

수행평가 정착 방안 : 방법 및 도구의 질 관리

白 淳 根*

1. 들어가는 말

교육부가 '98년에 발표한 '교육비전 2002: 새 학교문화 창조' 안(案)에서 강조하고 있는 핵심 내용 중의 하나는 우리나라 교육의 다양화·전문화·특성화를 지향하는 것이다(교육부, 1998). 즉, 교수·학습 방법 및 평가 방법의 다양화·전문화·특성화와 그에 따른 학생 선발 방식의 다양화·전문화·특성화를 지향하고 있다. 이러한 맥락에서 학생 평가의 다양화·전문화·특성화를 지향하고 있는 수행평가(遂行評價: performance assessment) 정책의 도입은 단순히 평가 방식만의 변화를 의미하는 것이 아니라 우리나라 교육 체제를 재구조화하기 위한 종합적인 시도 중의 하나라고 할 수 있다.

그리고, 지역의 실정이나 학교의 교육 여건에 따라 담당 교사가 자율적으로 수행평가를 시행하도록 한다는 원칙 하에, 가능한 한 수행평가를 확대 실시하고자 한 교육부의 수행평가 정책은, 단순히 학생평가 방법만의 변화를 의미하는 것이 아니라 21세기 지식·정보화 시대를 맞이하여 앞으로 우리나라의 교육정책 혹은 교육평가 정책이 나아가야 할 방향을 제시하는 하나의 새로운 관점(觀點)이라 할 수 있다(백순근, 2000).

그 동안 수행평가 정책의 추진 과정을 살펴보면, 교육부는 '98년 10월에 '교육비전 2002: 새 학교문화 창조'안을 발표하면서 '99년 3월부터 전국의 모든 초·중·고등학교에서 해당 학교의 여건이나 교과목의 특성을 고려하여 수행평가를 부분적으로 도입·실천하도록 하였다. 그리고, 교육부가 '99년 3월에 발간한 '99년도 시·도 교육청 평가 편람에서, 비록 적은 점수(600점 만점에서 12점이 배당됨)이기는 했지만, '평가방법의 다양화·전문화를 위한 노력'이라는 항목으로 ①수행평가 확대 실시를 위한 기본계획의 적절성 및 환류 실적, ②수행평가 관련 교원 연수의 적절성, ③수행평가 관련 (연수)자료 개발 실적, 그리고 ④수행평가 관련 학부모에 대한 홍보의 적절성 등을 통해 평가하도록 하여, 16개 시·도 교육청에서는 경쟁적으로 수행평가를 확대 실시하기 위해 노력하였다(시·도교육청평가위원회, 1999a, b). 참고로, '99년도에 수립한 K교육청의 장기계획안에 따르면, 개별 학교의 실정이나 교과목의 특성은 고려하지 않고, 99년에는 30%이상, 2000년에는 40%이상, 2001년에는 50%, 그리고 2002년에는 100%까지 수행평가를 확대 실시하도록 하고 있었다.

* 서울대학교 사범대학 교육학과

아울러, '99년 4월 교육부 훈령 제587호로 공포된 '학교생활기록부 전산처리 및 관리 지침'에는 학생들의 교과성적을 산출할 때 수행평가 결과를 포함시키도록 명시하였다. 요컨대, 교육부의 강력한 의지에 따라 '99년 3월부터 모든 학교에서, 비록 정도의 차이는 있지만, 학교생활기록부에 기록하기 위한 교과목별 성적을 산출할 때 기존의 선택형 검사의 결과와 함께 수행평가의 결과를 반영하기 시작하였다. 특히, 2002학년도 대학입학제도 개선안에서, 무시험 전형이 강조되고, 수행평가의 결과를 포함하는 고등학교 내 신성적이 대학입사에서 큰 비중을 차지하게 될 것이라는 예측과 함께 수행평가는 교사들을 포함하여 모든 국민들의 중요한 관심사가 되었다.

비교적 짧은 기간이었지만 교육부나 시·도교육청 등의 적극적인 홍보와 연수를 통해 대부분의 교사들이 21세기 지식·정보화 시대를 맞이하여 수행평가가 필요하다는 것을 인정하고 있으며, 가능한 한 수행평가를 확대 실시하기 위해 성실하게 노력하고 있다. 따라서 이제는 수행평가를 확대 실시해야 한다는 등의 원칙론적인 주장보다는, 담당 교사들이 수행평가를 제대로 시행할 수 있도록 학교 교육 여건을 개선하고 다양한 형태의 수행평가 도구들을 개발·보급하는 등 수행평가의 정착을 위한 실제적이면서도 구체적인 노력을 지속적이면서도 체계적으로 기울여야 할 때라고 할 수 있다(한국교육평가학회 외, 2000).

그러므로, 이 연구에서는 수행평가를 학교 교육현장에 정착시키기 위한 노력의 일환으로 수행평가 방법이나 도구의 질(質)을 관리하기 위한 방안에 대해 논의하고자 한다. 즉, 개별 교사가 수행평가 방법이나 도구를 선정하거나 직접 수행평가 도구를 개발할 때 그 질을 확보하기 위한 방안에 대해 논의하고자 하는 것이다. 특히 이 연구에서는 주어진 여건과 상황 속에서 개별 교사가 수행평가를 실시하기 위해 평가 방법이나 평가 도구를 선정할 때 고려해야 할 사항과 평가 도구를 개발하기 위한 구체적인 절차에 대하여 논의하고자 한다.

결국 수행평가 방법 및 도구를 선정하거나 직접 개발할 때 그 질을 관리하기 위한 이러한 논의는 학교 교육 현장에서 수행평가를 제대로 정착시키는 데 기여하게 될 것이다.

2. 수행평가 방법 및 도구의 선정 시 고려해야 할 사항

학교 교육 현장에서 수행평가를 제대로 실시하기 위해서는 학교의 교육 여건이나 실정, 그리고 교과목이나 교육 내용의 특성 등을 고려함과 동시에 그에 적절한 평가 방법 및 도구를 활용해야 한다. 다양한 형태의 수행평가 방법 중에서 어떠한 방법을 활용할 것인가를 결정하고 그에 적절한 평가 도구를 제작하기 위해서 고려해야 할 사항들은 다음과 같다¹⁾(백순근, 2000; Downing & Haladyna, 1997; Hogan, Benjamin & Brezinski, 2000; Linacre, 2000; Spalding, 2000).

첫째, 수행평가의 본질적인 목적이나 취지에 적합한 방법이나 도구이어야 한다. 즉 적합도(properness)가 높아야 한다. 둘째, 평가하고자 하는 구체적인 목표나 내용을 제대로 평가할 수 있는 방법이나 도구이어야 한다. 즉 타당도(validity)가 높아야 한다. 셋째, 평가 결과의 정확성과 일관성을 보장할 수 있는 방법이나 도구이어야 한다. 즉 신뢰도(reliability)가 높아야 한다. 넷째, 주어진 여건과 상황에서 실행할 수 있는 적절한 방법이나 도구이어야 한다. 즉, 실용도(usability)가 높아야 한다. 이에 대해 부연하여 설명하면 다음과 같다.

가. 적합도

평가 방법이나 평가 도구의 적합도(properness)란 학생들의 지식, 기능, 태도 등을 평가함에 있어서 특정한 평가의 본질적인 목적이나 취지에 적합한 정도를 의미하며, 이 글에서는 수행평가의 본질적인 목적이나 취지에 적합한 정도를 의미하는 것으로 사용하고 있다. 수행평가의 본질적인 목적이나 취지는 ①학생 개개인의 성장·발달을 돕기 위해 교수·학습 활동을 개선하고 지도·조언·충고하기 위한 것, ②교사의 전문적인 관찰과 판단을 중시하고 창의성이나 문제 해결력 등 고등 사고기능을 신장하기 위한 것, ③주어진 선택지 중에서 답을 선택하게 하는 것이 아니라 자기 스스로 답을 작성(서술 혹은 구성)하거나 행동으로 나타내도록 하여 인지적으로 '아는 것'뿐만 아니라 실제로 '적용할 수 있는 것', 즉 '할 줄 아는 능력'을 신장하기 위한 것 등이다. 이러한 수행평가의 목적이나 취지에 어느 정도 적합한지를 밝히는 것은 어디까지나 정도의 문제이기 때문에 적합도가 높거나 낮다고 할 수는 있지만, 있다거나 없다고 하기는 어렵다.

참고로 적합도가 높은 경우의 예를 들면, 학생의 말하기 능력을 평가하기 위해서는 선택형 시험을 통해 간접적으로 측정하는 방법보다는 실제로 말할 기회를 제공하여 직접적으로 측정하는 말하기 실기 시험이 좀더 적합하다고 하겠다. 그리고, 고등 사고기능의 습득여부를 파악하지 않고 간단한 지식이나 정보의 습득여부만을 파악하거나 전문적인 관찰이나 판단과정을 포함함이 없이 컴퓨터를 이용하여 기계적으로 채점하는 것 등은 적합도가 낮은 경우라고 할 수 있다(Spalding, 2000).

-
- 1) 여기에서 평가 방법이나 도구의 '객관도(objectivity)'를 언급하지 않는 이유는, 수행평가와 같은 질적 평가(혹은 충고형 평가, 즉 개별 학생에게 지도·조언·충고하기 위한 평가)에서는 '실시 방법이나 채점의 객관성'보다는 '상황과 맥락에 따른 적합성이나 전문성'을 더욱 강조하기 때문이다. 동시에 양적 평가(혹은 선발형 평가, 즉 학생들을 선발·분류·배치하기 위한 평가)에서는 '채점자간 일치도'를 주로 객관도의 범주에 포함시키고 있으나, 수행평가와 같은 질적 평가(혹은 충고형 평가)에서는 신뢰도의 개념에 포함시킬 수 있기 때문이다.

나. 타당도

평가 방법이나 평가 도구의 타당도(validity)란 평가하고자 의도했던 구체적인 목표나 내용을 제대로 평가하고 있는가를 나타내는 정도를 의미한다. 예컨대, 어떤 평가 방법이나 도구를 활용하여 학생의 쓰기 능력을 알아보고자 했을 경우, 그 방법이나 도구를 이용하여 쓰기 능력을 제대로 파악할 수 있었다면 타당도가 높은 평가 방법이나 도구라고 할 수 있겠지만, 반대로 그 방법이나 도구를 이용하여 쓰기 능력을 제대로 파악할 수 없었다면 타당도가 낮은 평가 방법이나 도구라고 할 수 있다. 앞서 언급한 적합도가 수행평가의 본질적인 목적이나 취지를 제대로 살릴 수 있는지의 문제라면, 타당도는 평가하고자 하는 구체적인 목표나 내용을 제대로 평가할 수 있는지의 문제라고 할 수 있다. 타당도도 적합도와 유사하게 어디까지나 정도의 문제이기 때문에 타당도가 높다거나 낮다고 할 수는 있지만, 있다거나 없다고 하기는 어려우며 대부분의 경우 숫자로 표시하기 어려운 경우가 많이 있다.

평가 방법이나 도구의 타당도를 알아보기 위한 방식에는 여러 가지가 있는데, 어떠한 측면을 강조하느냐에 따라 안면타당도, 내용타당도, 구인타당도, 공인타당도, 예언타당도, 체제적 타당도(혹은 결과타당도), 생태학적 타당도 등으로 분류할 수 있다(백순근, 2000; 성태제, 1999; Feldt & Qualls, 1999). 그러나 이러한 분류는 어디까지나 타당도를 다양한 방법으로 파악하기 위한 기능상의 구분이라 할 수 있으며, 타당도를 제대로 파악하기 위해서는 가능한 한 다양한 방법으로 통합적이면서도 전체적인 정보를 수집하는 것이 바람직하다. 주어진 상황이 동일하다면 평가도구의 타당성을 확보할 수 있는 증거는 많으면 많을수록 좋으며, 증거의 양도 중요하지만 증거의 질이 더욱 중요하다는 것도 인식할 필요가 있다.

다. 신뢰도

평가 방법이나 평가 도구의 신뢰도(reliability)란 그 평가 방법이나 도구를 이용하여 수집한 검사의 점수가 얼마나 정확하고 일관성이 있는가 하는 정도, 즉 측정의 오차(measurement error)가 얼마나 적은가 하는 정도를 의미한다. 예컨대, 어떤 평가 방법이나 도구를 이용하여 평가한 결과를 아침에 채점한 점수와 저녁에 채점한 점수가 동일하다면 신뢰도가 높은 평가 방법이나 도구라 할 수 있겠지만, 반대로 아침에 채점한 점수와 저녁에 채점한 점수간에 차이가 크다면 신뢰도가 낮은 평가 방법이나 도구라고 할 수 있다. 이러한 신뢰도는 어디까지나 정도의 문제이기 때문에 신뢰도가 높다거나 낮다고 할 수는 있지만 있다거나 없다고 하기는 어려우며, 타당도와는 달리 대부분의 경우 숫자(혹은 계수)로 표시하는 경우가 많다.

그리고, 수행평가와 같은 질적 평가(혹은 충고형 평가)의 경우에는 교사의 전문적인 관찰과 판단을 중시하기 때문에 신뢰도의 중요성이 양적 평가(혹은 선발형 평가)에서보다는 덜 강조되는 경향이 있다. 특히 학생 개개인의 성장·발달을 돕기 위해서 교수·학습 활동을 개선하고 지도·조언·충고하기 위해 수행평가를 시행하는 경우라면 평가 방법이나 도구의 적합도와 타당도를 확보하는 것이 더욱 중요하다는 사실은 지극히 당연한 일이다²⁾. 그런데, 학교 교육 현장에서는 지도·조언·충고를 목적으로 수행평가를 실시하기도 하지만, 학생들을 선발·분류·배치하기 위한 목적으로 수행평가를 사용할 가능성도 많이 있다. 후자의 경우처럼 학생들을 선발·분류·배치하기 위해 수행평가를 실시할 경우에는 평가 방법이나 도구의 신뢰도도 제대로 확보해야 할 것이다.

평가 방법이나 도구의 신뢰도를 알아보기 위한 방식에는 여러 가지가 있는데, 어떠한 방법과 절차를 이용하여 파악하느냐에 따라 채점자간 일치도, 재검사 신뢰도, 동형검사 신뢰도, 반분신뢰도, 크론박(Cronbach)의 알파계수(일반화 계수) 등으로 구분할 수 있다(백순근, 2000; Enders & Bandalos, 1999; Hogan, Benjamin & Brezinski, 2000; Messick, 1989; Spalding, 2000; Wiggins, 1989). 그러나 이러한 분류는 어디까지나 특정 평가 방법이나 도구의 신뢰도를 가능한 한 다양한 방법으로 파악하기 위한 기능상의 구분으로 생각해야 한다. 신뢰도를 제대로 파악하기 위해서는 가능한 한 다양한 방법으로 통합적이면서도 전체적인 정보를 수집하는 것이 바람직하다. 주어진 상황이 동일하다면 평가 방법이나 도구의 신뢰성을 확보할 수 있는 증거는 많으면 많을수록 좋으며, 증거의 양도 중요하지만 증거의 질이 더욱 중요하다는 것도 인식할 필요가 있다.

라. 실용도

평가 방법이나 평가 도구의 실용도(usability)란 평가 방법이나 도구의 제작 과정에서 뿐만 아니라 시행 방법이나 절차, 그리고 평가 결과를 채점하거나 분석하기 위해 소요되는 인적·물적 자원의 양과 질이 주변 여건에 비추어 실용적인가를 나타내는 정도를 의미한다. 비록 어떤 평가 방법이나 도구의 적합도, 타당도, 그리고 신뢰도가 높다고 하더라도 그러한 평가도구를 제작하고 시행하기 위해 소요되는 인적·물적 자원의 양과 질이 주변 여건에 비추어 지나치게 많거나 높다면 현실적으로 사용하기란 거의 불가능할 것이다. 예컨대 학생들의 야구 실력을 평가하기 위해서는 직접 야구 경기를 하는 것이 바람직하겠지만, 학교의 여건상 야구장이 없고 야구 경기를 진행할 각종 장비들이 없다면 그 방법은 실용도가 낮다고 할 수 있다. 마찬가지로 이유로, 과학교과의 특정 내용

2) 참고로, Feldt(1997)는 선택형 검사보다 수행평가가 신뢰도는 낮을 수 있지만, 오히려 타당도는 더 높을 수 있다는 것을 수리적으로 증명하기도 하였다.

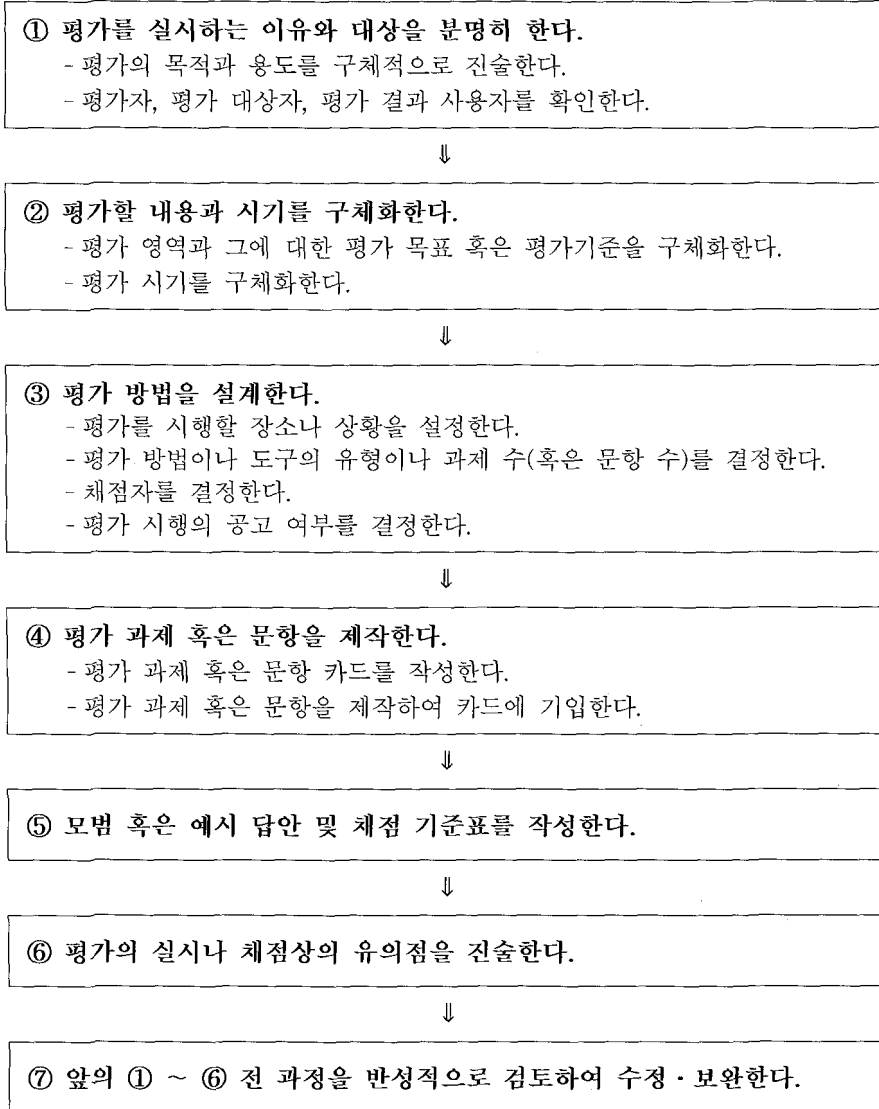
에 대해서 실험을 통하여 평가하는 것이 바람직하다고 할지라도 해당 학교에 과학실이나 실험 장비가 없다면 그 방법은 실용도가 낮다고 할 수 있다. 참고로, 과소규모 초등학교의 경우 담임 교사가 이미 학생들의 능력을 잘 알고 있음에도 불구하고 형식적으로 시험을 실시한다면 이 또한 실용도가 낮은 경우라고 할 수 있다. 이러한 경우에는 불필요하게 시험을 형식적으로 실시하는 것보다 오히려 실시하지 않는 것이 바람직할 것이다.

특히 수행평가에서는 주어진 여건과 상황 및 맥락에서 실제로 이용 가능한 평가 방법이나 도구를 활용할 것을 권장하고 있다. 흔히 평가 방법이나 도구의 실용도와 관련하여 고려해야 할 사항으로는, 그것이 평가의 목적이거나 취지에 부합하면서도 주어진 여건이나 상황에서 ① 실시 방법이 쉬울 것, ② 실시하는 데 소요되는 시간이 적절할 것, ③ 채점하기가 쉬울 것, ④ 결과를 해석하거나 활용하기가 쉬울 것, 그리고 ⑤ 검사의 실시 비용이 적절할 것 등이 있다(Gronlund & Linn, 1990).

3. 수행평가 도구의 개발 절차

수행평가 방법이나 도구의 질을 관리하기 위해 일선 학교 교사들이 앞서 언급한 ①적합도, ②타당도, ③신뢰도, 그리고 ④실용도 등을 고려하여 수행평가를 실시해야 한다는 것은 지극히 당연한 일이라는 하지만, 평가를 시행할 때마다 일일이 타당도나 신뢰도 등을 경험적으로 검증한 후에 평가를 실시하기란 현실적으로 거의 불가능한 일이라고 할 수 있다. 따라서 그러한 측면들을 고려하되 일선 교사들이 수행평가 방법이나 도구의 질을 관리하기 위한 하나의 방안으로 활용할 수 있는, 수행평가 도구 개발을 위한 일반적인 절차를 예시하면 <표 1>과 같다. 다만, 수행평가 도구를 개발하는 과정이나 절차는 창의성을 요구하는 작업이므로, 여기에 제시하고 있는 일반적인 개발 절차는 단위 학교의 여건이나 교사의 능력에 따라 수정·보완하는 것이 바람직하다. 아울러 수행평가 도구를 혼자서 개별적으로 개발하는 것보다 관련 교과나 동학년 교사들이 함께 상호 협의하여 공동으로 개발하는 것이 바람직하다.

<표 1> 수행평가 도구 개발의 일반적인 절차



가. 평가를 실시하는 이유와 대상을 분명히 한다.

대부분의 경우 평가(혹은 시험이나 검사)는 그 자체가 목적이 될 수 없다. 다만 특정한 목적을 달성하기 위한 하나의 수단일 따름이다. 따라서, 평가 도구의 개발은 평가를 실시하는 이유와 대상을 분명히 밝히는 일에서부터 출발해야 한다. 평가를 실시하는 이유와 대상을 분명히 밝히는 작업을 구체적으로 제시하면 다음과 같다.

1) 평가의 목적과 용도를 구체적으로 진술한다.

우선 평가의 목적과 용도를 구체적으로 밝히는 작업이 필요하다. 참고로 평가의 목적과 용도를 분류하는 방식 중의 하나는 교수·학습 활동의 진행 순서와 관련하여 분류하는 것이다. 즉, 교수·학습 활동을 시행하기 위해 반편성을 하거나 능력별로 집단을 구성하기 위하여 시행하는 배치고사(placement test), 학생들을 가르치기 직전 단계에서 교수·학습 목표에 대한 준비도 등을 확인하기 위해 시행하는 진단 평가(diagnostic evaluation), 학생들을 가르치는 중간 단계에서 교수·학습 활동을 개선하는 데 필요한 각종 정보(예: 학생들의 흥미도나 성취도 등)를 수집하기 위해 시행하는 형성평가(formative evaluation), 그리고 학습 단원이 끝났을 때, 혹은 학기나 학년이 끝날 무렵에 최종적으로 의도했던 교수·학습 목표의 달성도를 파악하거나 전체 교수·학습 활동에 대한 종합적인 피드백을 제공하기 위해 시행하는 총괄평가(summative evaluation)로 분류하는 것이다.

2) 평가자, 평가 대상자, 평가 결과 사용자를 확인한다.

평가의 목적과 용도를 밝히는 일과 함께 평가를 시행할 사람과 평가를 받을 사람, 그리고 평가 결과를 누가 활용할 것인가를 구체적으로 진술하는 것이 필요하다.

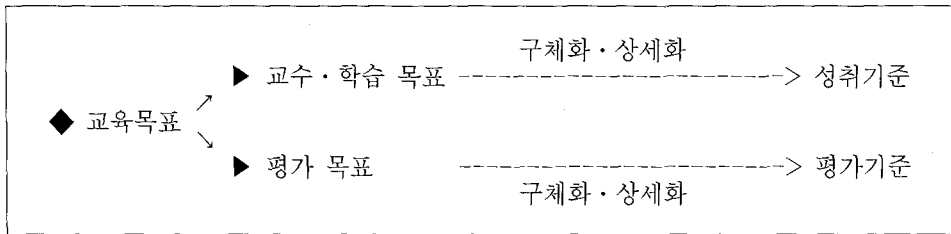
나. 평가할 내용과 시기를 구체화한다.

평가를 실시할 목적과 대상을 분명히 한 다음에 해야 할 일은, 평가할 내용과 시기를 구체화하는 작업이다. 여기서 평가할 내용을 구체화한다는 의미는 평가할 내용 영역을 정하고 그에 따른 평가 목표 및 평가기준(assessment standards)을 구체화하는 것이며, 평가 시기를 구체화한다는 의미는 언제, 얼마나 오랫동안 시행할 것인가를 구체화한다는 의미이다.

1) 평가 영역과 그에 대한 평가 목표 혹은 평가기준을 구체화한다.

평가를 시행하기 위해서는 어떠한 영역을 평가할 것인가를 구체화해야 한다. 예컨대 한 교과 전체가 될 수도 있을 것이며, 한 단원(혹은 대영역)이 될 수도 있고, 혹은 중영역이나 소영역이 될 수도 있다. 특정 영역이 평가 영역으로 결정되었을 경우에는 그 영역에 대한 평가 목표 혹은 평가기준을 구체화하는 것이 필요하다. 평가 목표 및 평가기준에 대해서 부연하여 설명하면 다음과 같다(백순근 외, 1999).

학교에서 교사들은 교육목표에 근거하여 학생들을 가르치기도 하고 평가하기도 한다. 이러한 교육목표를 ‘가르치는 상황’과 ‘평가하는 상황’으로 구분한다면, 교육목표를 ‘교수·학습 목표’와 ‘평가 목표’로 나눌 수 있다. 일선 학교에서는 대체로 가르치는 시간이 많고 평가하는 시간이 짧기 때문에, 평가 목표는 교수·학습 목표의 부분이 되기는 하지만 교수·학습 목표 그 자체가 되기는 어렵다. 어떤 평가 목표는 여러 개의 교수·학습 목표 중에 취사선택한 것일 수도 있고, 어떤 평가 목표는 여러 개의 교수·학습 목표를 통합적으로 진술한 것일 수도 있다. 따라서 교수·학습 목표를 보다 구체화·상세화한 것을 성취기준으로, 평가 목표를 좀더 구체화·상세화한 것을 평가기준으로 구분할 수 있다([그림 1] 참조).



[그림 1] 교육목표와 성취기준 및 평가기준과의 관계

2) 평가 시기를 구체화한다.

평가할 시기를 구체화한다는 것은 언제, 얼마나 오랫동안 시행할 것인가를 구체화한다는 것으로, 학년초나 학기초에 학년별 혹은 학기별 ‘수업 및 평가 계획서’⁴⁾를

4) 수행평가를 시행함에 있어서 일회적이면서도 즉흥적인 평가가 아니라 종합적이면서도 체계적인 평가가 될 수 있도록, 가능한 한 교과담당 교사들이나 혹은 동학년 교사들의 상호협의를 거쳐 학기별 혹은 연간 ‘수업 및 평가 계획서’를 사전에 작성하는 것이 필요하다. 이에 대한 자세한 사항은 백순근(2000)의 ‘수행평가의 원리’에 제시되어 있다.

작성하여 두었다면 그것에 근거하여 평가 시기를 구체화함으로써 수업과 평가가 잘 연계될 수 있도록 하는 것이 필요하다.

다. 평가 방법을 설계한다.

평가 방법을 설계한다는 의미는 평가를 시행할 장소나 상황을 구체적으로 설정하고, 구체적인 평가 방법과 평가 도구의 유형이나 과제 수(혹은 문항 수) 등을 결정하며, 누가 채점할 것인가를 결정하고, 평가의 시행에 대해 사전에 공고할 것인가의 여부를 결정하는 것을 의미한다.

1) 평가를 시행할 장소나 상황을 설정한다.

평가를 실시하는 이유나 용도, 평가 영역의 특성이나 평가기준의 성격, 실시 대상이나 실시 시기, 그리고 주변 여건이나 설정 등을 고려하여 평가를 시행할 장소나 상황을 구체화하는 작업이 필요하다. 예컨대, 일반 교실에서 시행할 것인지, 실험실이나 교과교육실에서 시행할 것인지, 운동장에서 시행할 것인지, 실내체육관에서 시행할 것인지 등을 결정하는 것이 필요하다. 그리고, 야외에서 실시하는 미술 경시대회 등과 같이 학교 밖의 특정 장소를 선정할 수도 있을 것이다.

2) 평가 방법이나 도구의 유형이나 과제 수(혹은 문항 수)를 결정한다.

이 단계는 평가를 실시하는 이유나 용도, 평가 영역의 특성이나 평가기준의 성격, 실시 대상이나 실시 시기, 평가 장소나 상황 등을 고려하여 구체적으로 어떠한 방법이나 도구를 사용할 것인지를 결정하는 단계이다. 예컨대, 학생들의 쓰기 능력을 평가하기 위해서, 쓰기 실기 시험을 실시하되 응답제한순형 2문항으로 평가하기로 결정하는 것과 같은 단계이다. 혹은 학생들의 연구보고서 작성 능력을 평가하기 위하여 연구 보고서법을 활용하되, 여러 개의 연구 과제 중에서 한 가지 과제를 임의로 선정하여 200자 원고지 100장 이내로 보고서를 작성하여 제출하도록 하는 것을 결정하는 것이다.

3) 채점자를 결정한다.

선택형 시험의 경우에는 OMR 답안지와 컴퓨터를 활용하여 채점할 수 있으나, 수행평가는 채점자의 전문적인 관찰이나 판단에 의존해야 하는 경우가 많기 때문에 누가 채점할 것인가를 결정하는 것이 매우 중요하다. 예컨대, 담임 교사가 채점할 것인지, 교과담당 교사가 채점할 것인지, 외부의 교과 전문가들이 채점할 것인지를 결정하는 것이 필요하다. 그리고 수행평가의 경우 가능한 한 2명 이상의 전문가가 채점하는 것이 바람

직하기는 하지만, 채점 기준표를 제대로 작성하였을 경우에는 한 사람이 채점할 수도 있다. 아울러, 채점에 앞서 채점자 훈련이 필요하다고 판단되면 적절한 시점에서 채점자 훈련을 실시해야 한다. 그리고, 평가해야 할 학생들이 많고 평가 과제나 문항이 많아서 여러 명의 채점자가 필요할 경우에는, 채점자들이 학생들을 나누어서 채점하도록 하는 것보다 특정 채점자가 특정 평가 과제나 문항의 채점을 전담하도록 하는 것이 바람직하다. 왜냐하면 특정 문항에 대해 동일한 채점자가 채점하도록 함으로써 채점의 공정성을 높이고, 동일한 과제나 문항에 대한 채점을 계속하면 익숙해지기 때문에 채점 속도가 점점 빨라지게 할 뿐만 아니라, 채점의 정확성도 높일 수 있기 때문이다.

4) 평가 시행의 공고 여부를 결정한다.

평가의 시행에 대해서 사전에 자세히 공고를 할 것인가 아니면 특별한 공고 없이 자연스럽게 시행할 것인가를 결정해야 한다. 지도·조언·충고를 목적으로 평가를 시행할 경우에는 특별히 공고를 할 필요가 없을 것이다. 그러나 평가의 결과를 중요한 의사결정에 활용하고자 할 경우(예: 합격/불합격 판정)에는 평가의 시행에 대해서 자세히 공고할 필요가 있다. 평가 시행에 대해서 공고하고자 할 때는 평가를 실시하는 이유나 용도, 평가 영역이나 실시 시기, 실시 장소 및 상황, 평가 방법이나 평가 도구의 유형 등을 자세히 언급하는 것이 바람직하다.

라. 평가 과제 혹은 문항을 제작한다.

이 단계는 실제로 평가 과제 혹은 문항을 직접 제작하는 과정이다. 이때 앞에서 언급한 내용들에 대해 요약·정리할 수 있는 평가 과제 혹은 문항 카드를 작성한 다음, 평가 과제나 문항을 제작하는 것이 바람직하다.

1) 평가 과제 혹은 문항 카드를 작성한다.

평가 과제 혹은 문항을 제작하기에 앞서 평가 방법이나 도구에 대한 전체적인 내용을 일목요연하게 파악할 수 있도록 평가 과제 혹은 문항 카드를 작성하는 것이 필요하며, 이를 예시로 제시하면 <표 2>과 같다.

<표 2> 평가 과제 혹은 문항 카드 (예시)

평가의 목적과 용도					
평가자		평가 대상자		결과 사용자	
평가 시기		시행 장소		실시공고 여부	
평가 내용					
평가 목표/기준					
평가 방법 (도 구)		예상 소요 시 간	(분)	배 점	
채점자		예 상 정답율(%)			
<p>< 평가 과제 혹은 문항 ></p> <p>< 모범 혹은 예시 답안 ></p> <p>< 채점 기준표 ></p> <p>< 평가 실시 및 채점 상의 유의점 ></p>					

2) 평가 과제 혹은 문항을 제작하여 카드에 기입한다.

이 단계는 평가 과제 혹은 문항을 카드에 직접 기입하는 단계이다. 참고로, 평가 과제가 토론법이나 실기시험 등과 같이 학생들이 직접 수행하는 것을 요구할 경우에는 학생들에게 요구하는 것이 어떠한 것인지를 구체적으로 진술함과 동시에 ‘평가 과정이나 절차’ 등에 대해서도 자세히 진술해야 한다. 그리고 문항의 배점과 함께 문항의 예상 정답율(%)을 제시함으로써 문항의 어렵고 쉬움의 정도를 제시하는 것이 바람직하다. 아울러, 각 평가 과제나 문항에 대해 예상되는 답변 소요 시간을 기록함으로써, 평가 과제나 문항들을 최종적으로 선제하거나 편집할 때 활용할 수 있도록 한다.

마. 모범 혹은 예시 답안 및 채점 기준표를 작성한다.

이 단계에서는 모범 혹은 예시 답안 및 채점 기준표를 작성한다. 일선 학교에서 채점 기준표를 작성할 때는 가능한 한 각 평가 요소별로 구체적으로 진술하여 분석적으로 채점할 수 있도록 하는 것이 바람직하다. 왜냐하면, 채점 결과를 이용하여 개별 학생들에게 구체적인 정보를 제공하기 위해서는 총체적인 채점 기준표보다는 분석적인 채점 기준표가 더 유용하기 때문이다(지은림, 1999; Johnson, Penny & Gordon, 2000). 특히, 출제자가 직접 채점하지 않을 경우에는 예상되는 여러 가지 형태의 답안을 고려할 수 있도록 채점 기준표를 구체적으로 작성해서 제공해야 할 것이다.

바. 평가의 실시나 채점상의 유의점을 진술한다.

이 단계에서는 평가 실시상의 유의점이나 채점상의 유의점 등에 대해 자세히 진술한다. 특히 실시상의 유의점에는 준비물에 대한 진술이나 계산기, 자, 영한사전, 옥편 등의 사용을 허용할 것인지 여부 등을 진술해야 한다. 또한, 채점상의 유의점에는 부분점수(혹은 채점 기준표에 제시되는 낮은 중간점수)를 허용할 것인지의 여부와 유사한 답안이 나왔을 경우에 처리하는 방법 등에 대해서 자세히 진술해야 한다. 그리고, 평가 실시 후에 가능한 한 일부학생들을 대상으로 가채점을 실시함으로써 채점 기준표에 대한 수정·보완의 필요성 여부를 판단하고, 필요하다면 본채점에 앞서 적절한 조치를 취하는 것이 바람직하다.

사. 앞에서 언급한 ‘가’~‘바’까지의 전 과정에 대해 반성적으로 검토하여 수정·보완한다.

수행평가 도구 개발의 마지막 단계인 이 단계에서는 평가를 실시하는 이유를 명확히 밝히는 단계에서부터 평가의 실시나 채점상의 유의점을 진술하는 단계에 이르는 전 과정에 대해서 반성적으로 검토하여 수정·보완하는 작업을 수행한다. 특히 모범 혹은 예시 답안이나 채점 기준표를 작성하는 과정에서 평가 과제나 문항의 진술문이나 응답 조건을 수정·보완해야 하는 경우가 많이 발생하기 때문에 수시로 반성적으로 검토하고 수정·보완하는 작업이 요청된다고 하겠다.

그리고, 평가가 시행된 후에, 그 결과 분석을 근거로 평가 과제나 문항 개발의 전(全) 과정에 대하여 반성적으로 검토·평가하여, 추후에 보다 바람직한 개발 과정과 절차를 마련함으로써 좀더 우수한 평가 과제나 문항을 개발할 수 있도록 연구·노력하는 것이 필요하다.

4. 맺는 말

21세기 지식·정보화 시대를 맞이하여 평가 방법을 다양화·전문화·특성화하기 위하여 학교 교육 현장에서 수행평가를 확대 실시해야 한다는 교육부의 수행평가 정책은 지극히 당연한 것이었다고 할 수 있다. 다만, 시행착오를 범하지 않고 교육 현장에서 수행평가를 제대로 확대 실시하기 위해서는 교육부를 비롯한 상급기관과 일선 학교 교사들 간의 긴밀한 의사소통이 가능하도록 함으로써 상호 불필요한 오해를 줄이고, 수행평가의 효과나 효율성이 극대화되도록 교육에 종사하는 모든 사람들이 서로 힘을 합쳐야 할 것이다. 그리고, 수행평가에 대한 연수 및 홍보 활동을 강화하고, 수업 및 평가를 위한 다양한 자료들을 개발·보급해야 할 것이며, 교육 현장의 주어진 여건과 상황에 따라 수행평가를 점진적으로 확대 실시해야 할 것이다.

학생들을 평가함에 있어서 수행평가를 위한 방법들은 선택형 검사에 비해 실시하기가 어려울 뿐만 아니라 인적·물적 자원이 많이 소요되며 많은 교수·학습 목표를 포괄하여 평가하기 어렵다는 단점이 있다. 아울러 수행평가의 신뢰성, 비교가능성, 공정성 등을 확보하기가 다소 어렵다. 그러나 수행평가는 교수·학습의 과정과 평가의 과정을 분리하지 않고 통합된 형태로 운영하기 쉽기 때문에, 교수·학습의 방법과 절차를 개선하기에 적절한 방법이다. 아울러 구체적인 상황하에서 학습자 개개인의 직접적인 수행을 통해서 지속적이면서도 종합적으로 평가하기 때문에 창의성이나 문제 해결력 등 고등 사고기능에 대한 평가나 인지구조의 변화 과정 혹은 학습의 과정에 대한 평가가 용이할

뿐만 아니라, 학생의 개성을 강조하면서도 전인적인 변화·발달에 대해 제대로 평가할 수 있다는 장점이 있다(백순근, 2000; Baron & Boschee, 1995; Braden, 1999; Burger & Burger, 1994; Spalding, 2000).

교육 현장에서 선택형 검사와 같은 양적 평가(혹은 선발형 평가)를 할 것인가 아니면 수행평가와 같은 질적 평가(혹은 충고형 평가)를 할 것인가는 주어진 과제의 특성과 상황에 따라 어느 정도는 선택해야 할 문제라고 할 수 있다. 그러나 21세기 지식·정보화 시대로 특징 지워지는 현대 사회에서 교수·학습의 질을 개선하고, 학생들의 다양한 개성을 존중하며, 창의성이나 문제 해결력 등 고등 사고기능을 신장시키고, 학생 개개인의 전인적인 성장·발달을 돕기 위해서는 수행평가와 같은 질적 평가(혹은 충고형 평가)를 교육 현장에 적극적으로 도입하기 위해 연구·노력해야 할 것이다.

그리고 교수·학습 방법의 변화 없이 평가 방법의 변화만으로는 수행평가를 제대로 정착시킬 수 없다는 것을 깊이 인식하고(Eisner, 1999; Haertel, 1999), 교수·학습 방법의 개선·발전을 위해서도 지속적으로 연구·노력해야 할 것이다. 특히, 일선 학교에서 교수·학습 방법의 개선 및 학생평가 방법의 개선이 성공적으로 이루어지기 위해서는 교사들의 적극적인 참여가 필수적이므로, 교사들의 사기 진작을 위해 최선을 다해야 할 것이다. 아울러 교사들도 21세기 지식·정보화 시대를 맞이하여 학교 교육 현장에서 수행평가의 정착이 반드시 필요하다는 것을 깊이 인식하고 질 높은 수행평가를 시행하기 위해 성실히 연구·노력해야 할 것이다. 또한 수행평가의 정착을 위해 현직 교사에 대한 연수가 강화되어야 할뿐만 아니라, 교사 양성기관에서도 예비 교사들에게 가르치는 방법 및 기술과 함께 평가하는 방법 및 기술에 대한 교육도 강화해야 할 것이다.

마지막으로, 수행평가 방법이나 도구를 선정하거나 직접 개발할 때, 그 질을 관리하기 위한 방안에 대해 논의한 이 글이 학교 교육 현장에서 수행평가를 제대로 정착시키는 데 조그마한 보탬이 되기를 기대한다. 그리고, 교육 현장에서 수행평가를 제대로 정착시킴으로써 학생 개개인의 성장·발달을 돕고 단위 학교의 교수·학습 활동을 개선하며, 나아가 국가 교육 발전에도 기여할 수 있기를 기대한다.

〈 참고 문헌 〉

- 교육부(1998). 새로운 대학입학제도와 교육비전 2002: 새 학교문화 창조. 교육부.
- 교육부(1999). 초·중·고등학교 학교생활기록부 전산처리 및 관리 지침. 교육부 훈령 제587호. 교육부.
- 백순근(2000). 수행평가의 원리. 서울: 교육과학사.
- 백순근 외(1999). **고등학교 각 교과별 수행평가의 이론과 실제 (총 10권)**. 서울: 교육진흥연구회.
- 백순근 편(1998). **중학교 각 교과별 수행평가의 이론과 실제**. 서울: 원미사.
- 성태제(1999). 교육평가 방법의 변화와 결과타당도에 대한 고려. **교육학연구**, 37(1), 197-218.
- 시·도교육청평가위원회(1999a). **1999년도 시·도 교육청평가 편람**. 시·도교육청평가위원회.
- 시·도교육청평가위원회(1999b). **1999년도 시·도 교육청평가 보고서**. 시·도교육청평가위원회.
- 지은림(1999). 사회과 보고서 수행평가를 위한 총체적 채점과 분석적 채점의 비교. **교육평가연구**, 12(2), 11-24.
- 한국교육평가학회 외(2000). **수업과 수행평가의 개선을 위한 질적 연구방법의 활용**. 한국교육평가학회, 한국교육과정학회, 한국교육인류학회, 한국초등교육학회.
- Baron, M. A. & Boschee, F. B.(1995). **Authentic Assessment: The Key to Unlocking Student Success**. Pennsylvania: Technomic.
- Braden, J. P.(1999). Performance assessment and diversity. **School Psychology Quarterly**, 14(3), 304-319.
- Burger, S. E. & Burger, D. L.(1994). Determining the validity of performance-based assessment. **Educational Measurement: Issues and Practice**. 13(1), 9-15.
- Downing, S. M. & Haladyna, T. M.(1997). Test item development: Validity evidence from quality assurance procedures. **Applied Measurement in Education**, 10(1), 61-82.
- Eisner, E. W.(1999). The use and limits of performance assessment. **Phi Delta Kappan**, 80(9), 658-62.

- Enders, C. K. & Bandalos, D. L.(1999). The effects of heterogeneous item distributions on reliability. **Applied Measurement in Education**, 12(2), 133-150.
- Feldt, L. S.(1997). Can validity rise when reliability declines? **Applied Measurement in Education**, 10(4), 377-387.
- Feldt, L. S. & Qualls, A. L.(1999). Variability in reliability coefficients and the standard error of measurement from school district to district. **Applied Measurement in Education**, 12(4), 367-381.
- Gronlund, N. E. & Linn, R. L.(1990). **Measurement and Evaluation in Teaching (Sixth Edition)**. NY: Macmillan Publishing Company.
- Haertel, E. H.(1999). Performance assessment and education reform. **Phi Delta Kappan**, 80(9), 662-666.
- Hogan, T. P., Benjamin, A. & Brezinski, K. L.(2000). Reliability methods: A note on the frequency of use of various types. **Educational and Psychology Measurement**, 60(4), 523-531.
- Johnson, R. L., Penny, J. & Gordon, B.(2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. **Applied Measurement in Education**, 13(2), 121-138.
- Linacre, J. M.(2000). New approaches to determining reliability and validity. **Research Quarterly for Exercise and Sport**, 71(2), 129-136.
- Messick, S.(1989). Validity. In R. L. Linn(ed.), **Educational Measurement (3rd Edition)**, 13-103. NY: Macmillan Publishing Company.
- Sawilowsky, S. S.(2000). Psychometrics versus datametrics: Comment on Vacha-Haase's "reliability generalization" method and some EPM editorial processes. **Educational and Psychology Measurement**, 60(2), 157-173.
- Spalding, E.(2000). Performance assessment and the new standards project: A story of serendipitous success. **Phi Delta Kappan**, 81(10), 758-764.
- Wiggins, G.(1989). A true test: Toward more authentic and equitable assessment. **Phi Delta Kappan**, May, 703-713.

< Abstract >

A Proposal for Successful Implementation of Performance Assessment :
Focus on the Quality Control of its Methods and Instruments

Sun-Geun Baek*

The major purpose of this paper is to suggest a proposal for successful implementation of performance assessment in primary and secondary schools. This proposal is developed with focusing on the quality control of methods and instruments for performance assessment.

This paper, at first, proposes four criteria for selecting or for developing methods and instruments for performance assessment. The first criterion is 'properness' for the intrinsic purposes of performance assessment. The others are 'validity', 'reliability', and 'usability' of methods and instruments.

In addition, this paper proposes a systematic procedure for developing methods and instruments for performance assessment in order to improve or to guarantee their quality. The procedure includes seven steps as follows:

Step 1: Define the purpose of performance assessment.

Step 2: Define the contents to be assessed.

Step 3: Design the assessment methods.

Step 4: Develop the assessment items.

Step 5: Construct exemplary answers and scoring rubrics.

Step 6: Describe matters to be attended to implement.

Step 7: Review reflectively and revise the previous steps.

Since the successful implementation of performance assessment is a quite critical matter for developing primary and secondary school education, the quality of its methods and instruments should be continually improved and be guaranteed. ⁵⁾

* Assistant Professor, Department of Education, College of Education, Seoul National University.