessment literacy instruments. Most of the standards and instruments were not specified by purpose and contexts. Knowledge and skills related to assessment processes were revealed the most. However, proportion of item related to formative assessment, classroom culture, and collaboration in instruments of assessment literacy were somewhat increased under the influence of curriculum revision. Development of assessment literacy was analyzed by scale of assessment literacy instruments. As result of that, most of the instruments could measure the level of literacy and development quantitatively through average of Likert-type scale. However, the result had limit to suggest direction to teachers to develop assessment literacy. The result of study would suggest implication in terms of aspect and development of assessment literacy for teachers who implement assessment.

Key Words: *Assessment literacy, aspect of teacher assessment literacy, development of teacher assessment literacy*