

초등학교 국가수준 학업성취도 평가의 결과타당도¹⁾

조재민(서울남천초등학교 교사)*

민경석(세종대학교 부교수)**

박인용(한국교육과정평가원 부연구위원)

《 요 약 》

결과타당도는 평가 시행에 따른 결과가 본래 목적과 부합되는지, 평가 결과의 직·간접적인 사회/교육적 영향은 어떠한지에 대한 가치판단을 반영한다. 그러므로 국가차원에서 공교육의 질 관리를 위해서 대규모 평가를 활용하는 경우, 평가도구의 결과타당도는 더욱 강조되어야 한다. 이 연구는 초등학교 국가수준 학업성취도 평가의 결과타당도를 단위학교 교사의 관점에서 논의하기 위하여, 성취수준 진술문에 대한 교사의 분류 일치도, 국가수준 학업성취도 평가 결과와 교사의 예측 성취수준의 일치도, 단위학교의 교육 활동 속에서 평가 목적의 구현 정도를 분석하였다. 이 연구의 주요 결과는 다음과 같다. 첫째, 성취수준 진술문에 대한 교사의 분류는 국어과와 수학과 모두에서 국가수준 학업성취도 평가의 성취수준 분류에 대한 일치율이 낮았다. 둘째, 개별 학생의 국가수준 학업성취도 평가 결과와 교사의 예측 성취수준은 우수학과 기초학력 미달에서는 상대적으로 일치율이 높았지만, 기초학력과 보통학력에서는 낮게 나타났다. 셋째, 학업성취도 평가 목적의 구현에 대하여 교사는 전반적으로 낮은 인식을 보였다. 구체적으로, 단위학교 교사들은 평가 시행 결과가 학업 성취수준 파악, 학업 성취 추이 분석, 교수·학습 개선, 배경변인 정보 활용, 평가 전문성 신장에 미치는 영향은 중립적 또는 부정적으로 인식하고 있으나, 기초학력 보장에 미치는 영향에 대해서는 긍정적인 가치를 부여하는 것으로 판단된다. 이러한 연구 결과로 볼 때, 현재 국가수준 학업성취도 평가의 개선(초등학교 재시행 포함)에 대한 논의에서는 평가 절차(학년, 과목, 전수/표본조사 등) 뿐만 아니라 평가 결과의 교육적 활용(결과타당도) 측면에서도 타당성이 지속적으로 검토되어야 한다.

주제어: 초등학교 국가수준 학업성취도 평가, 결과타당도, 성취수준

1) 이 논문은 제1저자의 학위 논문(세종대학교 교육학과)을 재구성한 것이며, 일부 내용이 2013 한국교육평가학회 추계 학술대회에서 발표되었음.

* 제1저자, handache@sen.go.kr

** 교신저자, minkyungseok@sejong.ac.kr

I. 연구의 필요성 및 목적

학생평가를 실시한 이후 평가 결과가 본래의 목적과 일치하는지, 평가 결과를 사용함으로써 평가 목적이 달성되는지, 또한 평가 결과가 사회에 미치는 영향이 적절한지 등을 고려하여 평가 도구의 적절성을 판단하는 것을 결과타당도(consequential validity)라고 한다(Messick, 1989; Shepard, 1997). 이는 타당도를 측정 도구 자체의 문제로 국한시킨 제한적 관점에 다양한 사회적 결과나 가치판단의 문제를 포함시킨 개념으로 볼 수 있다. 이러한 결과타당도는 검사 활용의 목적 달성 및 직·간접적인 사회적 영향을 포함하는 결과적 상황에 의해 판단되는 타당도로, 평가 결과가 사회적으로 중요한 의사결정의 근거로 활용될 때 반드시 고려되어야 할 필요가 있다.

국가수준 학업성취도 평가(이하 학업성취도 평가)는 국가수준에서 교육의 질을 향상시키는 데 목적을 두고 있다. 국가수준에서 학생들의 학업성취도를 파악하고, 그 결과를 통해 교육과정 이 적절하게 기능하고 있는지 확인하고, 나아가서 교수·학습을 개선하여 학생들의 학업 성취를 향상시키는 것이 학업성취도 평가의 주요 목적이다(교육부, 2013). 지금까지 우리나라 학생들의 학업성취도는 지속적으로 향상되는 경향을 보였다. 특히 2008년 학업성취도 평가가 전수체제로 전환된 이후의 교육지표들을 살펴보면, 학업성취도 평가가 학력의 상향 평준화, 기초학력 미달 비율 감소(초등학교에서 3% 미만), 도·농간 학력 격차 감소 등에 긍정적인 영향을 미친 것으로 나타났다(김경희 외, 2011, 2012). 이러한 사실로 볼 때, 학업성취도 평가의 결과는 교육의 질 관리로 대표되는 평가 목적에 부합하고 있는 것으로 볼 수 있다. 그러나 2013년 교육부는, 행복교육 정책에 따라 고등학교에서는 현행 방식을 유지하고, 중학교는 평가 대상 교과를 5개 교과에서 3개 교과로 축소하며, 초등학교의 전수평가는 폐지하는 학업성취도 평가 기본 계획을 발표했다(교육부, 2013). 이와 같은 현 정부의 학업성취도 평가 정책 방향은 거시적인 교육 지표상으로는 학업성취도 평가가 학교교육에 긍정적인 영향을 미치고 있지만, 학생들의 시험부담, 학교·교육청 간 과열 경쟁 등의 부정적인 측면을 고려한 결과라 할 수 있다. 이러한 정책적 변화는 학업성취도 평가의 타당성을 거시적인 교육지표를 통해서만 고려할 것이 아니라, 단위학교 수준에 미치는 다양한 관점도 고려해야 함을 의미한다. 또한, 이러한 문제점들이 초등학교에서만 나타나는 것인지, 또는 초등학교에서는 국가수준의 학력 진단이 필요 없다는 것인지, 아니면 전집평가에서 표집평가 방식으로 지속될 필요성이 있다는 것인지 등에 대해서 보다 명확한 근거에 기반을 두고 논의될 필요가 있다(김경희 외, 2014; 김경성 외, 2013; 남현우, 2014).

학업성취도 평가 결과의 경우 직접적으로 학교 평가, 국가 교육의 질 관리 정책의 중요한 근거로 활용될 뿐만 아니라, 간접적으로도 교육 현장 및 관련 의사결정에 강력한 영향을 미칠 수

있다. 이러한 점에서 학업성취도 평가의 결과타당도는 중요하게 다루어질 필요가 있다. 학업성취도 평가가 2008년 표집평가에서 전수평가 체제로 전환된 주요 목적은 기초학력 지원 체계 구축, 개별 학생의 학업성취 향상과 더불어, 학생 개개인의 학습에 직결되는 정보를 제공함으로써 교수·학습 활동 전반을 지원하려는 것이다(김성숙 외, 2010). 구체적으로, 단위학교 교사는 학업성취도 평가 결과를 학교 및 학생 수준에서 해석하고, 교수·학습 활동의 개선을 위한 정보로 재구성하여 활용하며, 경우에 따라서는 관련 내용을 학생이나 학부모에게 제공한다. 따라서 학업성취도 평가 결과를 주도적으로 다루는 교사의 교육 활동을 통하여, 단위학교 수준에서 학업성취도 평가 시행 결과와 평가 목적이 부합되는지, 평가 목적이 구현되는지, 그리고 학업성취도 평가가 단위학교 수준의 교육 활동에 미치는 영향이 긍정적인지 등에 대한 추론이 가능하다.

이 연구는 전수체제로의 전환 목적에 따라 단위학교에서의 평가 결과 활용도에 초점을 두고, 교사의 학업성취도 평가 결과 해석 및 평가 목적 관련 교육 활동을 분석함으로써 단위학교 수준에서 학업성취도 평가의 결과타당도를 검증하는 데 목적을 두고 있다. 구체적으로, 이 연구에서는 초등학교 학업성취도 평가 결과에 대한 단위학교 교사의 이해 및 해석, 평가 목적과 관련된 교사의 활용도를 확인하기 위해 교사를 대상으로 학업성취도 평가의 성취수준별 진술문에 대한 교사 분류의 일치도, 학생의 학업성취도 평가 결과와 교사의 예측 성취수준의 일치도를 분석하며, 설문조사를 통해 단위학교 수준의 교육 활동 속에 학업성취도 평가 목적의 구현 정도를 경험적으로 검증하고자 하였다.

Ⅱ. 초등학교 국가수준 학업성취도 평가와 결과타당도

1. 국가수준 학업성취도 평가

국제적으로 교육의 질을 향상시키기 위한 국가수준의 대규모 평가 및 국가 간 비교 평가가 확대되고 있다. 미국의 NAEP(National Assessment of Educational Progress), 영국의 NCA(National Curriculum Assessment), 호주의 NAP(National Assessment Programme), 일본의 '전국학력·학습상황 조사' 등과 같은 대규모 평가를 통하여 각 국가는 학교 교육의 성과를 확인하고, 교육과정 개정과 교수·학습 개선을 위한 교육정보를 공유하고 있다(민경석, 2010, 2013). 우리나라에서도 국가수준의 학업성취도 평가가 한국교육과정평가원을 중심으로, 1998년 기본 계획이 마련되고, 1999년(예비평가: 사회, 수학), 2000년(사회, 수학), 2001년~2007년(국어, 사회, 수학, 과학, 영어), 2008년~2010년(전집평가로 전환), 2011년~2012년(전집평가: 국어, 수학, 영어, 표집평가: 사회, 과학)에 이르기까지 13년간 시행되어왔다. 더불어 학업성취도와 관련된 배경변인 분석, 2003년 개발된 평가틀 및 검사 동등화를 기반

으로 보다 체계적·실증적으로 학교와 교실의 교육 상황을 반영하려는 노력이 계속되어 왔다(김경희 외, 2011, 2012; 김성숙 외, 2010). 이러한 노력의 결과로 우리나라 초등학교 6학년의 기초학력 미달 비율은 2008년에 2.3%, 2010년에 1.5%, 2012년에 0.7%로 꾸준한 감소세를 보이는 등 학생들의 학력 향상에 기여했다(교육과학기술부, 2012). 그러나 2013년 초등학교 전집평가는, 학생들의 시험부담, 학교 간 경쟁 등을 이유로, 그 시행이 보류되었다. 이것은 대규모 평가 도구로서 학업성취도 평가의 실효성과 가치에 대하여 전국 수준의 평가 결과뿐만 아니라, 학교 현장의 구체적인 평가 결과에 대해서도 분석적 접근이 필요함을 시사한다.

학업성취도 평가의 목적은 국가수준에서 설정하여 각 시·도교육청 및 단위학교에 제공되며, 실제 시행 단계에서 일선 교육 현장의 교육 여건에 맞게 평가 목적을 재구성하여 진술할 수 있다. 그러나 국가수준 교육과정을 근거로 시행하는 대규모 평가이기 때문에 시·도교육청 및 단위학교 수준에서의 평가 목적 또한 국가수준의 목적을 그대로 적용하는 경향을 보인다(교육과학기술부, 2011; 서울시교육청, 2011). 학업성취도 평가 및 평가 결과에 대한 선행 연구에서 다루어진 목적 및 결과 활용 영역을 요약하면, <표 1>과 같다(김신영 외, 2011; 정은영 외, 2009).

〈표 1〉 학업성취도 평가 목적

학업성취도 평가의 주요 목적	평가의 결과 활용 영역
학업성취도 추이 점검을 통한 교육과정 질 관리	학업 성취 변화 분석을 통한 학업 성취수준 파악
	교육정책 시행 및 장학
	교육과정 운영 및 개선
	교수·학습방법 개선
기초학력 보장	기초학력 향상을 위한 기초학력 보장 정책 지원
학업성취도 배경 변인 분석	교육 효과성 분석
학교 현장의 평가 방법 개선	평가 절차 개선 및 교사 전문성 신장

학업성취도 평가의 목적은 2008년 표집평가에서 전수평가로 전환되면서, 학생 개인별 학업 성취도 파악 및 기초학력 보장의 목적이 추가되었고, 그 외의 목적은 지속적으로 유지되었다. 이러한 목적을 토대로 학업성취도 평가의 결과는, 국가수준의 교육과정 내용을 구체적으로 명시한 성취기준(achievement standards)에 근거하여, 학생들의 학업성취가 어느 정도 이루어졌는지 성취수준으로 제시된다. 평가의 근거가 되고 있는 성취기준은 이후 이루어지는 평가 결과에 담긴 정보의 질, 결과 활용의 질 그리고 최종적으로 평가 도구 자체의 목적 달성 여부에 영향을 미친다. 따라서 학업성취도 평가의 결과타당도는 성취기준의 설정과 적용을 고려할 필요가 있다.

학업성취도 평가 결과는 국가수준 교육과정에 따라 과목별로 성취기준에 도달된 정도를 ‘우수

학력, 보통학력, 기초학력'의 성취수준으로 구분된다. 이러한 구분은 2000년 학업성취도 평가부터 최소능력(minimum competency)을 적용하여 설정된 3지점의 기준점을 근거로 하고 있다. 각 성취수준이 나타내는 특성은, <표 2>와 같이, 평가 대상 학생들이 성취하기를 기대하는 기본 내용에 대한 이해 정도에 따라 기술되며, 기초학력에 이르지 못한 경우는 기초학력 미달로 분류된다(박정 외, 2006).

<표 2> 학업성취도 평가의 성취수준 (일반적 정의)

최소능력	성취수준	특성
도달	우수학력	평가 대상 학년 급 학생들이 성취하기를 기대하는 기본 내용을 대부분 이해한 수준
	보통학력	평가 대상 학년 급 학생들이 성취하기를 기대하는 기본 내용을 상당부분 이해한 수준
	기초학력	평가 대상 학년 급 학생들이 성취하기를 기대하는 기본 내용을 부분적으로 이해한 수준
미도달	기초학력 미달	교육과정의 일반적 내용과 범위의 기초가 되는 것을 달성하지 못한 수준

성취수준이 교육과정의 성취기준에 대한 학업 성취 정도를 의미하는 정보로 구현되기 위해서는, <표 2>와 같이, 포괄적인 성취기준을 수준별로 세분화한 성취수준 진술문으로 재구성되어야 한다. 이렇게 성취수준 진술문은 성취기준과 그 맥락이 동일하기 때문에 국가수준 교육과정, 단위학교 교육과정, 학업성취도 평가 등에 동등하게 적용될 수 있으며, 평가 내용의 연속성과 계열성 측면에서도 일관성이 유지되어야 한다.

2. 대규모 평가에서의 결과타당도

결과타당도는 1989년에 Messick에 의해 제안되었다(Shepard, 1997). Messick은 타당도 개념에 대해 '경험적인 증거와 이론적 타당성이 검사 결과에 기반을 둔 해석 및 행위의 정당성과 적절성을 지지하는 정도'로 보다 통합적인 판단 과정으로 이해하고, 그 판단의 주요한 부분으로서 검사 결과 활용으로 나타나는 사회적 영향력까지 고려할 것을 제안하였다(Messick, 1990). Messick의 결과타당도에 대한 제안은 결과타당도 개념의 적절성과 타당도 개념의 범위에 대한 논쟁을 촉발시켰다. 검사 결과 활용 및 사회적 영향을 고려하는 것의 당위성 자체에는 대다수의 학자들에게서 이견이 없었으나, 그러한 측면을 타당도의 한 부분으로 포함하는 것의 적절성 여부가 쟁점이 되었다(Popham, 1997).

결과타당도 개념에 제기된 비판에 대한 반박은 Shepard(1997)에 의해 구체적으로 이루어졌다. 먼저 Mehrens(1997)등이 타당도의 범위를 구인에 대한 추론의 정확성에 국한하려는 입장

에 대해, 이미 1950년대부터 검사 활용에 따른 영향이 타당도 판단에서 중요하게 고려되어 왔다고 반론을 제기했다. Shepard는 이러한 선행 연구들을 통해 결과타당도가 타당도 개념에 이미 포함되어 있으며, 검사를 개발한 의도 자체가 논리적으로 검사를 활용하는 맥락의 목적성과 결부되어 설명될 수밖에 없기 때문에 결과타당도에 대한 고려는 당연하다고 주장했다.

학업성취도 평가와 같은 대규모 평가의 결과타당도는 다음과 같은 점에서 특히 중요하게 다루어질 필요가 있다. 먼저, 대규모 평가 및 그 결과의 사회적 영향에 대한 가치판단이 중요하다. 결과타당도는 평가 개발자 및 사용자들이 평가 결과의 활용에 따른 사회적 영향에 대해 자각함으로써, 그것을 예상하고 점검하는 책임을 갖도록 하는 데 중요한 의미가 있다(성태제, 2003; Linn 1997). 대규모 평가 결과는 학교 평가, 국가 교육의 질 관리 정책, 교육 현장 및 관련 의사결정 등의 중요한 근거로 활용된다. 이러한 점에서 대규모 평가의 경우 그 영향에 대한 사회적 가치판단의 중요성이 강조된다. 둘째, 대규모 평가의 결과타당도에서 평가 결과와 목적과의 일치 여부가 중요하다. 교육 책무성을 위한 대규모 평가의 목적은 학생들의 성취를 측정하는데 국한되는 것이 아니라, 그 결과에 기반하여 교수·학습을 개선하고 학생들의 학업성취도를 향상시키는 데 있다. 그러므로 대규모 평가의 결과타당도에 대한 위의 두 측면을 충분히 고려하기 위해서는 결과타당도의 범위를 측정 도구와의 관련성으로 국한하기보다는 실행 과정 전반과 활용에 따른 직·간접적인 결과 및 사회적인 영향까지 넓게 포괄하는 관점이 더 적합하다고 볼 수 있다(Kane, 2006).

Lane과 Stone(2002)은 주 단위 평가(State Assessment)와 책무성 프로그램에서 확인될 수 있는 결과타당도의 증거와 구체적인 검증 방식을 제안하였다. Lane과 Stone은 먼저 주 단위 평가 및 책무성 프로그램의 결과타당도 대상으로 포함될 수 있는 의도적 결과와 비의도적이며 부정적인 결과를 <표 3>과 같이 제시하였다.

<표 3> 주 단위 평가 프로그램의 의도적 결과와 비의도적 결과

의도적 영향	비의도적(부정적) 영향
<ul style="list-style-type: none"> • 학교, 교사, 행정가의 동기 부여와 노력 • 교육과정과 교수 내용, 전략 개선 • 교실 평가의 내용, 형식 발전 • 전체 학생의 학습 향상 • 전문성 발달 지원 • 시험 준비 활동의 활용 및 성격 • 평가, 평가기준, 평가 결과 활용에 대한 학생, 교사, 행정가, 대중의 인식 및 신념 	<ul style="list-style-type: none"> • 평가기준에 직결되는 시험문제 예상 범위의 교육과정 및 교수방법의 파행 운영 • 교수방법의 개선없이 평가내용과 유사한 시험 준비 자료만 활용 • 비윤리적인 시험 준비 자료 사용 • 부적절하고 불공정한 방식의 시험점수 활용 • 학생 하위 집단(인종 등) 간에 수행 향상도 차이 • 과거의 시험 경험으로 인한 학생의 동기와 자신감 저하

출처: Lane & Stone(2002, p. 24), 좌혜정과 성태제(2003, p. 334)를 참고하여 재구성함

Lane과 Stone(2002)은 평가 프로그램의 결과타당도에 대한 경험적인 검증 절차를 구체화하기 위하여 <표 4>와 같이, 평가의 목적이 교수·학습 개선과 책무성 관리일 경우, 필수적으로

확인되어야 할 내용을 진술하고 각 항목에 따른 검증 방식을 제시하였다.

〈표 4〉 진술된 결과 및 결과타당도 증거

결과와 이해관련자	결과적 증거	조사 방법
<ul style="list-style-type: none"> 학교행정가와 교사가 기준(standard)에 따라 교수 및 교육과정을 개선하도록 동기화되었다. 	<ul style="list-style-type: none"> 기준과 평가에 대한 친숙성 기준과 평가에 대한 신념과 태도 	<ul style="list-style-type: none"> 교사와 교장 설문지
<ul style="list-style-type: none"> 전문성 향상을 위한 지원이 이루어지고, 자원 활용이 가능하다. 	<ul style="list-style-type: none"> 전문성 향상 지원의 성격 전문성 향상 지원의 규모 	<ul style="list-style-type: none"> 교사와 교장 설문지
<ul style="list-style-type: none"> 교수와 교육과정이 기준에 맞게 개선되었다. 	<ul style="list-style-type: none"> 교수 및 교실 평가 내용이 주단위 평가와 기준을 반영하는 정도 교수와 교실평가 전략이 기준에 맞게 설정되는 정도 	<ul style="list-style-type: none"> 교사, 교장, 학생 설문지 교수와 교실 평가 과제, 채점 기준, 시험 준비 활동과 같은 교실 환경
<ul style="list-style-type: none"> 학생들은 평가에 최선을 다할 뿐만 아니라, 학습동기가 높다. 	<ul style="list-style-type: none"> 기준, 교수, 평가에 대한 신념과 태도 평가에 발휘되는 학생의 노력 	<ul style="list-style-type: none"> 교사와 학생 설문지
<ul style="list-style-type: none"> 수행의 향상이 교수·학습과 관계가 있다. 	<ul style="list-style-type: none"> 학교와 학습 향상과 동기, 교수와 교실 평가, 전문성 향상 변수 간의 관계 	<ul style="list-style-type: none"> 장기간에 걸친 교사, 교장, 학생 설문지 장기간에 걸친 학급 및 학교의 수행 결과

출처: Lane & Stone(2002, p. 27), 좌혜정과 성태제(2003, p. 335)를 참고하여 재구성함

〈표 4〉에서 보는 바와 같이 Lane과 Stone(2002)은 주 단위 평가의 결과타당도에 학생, 교사, 학교행정가 관련 내용을 모두 포함시키고 있다. 또한 교육·학습 개선과 수행 향상의 결과적 측면과 교수·학습 주체들의 동기적인 측면을 포괄하고 있다. 그리고 각 항목에 대해 결과타당도를 확인할 수 있는 증거와 자료 수집의 출처를 나열하고 있다. 예를 들어 〈표 4〉의 ‘학생들의 평가 참여 및 학습의 동기 부여’에 관련된 증거는 교사와 학생의 설문지를 통해 평가에 발휘되는 학생의 노력에 대해 조사함으로써 얻을 수 있음을 보여준다.

학업성취도 평가는 기본적으로 그 주요한 목적을 기초학력 보장 지원, 학업 성취수준 향상에 두고 있기 때문에 단위학교의 교수·학습 개선이 결과타당도의 주요한 관심이 된다. 이에 따라 본 연구에서는 현직 교사들을 대상으로 학업성취도 평가 결과의 해석과 활용에 대해 보다 면밀한 분석을 시도하였다. 교사는 단위학교 교수·학습 과정에서 학업성취도 평가의 목적을 달성하는 데 있어서 핵심적인 역할을 하며, 학업성취도 평가 결과에 대한 교사의 정확한 해석은 평가도구의 목적 달성을 위한 중요한 전제조건이다. 그리고 그러한 해석의 기반은 학업성취도 평가와 단위학교 교사 간의 성취기준과 성취수준 등의 준거 공유를 통해 가능해질 수 있다. 이 연구에서는 학업성취도 평가 결과타당도의 근거로서 성취수준 진술문에 대한 교사의 이해, 학업성취도 평가 결과와 교사의 학생 평가 결과의 일치 정도, 단위학교에서의 학업성취도 평가 결과 활용 실태 등을 분석했다.

Ⅲ. 연구 방법

1. 표본 및 자료 수집

초등학교 학업성취도 평가의 결과타당도를 확인하기 위하여 서울시 초등학교의 교사 및 학생 자료를 수집하였다. 먼저, 2012년 학업성취도 평가 결과와 학교의 사회·경제적 여건을 반영하여 표집 대상 학교를 선정했다. 서울 지역의 공립 초등학교 545개교 중 521개교의 2011년 학업성취도 평가 결과에서 보통학력 이상 비율을 기준으로 3개 그룹으로 나눈 후, 상위그룹에서 4개교, 중위그룹에서 4개교, 하위그룹에서 6개교를 표집하였다. 자료 수집 단계에서 기초학력 미달 학생들의 표본을 충분히 확보하기 위해 하위그룹의 학교 수를 2개교 더 많이 설정하였다. 이러한 기준에 따라 서울 지역 5개 교육지원청, 14개 초등학교, 70명의 6학년 담임교사 그리고 1,840명의 학생에 대한 자료를 수집하였다.

〈표 5〉 표본 특성과 자료 수집

학교 (공립학교)	교육지원청	교사	학생
14개 학교	5개 교육지원청	70명 (학교 당 약 5명의 6학년 담임)	1,840명
조사 내용		성취수준 진술문 분류, 학업성취도 평가 결과의 적용	2012년 학업성취도 평가의 성취수준, 교사가 예측한 성취수준 (국어, 수학)

표집 대상 교사는 모두 6학년 교사들로 한정하였으며, 학생은 교사 70명의 담임반 학생들 중에서, 학생 및 학부모가 자료 수집에 동의한 1,840명을 조사 대상으로 하였다. 표집 교사는 6학년 담임 경력을 기준으로 1년 이상 5년 미만이 약 66% 정도이며, 약 33%가 5년 이상의 경력을 가지고 있었다. 총 교육 경력을 기준으로 1년 이상 5년 미만이 약 20% 정도였으며, 약 79%가 5년 이상의 경력을 가지고 있었으며 이 중 20년 이상의 경력을 소유한 교사가 20%를 차지하고 있었다.

학생들의 학업 성취수준 관련 자료는 국어와 수학 과목으로 한정하였으며, 2012년 6월에 교사의 예측 성취수준 자료를, 9월에 학업성취도 평가 결과(성취수준)를 조사하였다. 최초 1,840명의 학생 중 학업성취도 평가 결과 자료가 누락된 학생을 제외한 1,835명의 자료가 분석에 포함되었다. 즉, 교사가 예측한 학생의 성취수준과 학업성취도 평가 결과로 제시된 성취수준을 비교함으로써 학업 성취수준 진단이라는 평가 목표의 달성 여부를 분석하였다.

2. 교사 설문조사

이 연구에서는 교사 설문을 통해 학업성취도 평가의 성취수준 진술문에 대한 교사의 분류 정확성과 단위학교 수준의 교육 활동 속에서 학업성취도 평가의 목적이 어느 정도 달성되고 있는지를 조사하였다.

학업성취도 평가 성취수준 진술문 관련 설문에서는 과거 3년(2009, 2010, 2011) 동안의 평가 결과표를 근거로 성취수준별 진술문 30문항(국어, 수학 과목 각 15개 성취수준 진술문)을 제시하였다. 교사는 해당 진술문을 <표 1>에 제시된 학업성취도 평가 성취수준의 일반적 정의에 기반하여 우수학력, 보통학력, 기초학력 중 가장 적절한 성취수준을 선택하였다.

<표 6> 성취수준별 진술문 대한 설문지 문항 구성

성취 수준 진술문	문항 내용	분류
국어과 성취수준 진술문	'글을 읽고, 글 구조의 적절성을 평가할 수 있다.' 외 14개 진술문	각 진술문을 세 가지 성취수준(우수, 보통, 기초) 중 하나로 분류
수학과 성취수준 진술문	'문제해결과정의 타당성을 검토할 수 있다.' 외 14개 진술문	

학업성취도 평가 결과 활용에 대한 설문지는 선행 연구(김성숙 외, 2009; 민경석, 2010; Mertler, 2011), 그리고 학업성취도 평가의 설문지(이상하 외, 2010)등을 통하여, 학업성취도 평가 목적이 교사의 교육 활동 속에서 구현되고 있는 실태를 파악할 수 있는 문항으로 <표 7>과 같이 구성하였다.

<표 7> 학업성취도 평가 결과 활용에 대한 설문지 문항 구성

구분	문항 내용	문항수
학업 성취수준 파악	평가 결과를 통한 성취수준 파악	4
교수·학습방법 개선	수업방법, 수업내용, 수업운영 등의 조정	5
학업 성취 추이 분석	'기초학력미달' 학생 비율의 변화 파악	3
학업성취 배경변인 분석	학교특성에 대한 설문 분석 결과의 활용	4
교육정책 시행 및 장학	학교별, 교육청별 학업성취 비율 공개의 긍정적 활용	4
평가 전문성 신장	새로운 평가 방법, 결과 제시 방법 개발	3
교육과정 정착	교육과정의 충실한 운영	4
교사의 성취수준 인식	성취수준 분할 기준의 적절성	3
교사의 평가 성향	절대/상대 평가 활용	3
일반적인 학업성취 결정 요인	학교 및 학급 환경	6
계		40문항

설문지 문항은 교사의 교육 활동 중에서 학업성취도 평가의 목적이 되는 7가지 영역과 관련 되는 상황을 고려하여 현직 교사와의 협의를 통해 개발하였다. 그리고 교사의 성취수준 인식, 교사의 평가 성향, 일반적인 학업 성취 결정 요인 등의 내용을 설문에 포함하였다.

IV. 연구 결과

1. 성취수준 진술문 분류와 학생 성취수준 예측

학업성취도 평가의 성취수준 진술문에 대한 교사 분류의 일관성을 분석한 결과는 <표 8>과 같다.

<표 8> 학업성취도 평가 성취수준 진술문에 대한 교사의 분류 결과

구분		국어과(일치율)	수학과(일치율)
성취수준 진술문 분류	교사의 분류		
우수학력	우수학력	31.2%	22.1%
	보통학력	59.7%	63.2%
	기초학력	9.1%	14.7%
보통학력	우수학력	48.7%	23.1%
	보통학력	47.3%	64.6%
	기초학력	4.0%	12.3%
기초학력	우수학력	19.7%	6.0%
	보통학력	61.6%	52.4%
	기초학력	18.7%	41.6%

성취수준별로 설정된 학업성취도 평가 성취수준 진술문과 교사 분류 간의 일치율은 수학과가 국어과보다는 상대적으로 높았다. 그러나 수학과에서도 보통학력(64.6%)을 제외하고는 높은 일치율이 관찰되지 않았다. 전반적으로 교사들은 학업성취도 평가의 성취수준 진술문을 보통학력 수준으로 분류하는 경향을 보였다. 세부적으로 살펴보면 우수학력 진술문의 경우, 국어과 9.1%, 수학과 14.7%의 교사들은 기초학력 수준으로도 분류하였다. 또한 기초학력 진술문의 경우에도 국어과 19.7%, 수학과 6.0%의 교사들이 우수학력으로 분류한 것으로 나타났다. 결과적으로 학업성취도 평가 및 교사의 분류 기준이 모두 국가수준 교육과정에 근거하고 있음에도 불구하고, 성취수준 진술문에 대한 양측의 분류 편차가 적지 않은 것으로 나타났다.

표본에 포함된 14개 학교 1,835명 학생의 2012 학업성취도 평가 결과와 교사 예측 성취수준의 일치도 분석 결과를 <표 9>와 <표 10>에 제시하였다.

<표 9> 학업성취도 평가 결과와 교사 예측 성취수준(국어과)

빈도(%)

구분		학업성취도 평가 결과				전체
		우수학력	보통학력	기초학력	기초미달	
교사 예측 결과	우수학력	692(37.7%)	231(12.6%)	50(2.7%)	0(0.0%)	973(53.0%)
	보통학력	209(11.4%)	248(13.5%)	128(7.0%)	1(0.1%)	586(31.9%)
	기초학력	19(1.0%)	82(4.5%)	87(4.7%)	6(0.3%)	194(10.6%)
	기초미달	1(0.1%)	11(0.6%)	58(3.2%)	12(0.7%)	82(4.5%)
전체		921(50.2%)	572(31.2%)	323(17.6%)	19(1.0%)	1,835(100%)

$$\chi^2 = 1822.970(9)^{**}$$

** p<.01

<표 10> 학업성취도 평가 결과와 교사 예측 성취수준(수학과)

빈도(%)

구분		학업성취도 평가 결과				전체
		우수학력	보통학력	기초학력	기초미달	
교사 예측 결과	우수학력	585(31.9%)	260(14.2%)	31(1.7%)	0(0.0%)	876(47.7%)
	보통학력	106(5.8%)	332(18.1%)	120(6.5%)	2(0.1%)	560(30.5%)
	기초학력	11(0.6%)	83(4.5%)	146(8.0%)	8(0.4%)	248(13.5%)
	기초미달	2(0.1%)	19(1.0%)	107(5.8%)	23(1.3%)	151(8.2%)
전체		704(38.4%)	694(37.8%)	404(22.0%)	33(1.8%)	1,835(100.0%)

$$\chi^2 = 1119.523(9)^{**}$$

** p<.01

먼저, <표 9>에 제시된 국어과의 '우수학력, 보통학력, 기초학력, 기초학력 미달' 학업성취도 평가 결과 인원은 각각 921명(50.2%), 572명(31.2%), 323명(17.6%), 19명(1.0%)이며, 교사가 학교 교육 활동에 기반하여 예측한 성취수준은 973명(53.0%), 586명(31.9%), 194명(10.6%), 82명(4.5%)으로 나타났다. 두 가지 학생 성취수준의 전체적인 분포 경향은 성취수준이 높아질수록 빈도가 증가하는 특성을 보인다. 학업성취도 평가 결과 대비 교사 예측 성취수준 비율이 기초학력 미달은 4.32배, 보통학력은 1.02배, 우수학력은 1.06배로 더 높으나, 기초학력에서는 학업성취도 평가 결과보다 0.6배 낮은 비율을 보였다. 또한 학업성취도 평가 결과에서는 전체 학생 중 1.0%만을 기초학력 미달로 분류하고 있지만, 교사는 이에 비해 약 5배에 가까운 4.5%를 기초학력 미달로 예측하고 있다. 국어과에서 교사 예측과 실제 평가 결과 간

의 일치 정도는 평균 56.6%로 나타났다. 그러나 구체적으로 살펴보면, 학업성취도 평가에서는 우수학력으로 평가된 921명(50.2%)에 대하여 교사는 기초학력 미달에 1명(0.1%), 기초학력에 19명(1.0%)으로 분류하는 편차를 보였다.

〈표 10〉에 제시된 바와 같이, 수학과 실제 평가 결과의 ‘기초학력 미달, 기초학력, 보통학력, 우수학력’ 인원은 각각 33명(1.8%), 404명(22.0%), 694명(37.8%), 704명(38.4%), 교사 예측 성취수준은 151명(8.2%), 248명(13.5%), 560명(30.5%), 876명(47.7%)으로 나타났다. 이러한 결과는 국어과와 동일하게 성취수준이 높아질수록 빈도가 증가하는 분포 특성을 보였으며, 기초학력 미달의 경우, 학업성취도 평가 결과 대비 교사 예측 성취수준 비율 또한 약 5배의 차이가 나타났다. 수학과 학업성취도 평가 결과와 교사 예측 성취수준 간의 일치 정도는 59.2%로 나타났다. 이를 성취수준별로 살펴보면, 우수학력이 585명(31.9%), 보통학력에서 332명(18.1%), 기초학력에서 146명(8.0%), 그리고 기초학력 미달에서 23명(1.3%)의 분포가 나타났다.

요약하면, 국어과 및 수학과에 대한 학생 성취수준의 진단이라는 측면에서, 2012 학업성취도 평가와 비교하여 교사는 자신이 가르치는 학생에 대하여 우수학력과 기초미달을 과대하게 예측하며, 보통학력 비율을 과소 평가하는 특성을 보인다. 특히, 기초학력 보장이라는 학업성취도 평가의 주요 목표를 고려할 때 연구 결과의 교사 예측에서 나타난 기초학력 미달 학생의 높은 비율은 학업성취도 평가 결과의 신뢰성과 교육적 활용에 부정적인 영향을 미칠 가능성이 있다.

2. 국가수준 학업성취도 평가 결과의 활용

학업성취도 평가 목적에 기반한 평가 결과 활용에 대한 교사 설문은 총 7개 영역으로 이루어져 있다. ‘학업 성취수준 파악’, ‘교수·학습 방법 개선’, ‘학업 성취 추이 분석’, ‘학업성취 배경변인 분석’, ‘교육정책 및 장학’, ‘평가 전문성 신장’, ‘교육과정 정착’의 영역별로 교사들의 학업성취도 평가 결과 활용에 대한 실태를 조사한 결과는 〈표 11〉과 같다. 각 영역별로 문항들의 평균 점수는 2.8~3.2로 뚜렷한 차이 없이 중간(보통이다) 정도로 나타났다.

〈표 11〉 학업성취도 평가 결과의 활용(교사 인식)

빈도(%)

영역	문항	전혀 그렇지 않다(1점)	그렇지 않다(2점)	보통이다(3점)	그렇다(4점)	매우 그렇다(5점)	평균
학업 성취 수준 파악	3	1.0(1.3%)	7.0(9.2%)	26.0(34.2%)	39.0(51.3%)	3.0(3.9%)	3.5
	4	3.0(3.9%)	17.0(22.4%)	25.0(32.9%)	23.0(30.3%)	5.0(6.6%)	3.0
	15	3.0(3.9%)	14.0(18.4%)	28.0(36.8%)	31.0(40.8%)	0.0(0.0%)	3.1
	19	4.0(5.3%)	25.0(32.9%)	31.0(40.8%)	15.0(19.7%)	0.0(0.0%)	2.7
	평균	2.8(3.6%)	15.8(20.7%)	27.5(36.2%)	27.0(35.5%)	2.0(2.6%)	3.1

영역	문항	전혀 그렇지 않다(1점)	그렇지 않다(2점)	보통이다(3점)	그렇다(4점)	매우 그렇다(5점)	평균
교수·학습 방법 개선	1	7.0(9.2%)	27.0(35.5%)	18.0(23.7%)	21.0(27.6%)	3.0(3.9%)	2.8
	2	5.0(6.6%)	25.0(32.9%)	20.0(26.3%)	22.0(28.9%)	2.0(2.6%)	2.8
	13	3.0(3.9%)	14.0(18.4%)	33.0(43.4%)	25.0(32.9%)	1.0(1.3%)	3.1
	26	5.0(6.6%)	18.0(23.7%)	37.0(48.7%)	12.0(15.8%)	4.0(5.3%)	2.9
	27	2.0(2.6%)	11.0(14.5%)	21.0(27.6%)	38.0(50.0%)	4.0(5.3%)	3.4
	평균	4.4(5.8%)	19.0(25.0%)	25.8(33.9%)	23.6(31.1%)	2.8(3.7%)	3.0
학업 성취 추이 분석	6	6.0(7.9%)	16.0(21.1%)	25.0(32.9%)	27.0(35.5%)	2.0(2.6%)	3.0
	20	9.0(11.8%)	24.0(31.6%)	22.0(28.9%)	18.0(23.7%)	2.0(2.6%)	2.7
	25	7.0(9.2%)	15.0(19.7%)	29.0(38.2%)	22.0(28.9%)	2.0(2.6%)	2.9
	평균	7.3(9.6%)	18.3(24.1%)	25.3(33.3%)	22.3(29.4%)	2.0(2.6%)	2.9
학업 성취 배경변인 분석	9	11.0(14.5%)	27.0(35.5%)	24.0(31.6%)	13.0(17.1%)	0.0(0.0%)	2.5
	11	1.0(1.3%)	3.0(3.9%)	15.0(19.7%)	44.0(57.9%)	13.0(17.1%)	3.9
	12	1.0(1.3%)	5.0(6.6%)	36.0(47.4%)	33.0(43.4%)	1.0(1.3%)	3.4
	18	7.0(9.2%)	28.0(36.8%)	31.0(40.8%)	10.0(13.2%)	0.0(0.0%)	2.6
	평균	5.0(6.6%)	15.8(20.7%)	26.5(34.9%)	25.0(32.9%)	3.5(4.6%)	3.1
교육정책 및 장학	7	9.0(11.8%)	28.0(36.8%)	31.0(40.8%)	8.0(10.5%)	0.0(0.0%)	2.5
	21	1.0(1.3%)	13.0(17.1%)	27.0(35.5%)	32.0(42.1%)	2.0(2.6%)	3.2
	22	11.0(14.5%)	19.0(25.0%)	24.0(31.6%)	20.0(26.3%)	1.0(1.3%)	2.7
	23	0.0(0.0%)	17.0(22.4%)	37.0(48.7%)	22.0(28.9%)	0.0(0.0%)	3.1
	평균	5.3(6.9%)	19.3(25.3%)	29.8(39.1%)	20.5(27.0%)	0.8(1.0%)	2.9
평가 전문성 신장	8	3.0(3.9%)	22.0(28.9%)	34.0(44.7%)	17.0(22.4%)	0.0(0.0%)	2.9
	14	7.0(9.2%)	30.0(39.5%)	20.0(26.3%)	17.0(22.4%)	1.0(1.3%)	2.6
	24	3.0(3.9%)	24.0(31.6%)	25.0(32.9%)	23.0(30.3%)	1.0(1.3%)	2.9
	평균	4.3(5.7%)	25.3(33.3%)	26.3(34.6%)	19.0(25.0%)	0.7(0.9%)	2.8
교육과정 정착	5	8.0(10.5%)	32.0(42.1%)	16.0(21.1%)	15.0(19.7%)	5.0(6.6%)	2.7
	10	0.0(0.0%)	2.0(2.6%)	8.0(10.5%)	43.0(56.6%)	22.0(28.9%)	4.1
	16	1.0(1.3%)	36.0(47.4%)	26.0(34.2%)	11.0(14.5%)	1.0(1.3%)	2.6
	17	2.0(2.6%)	11.0(14.5%)	29.0(38.2%)	33.0(43.4%)	0.0(0.0%)	3.2
	평균	2.8(3.6%)	20.3(26.6%)	19.8(26.0%)	25.5(33.6%)	7.0(9.2%)	3.2

학업성취도 평가 결과 활용에 대한 설문지의 주요 결과는 다음과 같다. 첫째, 학생의 학업 성취 수준 파악 관련 설문에 대하여 평균 38.1%의 교사들이 학업성취도 평가 결과를 활용한다고 응

답했다. 특히 교사 자신의 평가 결과와 학업성취도 평가 결과를 통합하여 학업 성취를 진단한다고 응답한 비율이 55.2%로 나타났다(문항 3). 또한, 학생들의 학업성취도 평가 결과와 교사의 교수·학습 능력과의 관련성에 대해서는 긍정적인 응답이 19.7%, 부정적인 응답은 38.2%로 학업성취도 평가 결과에 대한 교사 영향력을 높게 평가하지 않았다(문항 19). 둘째, 교수·학습 방법 개선에 대한 설문에 대하여 평균 34.8%의 교사들이 학업성취도 평가 결과를 활용한다고 응답했다. 특히 학업성취도 평가 결과에 대하여 동학년 수준, 또는 학교 수준에서 협의가 이루어진다고 응답한 비율이 55.3%로 나타났다(문항 27). 그리고 31.5%의 교사들이 학업성취도 평가 시행으로 인해서 교육 활동이 위축된다는 반응을 하였다(문항 1). 셋째, 학업 성취 추이 분석에 대한 정보 활용에 대한 설문에 대하여 평균 32%의 교사들이 긍정적으로 응답했다. 특히 학교별 학업성취도 평가 결과가 공시되는 ‘학교알리미’ 사이트를 통해서 38.1% 교사들이 학생들의 학업 성취 추이 파악에 관심을 두는 것으로 나타났다(문항 6). 한편, 주변 학교들에 대한 학업 성취 추이 정보에 관심을 두는 교사는 26.3%로 비교적 낮게 나타났다(문항 20). 이는 교사들이 ‘학교알리미’ 사이트의 학교수준 성취도 정보에 관심을 가지고 있으나, 주변 학교와의 비교를 위한 정보로 활용하는 비율은 낮은 것으로 판단된다. 넷째, 학업 성취 배경변인에 대한 정보를 활용하느냐는 설문에 대하여 평균 37.5%의 교사들이 긍정적으로 응답했다. 특히 근무 여건 등 학교 특성에 대해서는 높은 관심을 보인다(문항 9, 18). 교수·학습 관련 정보로서 배경변인의 활용도는 낮은 것으로 나타났다(문항 11, 12). 이러한 결과는 교사들이 학업성취도 평가와 병행하여 실시되는 학업 성취 배경변인 정보의 가치를 낮게 인식하고 있음을 보여준다. 다섯째, 교육정책 및 장학에 대한 정보를 활용하느냐는 설문에 대하여 평균 28.0%의 교사들이 긍정적으로 응답했다. 기초학력 보장을 위한 지원의 효과성에 대해서는 44.7%의 교사들이 긍정적인 평가를 하는 것으로 나타난(문항 21) 반면, 학업성취도 평가 결과의 ‘학교알리미’ 사이트 공개에 대해서는 10.5%의 교사들만이 학교의 교육력 향상에 도움이 된다고 응답했다(문항 7). 여섯째, 학업성취도 평가와 평가 전문성 신장과의 관계를 묻는 설문에 대해서는 평균적으로 25.9%의 교사들만이 긍정적으로 응답했다. 학업성취도 평가를 통해 문항 개발의 전문성이 신장되었는지에 대한 설문에 31.6%의 교사가 긍정적으로 응답했다(문항 24). 이러한 결과는 전반적으로 교사들은 자신의 평가 전문성에 대해서 주관적 자신감을 가지고 있으나, 학업성취도 평가를 통해서 평가 전문성을 개선해야 한다는 의식은 높지 않을 것으로 판단된다. 일곱째, 교육과정에 미치는 영향에 대한 설문에 대하여 42.8%의 교사들이 긍정적으로 응답했다. 특히 학업성취도 평가 시행과 무관하게 학급 교육과정을 운영하는 비율이 85.5%에 이르는 것으로(문항 10) 볼 때, 학업성취도 평가 시행으로 인해서 교사의 부담이 과중되거나 파행적인 교육과정 운영 등 부정적인 특성이 발생할 가능성은 높지 않은 것으로 나타났다.

V. 결론 및 논의

학업성취도 평가로부터 도출된 국가 교육지표들은 교육의 질적 향상을 시사하지만(교육과학기술부, 2011, 2012), 학업성취도 평가의 목적이 달성되고 있는지에 대한 판단은 국가수준의 거시적 교육지표와 더불어 단위학교 수준의 실증적 자료에 대한 분석 등을 근거로 이루어져야 한다(성태제, 1999; 이원석, 2008, 2009). 이 연구는 학업성취도 평가의 결과타당도를 단위학교 수준에서 분석하기 위하여, 서울시 초등학교 교사를 대상으로 성취수준 진술문의 분류 일치도, 교사 예측 성취수준과 학업성취도 평가 결과의 일치도, 현장 학교에서의 학업성취도 평가 목적 구현 수준을 조사했다. 이에 따른 주요 연구 결과는 다음과 같다.

첫째, 학업성취도 평가의 성취수준에 대해 교사의 성취수준 진술문 분류 일치율은 수학과가 국어과보다는 상대적으로 높았지만 전반적으로 낮은 수준으로 나타났다. 이러한 결과의 원인 중 하나는 최근 4년간의 국어과 및 수학과 성취수준 진술문은 대부분 해마다 변경되었고, 일부의 경우에는 동일한 진술문에 시행 시기별로 성취수준이 달리 적용되어 나타난 현상으로 판단된다.

둘째, 국어과와 수학과 모두 학업성취도 평가의 성취수준과 교사의 예측 성취수준 간에 유사한 경향이 우수학력 수준에서는 나타났지만 기초학력 미달에서는 교사의 평가와 학업성취도 평가 결과 간에는 명확한 차이점이 나타났다. 이는 교사가 학업성취도 평가에 비해서 성취수준의 기준점을 최하위 집단에서는 높게 설정한 결과라고 생각된다. 특히, 교사들은 학생 성취수준에 대한 자신의 판단에 대해 강한 확신을 갖고 있으며, 이는 실제 학업성취도 평가 결과와 교사의 예측 성취수준 간의 일치율 및 성취수준 진술문 분류 일치율을 낮추는 요인으로 판단된다.

셋째, 학업성취도 평가 목적이 단위학교 교사 수준의 교육 활동 속에서 구현되는 정도에 대하여, 전반적으로 교사들은 초등학교 학업성취도 평가의 목적 구현에 대하여 부정적 혹은 중립적 의견을 보였다. 또한 학업성취도 평가 결과에 대해서는 큰 부담을 갖고 있지 않으며, 교육 정보로서의 학업성취도 평가 결과의 교육적 활용 가치를 높게 인식하지 않는 것으로 나타났다.

결국, 교사의 교육 활동 측면(객관적 성취기준 설정, 학생의 성취수준 진단, 교육과정에서 평가 결과 활용 등)에서 볼 때, 초등학교 학업성취도 평가의 교육적 활용도(결과타당도)는 재고가 필요한 것으로 나타난다. 신혜진(2014)은 초등학교 학업성취도 평가 자료가 단위학교 수준에서 긍정적으로 활용될 경우 학력 향상에 효과적이라고 보고하였는데, 이러한 측면을 고려할 때 기초학력 보장이라는 학업성취도 평가 목적의 달성을 위해서는 교사의 교육 활동에서 학업성취도 평가 결과 활용도를 높일 필요가 있다. 이를 위해서는 성취수준 기술문 등 평가 결과의 내용에 대한 지속적인 개선이 필요한데, 먼저 학업성취도 평가의 주요 평가 결과로 제시되는 성취수준을 정의하는 성취수준 진술문에 대한 명확한 재정립이 필요하다. 2009-2011 성취수준 진술문에 대한 교사의 분류 일치율은 국어과가 보통학력, 우수학력, 기초학력 순으로 46.2%,

30.9%, 18.4%였고, 수학의 경우는 63.6%, 41.1%, 21.9%로 낮았다(〈표 8〉 참조). 성취수준 진술문에 대한 교사의 분류 일치 정도가 낮다는 점은 성취수준의 구체적인 준거 내용에 대한 교사의 이해가 학업성취도 평가와 차별적임을 보여주는 결과이며, 학업성취도 평가 결과에 대한 교사들의 해석이 실제 학생들의 특성을 정확하게 반영하지 못한다는 것이다. 학생들에게 적절한 학습 처치는 학생의 학업 성취에 대한 정확한 진단에 근거할 때 가능하다는 점을 고려하면 학업성취도 평가의 결과타당도 확보를 위하여 학업성취도 평가의 성취수준에 대한 교사의 이해를 높이는 것이 필요하다. 특히 일부 동일한 진술문이 평가 시행 연도별로 제각각 다른 성취수준으로 분류되는 상황은 성취수준의 의미를 모호하게 한다.

또한, 교사와 학업성취도 평가 간의 성취수준을 결정하는 기준점에 대한 관점이 다르다는 점에 주목할 필요가 있다. 비록 이 연구에서 연구 대상으로 선정된 교사가 전국수준의 모집단을 완벽히 대표할 수는 없지만, 연구에서 활용된 교사의 교육 경력을 고려할 때 교사의 전문성을 확보했다고 할 수 있다. 따라서 이 연구 결과 중 개별 학생에 대한 학업성취도 평가의 성취수준과 교사 예측 성취수준 간의 비교에서 나타나는 차이는 성취수준을 설정하는 기준점의 차이가 존재할 가능성이 높음을 보여주고 있다(〈표 9〉와 〈표 10〉 참조).

즉, 국가수준 평가와 교사의 평가 사이에 차이가 발생한 원인과 타당한 기초학력 미달 기준에 대해서는 충분한 논의가 요구된다. 예를 들면, 학업성취도 평가에서는 국가 차원의 지원 대상이 되는 학생들만을 선별하기 위해 최소한의 성취기준을 설정하고 있는 반면, 학교 현장 교사들은 전반적인 성취도 및 학습 태도, 향후의 향상 가능성 등을 고려하여 학습 지원의 대상을 확대하려고 기준점을 가능한 높게 설정할 수 있다. 그러므로 학업성취도 평가와 교사의 기초학력 기준 설정 과정에 어떠한 관점의 차이가 존재하며, 어떻게 조율될 수 있을 것인지는 향후 더 구체적으로 검토되어야 한다.

부가적으로 초등학교 학업성취도 평가의 폐지에 대한 논의는 보다 실증적인 근거에 기반을 둔 숙의가 필요하다(김경성, 2013; 남현우, 2014). 초등학교 학업성취도 평가 폐지의 근거로 제시한 시험부담 및 과열경쟁 등은 결과적으로 일선 학교 현장의 교육과정 파행 운영 등을 야기할 가능성이 높다(교육부, 2013). 그러나 실제로 교사들은 전반적으로 학업성취도 평가에 대해서 강한 부담감을 갖거나 학교 간 경쟁의 주요 근거 자료로 생각하지 않으며, 학업성취도 평가 시행과는 무관하게 교육과정을 운영한다고 응답한 비율도 매우 높았다(〈표 11〉 참조). 물론 연구 결과에서 언급한 바와 같이, 학업성취도 평가가 제공하는 정보의 가치와 활용 가능성에 대해서 교사들이 낮게 평가하고 있으며, 학업성취도 평가의 결과타당도에 대해서 부정적이거나 중립적인 의견을 보이는 등의 학업성취도 평가 시행 보류를 지지하는 결과도 있다. 이러한 상반된 결과는 그만큼 초등학교 학업성취도 평가의 지속 여부는 보다 실증적인 검토와 다각적인 관점에 바탕을 둔 합의가 필요함을 의미한다.

참 고 문 헌

- 교육과학기술부(2011). 2011년 국가수준 학업성취도 평가 기본계획. 교육과학기술부 교육정보기획과-646, 2011.2.17. 서울: 교육과학기술부.
- 교육과학기술부(2012). 2012년 국가수준 학업성취도 평가 기본계획. 교육과학기술부 교육정보기획과-4940, 2011.12.29. 서울: 교육과학기술부
- 교육부(2013). 2013년 국가수준 학업성취도 평가 기본계획 발표. 2013.4.23. 교육부 보도자료.
- 김경성, 김도기, 김성식, 김준엽, 상경아, 양길석(2013). **국가수준 학업성취도 평가 개선 연구**. 교육부 수탁과제 2013-1.
- 김경희, 김성숙, 시기자, 김수진, 김완수, 신진아, 박인용, 구남옥, 구슬기, 노은희, 이인호, 김도남, 김부미, 우석진(2014). 국가수준의 기초학력 점검을 위한 초등학교 학업성취도 평가 방안. 한국교육과정평가원. 연구보고 CRE 2014-2.
- 김경희, 김환수, 최인봉, 상경아, 김희경, 신진아, 김준엽, 손원숙(2011). 국가수준 학업성취도 평가에 나타난 우리나라 학력 향상의 특성 분석. 한국교육과정평가원. 연구보고 RRE 2011-2-4.
- 김경희, 송미영, 김완수, 최인봉, 신진아, 박인용, 김종훈, 김성훈(2012). **2011 국가수준 학업성취도 평가 결과: 초등학교 학업성취도 변화 추이 분석**. 서울: 한국교육과정평가원.
- 김성숙, 김도남, 김소영, 송미영, 이봉주, 최인봉, 김성식, 양성관(2009). 국가수준 학업성취도 평가의 전수시행에 따른 교육맥락 변인 탐색과 설문지 개발. 한국교육개발원, 연구보고 RRE 2009-1.
- 김성숙, 송미영, 최인봉, 김희경, 김성훈(2010). 2009년 국가수준 학업성취도 평가 전수분석 결과 -초등학교 6학년-. 한국교육과정평가원, 연구보고 RRE 2010-7-2.
- 김신영, 강태훈, 김정민, 양길석(2011). 학교평가에서의 대규모 학업성취도 평가 결과 활용 방안. **교육평가연구**, 24(1), 1-30.
- 남현우 (2014, 12월). **국가수준 학업성취도 평가 체제 재구조화 방안**. 국가수준 학업성취도 평가 체제 재구조화 방안을 위한 공청회, 국가수준 학업성취도 평가 개선 연구팀.
- 민경석(2010). 국가수준 학업성취도 평가-외국 사례와 비교-. 한국교육개발원. 연구보고 OR 2010-04-3.
- 민경석(2013). 한국과 미국 국가수준 학업성취도 평가의 성취수준 비교. **중등교육연구**, 6(1), 111-136.
- 박정, 김경희, 김수진, 손원숙, 송미영, 조지민(2006). 국가수준 학업성취도 평가 -기술보고서-. 교육과정평가원, 연구보고 PRO 2006-4.

- 서울시교육청(2011). **국가수준 학업성취도평가 교감 및 담당 부장 연수 자료**. 서울: 서울시교육청.
- 성태제(1999). 교육평가 방법의 변화와 결과타당도에 대한 고려. **교육학연구**, 37(1), 197-218.
- 성태제(2003). 검사나 평가활동에 대한 메타평가적 관점에서의 결과타당도. **교육학연구**, 41(1), 91-110.
- 신혜진(2014). 초등학교 국가수준 학업성취도평가자료 활용정책. **교육평가연구**, 27(5), 1157-1188.
- 이상하, 김수진, 김완수, 신선희, 상경아(2010). 국가수준 학업성취도 평가의 설문지 모음집 - 2003년~2009년-. 한국교육과정평가원, 연구보고 RRE 2010-12-5.
- 이원석(2008). 타당도에 대한 통합적 관점. **교육평가연구**, 21(4), 67-79.
- 이원석(2009). 타당도 분석을 위한 평가적 접근 모형: 통합적 관점. **교육평가연구**, 22(4), 1079-1099.
- 정은영, 최인봉, 기희경, 기소영, 유진은(2009). 국가수준 학업성취도 평가 결과추이(2003년~2008년) -초등학교6학년-. 한국교육과정평가원, 연구보고 RRE 2009-8-1.
- 좌혜정, 성태제 (2003). 2002학년도 대학수학능력시험 언어, 수리 영역 결과타당도 분석과 2003학년도 출제 방향. **교육학연구**, 41(1), 329-361.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed). *Educational measurement* (4th ed., pp. 17-64). Washington, DC: American Council on Education/Praeger.
- Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice*, 21(1), 23-30.
- Linn, R. L. (1997). Evaluating the validity of assessments: the consequences of use. *Educational Measurement: Issue and Practice*. 16(2), 14-15.
- Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practice*, 16(2), 16-18.
- Mertler, C. (2011). Teachers' Perceptions of the Influence of No Child Left Behind On Classroom Practices. *Current Issues in Education*, 13(3). Retrieved from <http://cie.asu.edu/>
- Messick, S (1989). Validity, In R. L. Linn(Ed). *Educational Measurement(3rd ed)*, Washington D.C: American council on education & National council on Measurement in education. 13-103.
- Messick, S. (1990). *Validity of Test Interpretation and Use*. Princeton, NJ: Educational Testing Service.

- Popham, W. J. (1997). Consequential validity: right concern-wrong concept. *Educational Measurement: Issues and Practice*, 16(2), 9-13.
- Shepard, L. A. (1997). The Centrality of test use of consequences for test validity. *Educational Measurement: Issue and Practice*. 16(2), 5-8.

· 논문접수 : 2015-01-08/ 수정본접수 : 2015-02-05/ 게재승인 : 2015-02-23

ABSTRACT

Consequential Validity of National Assessment of Educational Achievement in Elementary Schools

Jae-Min Cho

(Teacher, Seoul NamCheon Elementary School)

Kyung-Seok Min

(Associate Professor, Sejong University)

In-Yong Park

(Associate Research Fellow, Korea Institute for Curriculum and Evaluation)

The purpose of this study was to examine the consequential validity of NAEA(National Assessment of Educational Achievement) in terms of school practices. To accomplish the research purposes, the consistency of teacher's classification of NAEA achievement level descriptions, the accuracy of predicted achievement levels for each student, and teachers' application of NAEA results for teaching were evaluated. The research results were found that 1) teachers categorized achievement level descriptions somewhat differently from NAEA, 2) teachers expected more below-basic level students compared with NAEA, and 3) NAEA results were far less used than expected in school practices. Also, teachers kept focusing on the national curriculum regardless of the implementation of NAEA. In sum, we found the impacts of NAEA on schools were not in the intended directions and degrees, in terms of consequential validity defined by Messick. Based on research results, we discussed issues on the recent cessation of Elementary School NAEA and suggested several research points for further studies.

Key Words : National assessment of Educational Achievement(NAEA), consequential validity, achievement level