

고등학교 국어과 평가의 문제점과 개선 방안

—평가 사례를 중심으로—

김혜정*

<차 례>

1. 머리말 : 고등학교 평가에 대한 이해
2. 고등학교 국어과 평가의 문제점과 개선안
3. 맺음말

1. 머리말 : 고등학교 평가에 대한 이해

주지하다시피, 학교 교육에서 평가는 구조적으로 얽혀있다. ‘목표 → 내용 → 방법 → 평가’라는 교육과정상에서 순방향적 의사결정의 마지막 단계이며, 역으로 평가 결과가 다시 ‘방법 · 내용 · 목표’ 수정에 대한 판단 근거이자, 개선 방향을 제시한다는 점에서, 평가는 물고 물려 있는 연쇄적인 구조 속에서 이해되며, 교육과정 체계의 일관성이 지켜진다.

그러나 평가는 교육과정의 내적 구조 속에서만 이해되지 않는다. 사회적 측면에서 그 어느 것보다 제도적이고 정치적인 특성을 여실히 드러낸다. 평가의 본질상 사회화로의 관문이라는 성격을 지닐 수밖에 없다.¹⁾ 특

* 경북대학교

1) 어떤 수행 기준점을 정하여 어떤 사람이 어떻게 수행하는가를 평가하는 것은 중요한 사회 제도가 되었으며, 많은 중요한 사회 역할로의 입구를 통제한다는 점에서 그것은 관문의 기능을 담당하고 있다. 그러나 평가의 중요성을 생각할 때 평가의 실체에 대해 거의 이해되고 있는 것이 없다는 것은 놀라운 일이다(McNamara, 2001 : 3).

히 고등학교에서의 평가는 교육과정의 내적 구조 속에서 이해되기보다는 중요한 사회 역할로의 입구를 통제한다는 점에서 선발과 구분을 위한 관문의 성격이 더 강하다. 그래서 고등학교 평가에서 서열은 유의미한 정보지만, 어떤 학습자가 ‘무엇을 왜’ 잘하고 못하는지에 대한 과정, 그에 대한 교사의 기술과 판단은 덜 중요한 정보가 되기 쉽다. 즉 대입이라는 상급학교 진학에 어떻게 편리하고 효과적으로 적응할 수 있을 것인가에 평가가 맞춰져 있다. 이러한 좌표 설정은 대학 진학을 위한 학습자, 학부모와 지역 사회의 요구를 충실히 반영했다는 이유만으로 정당성을 확보한다. 그래서 대입에 효과적이라면 적어도 ‘이론상’으로 평가되어야 할 본질적인 부분이 축소되거나 등한시 되고, 일종의 자격시험이 학업 성취도인양 과장되거나 간접 평가로도 수행 능력을 평가한 것으로 전용될 수 있다. 이로부터 고등학교 평가의 외형을 결정하는 첫 번째 변인은 대입 전형이 어떤 평가 자료들을 요구하느냐 하는 것이지,²⁾ 학습의 결과를 어떻게 적정하게 측정할 것인지, 평가를 통해 학습 능력을 어떻게 개선시킬 수 있는지는 차선으로 미뤄진다. 평가가 사회적, 정치적 성향을 띠고 사회제도적으로 불가피하게 얽매여 있는 측면이다. 이것은 아이러니하게도 현재와 같이 편향된 고등학교 평가 관행의 책임에 대해 쉽게 비판할 수 없는 이유이기도 하다.

여기서 과연 우리나라 교육 관계자들이 ‘고등학교’를 어떻게 이해하고 있는지 반문해 볼 필요가 있다.³⁾ 고등학교는 더 이상 중등 교육 기관으로 독립성과 완결성을 갖지 못하고, 적어도 우리나라와 같이 대학 진학률이 80%가 넘는 나라에서는 교육의 최종 종착점이자 목표인 대학 교육으로 가는 가교 역할만 한다. 고교생이면 누구나 대입 준비를 위해 교과를 공

2) 이른바 “WYTIWYG(what you test is what you get) 현상, 일명 ‘시험역류현상’으로 설명하기도 한다(최호성, 1996 : 155, 류수열 2004 : 68 재인용).

3) 우리나라의 대학 진학률은 응시자의 83%(2007년 12월 통계청 자료)로, OECD국가 중에 최상위에 속한다. 진학을 자체만 보더라도, 우리나라 고등학교의 목적과 기능에 대해 생각해볼 수 있다. 중학교에 이은 중등 교육의 종착 또는 종결 기관으로서 완결성을 가지지 못하고, ‘일반적 교양교육과 기초적인 전문교육을 담당한다’는 고등학교의 자기 충족적 역할을 다하지 못한다.

부하며, 교사는 정규 수업 내용보다는 대학이 전형 자료로 할 만한 평가(예컨대, 수능과 논술)를 준비시키며, 학교는 명문대 진학률로 평가받는다. 이러한 구조는 초·중·고등학교에도 연쇄적으로 맞물려있다.⁴⁾

본고는 고등학교 교육의 이러한 특수성을 주지하고, 현재의 사회문화적 맥락에서 국어과 평가가 어떻게 이뤄지고 있으며, 개선의 여지가 있다면 무엇인지 살펴보고자 한다. 논의를 펼치기 전에 결론은 무기력하게도 예상가능하다. 즉, 학교 안이든 밖이든 평가의 변화는 사회 체계의 요소로서 구조 전체가 함께 변화해야지, 고등학교의 평가 정상화만 강조하거나 다양한 평가 시행법에 대한 이론적 안내를 아무리 제공한다 해도 인식의 전환이 쉽게 따르지 않는다. 대학이라는 고등교육기관에 들어가기 위한 경유지이자 수단이 되어버린 고등학교에서, 대학 입시와의 연관성 부족, 채점 공정성에 따른 안팎의 불신, 과제의 수준과 수행 시간의 확보 어려움, 채점의 불편함과 시간 소요 등 칩칩산중의 문제들을 감수해가며 수행 평가를 치르기가 쉽지 않기 때문이다. 게다가 수행평가 하나로 평가가 완성되는 것이 아니라, 지필평가에 대한 보완적 평가로서, 평가에 대한 양적 부담은 배가 된다.

‘수행평가’와 ‘국어과 평가의 문제점’이라는 두 가지 주제에 대해서는 이미 많은 연구가 수행되어왔다. 원론적 수준이든 아니든 문제점에서 개선방안으로 나가는 방식은 더 이상 의미가 없을 것이다. 그럼에도 불구하고, ‘고등학교’ 상황에 보다 더 초점을 둔, 교사들의 목소리를 통해 이미 우리가 문헌으로 알고, 분위기로 짐작하고 있었던 문제들을 확인해보고, 또 각 경우마다 개선안을 붙여보았다.

4) 이렇듯 초·중·고교도 입시 위주의 교육에 내몰리곤 하지만, 초등학교와 중학교가 대학 입시를 목전에 두지 않았다는 점만으로도 중학교의 분위기는 고등학교에 비해 자유롭다 할 수 있다.

2. 고등학교 국어과 평가의 문제점⁵⁾과 개선안

(1) 선다형 지필평가

중학교에 근무하고 있는 교사들은 대부분 수행평가의 취지나 구체적인 방법에 관심과 지식을 갖고 의욕적으로 시도하는 편인데 반해, 고등학교의 경우는 선다형 지필평가 문항제작 기술에 관심이 더 많다. 중학교보다 지필평가 비율이 높으며, 중학교에 비해 선다형이 단답형이나 서술형보다 많다. 이러한 현상은 대입수능시험을 앞둔 고등학교의 특수한 사정을 반영한 것이지만, 이 때문에 오히려 단순 암기식 선다형 문항이 지양되고 일명 ‘수행형(최미숙, 2004 : 244-245)’ 객관식 문항의 도입된 점은 한편 바람직한 현상이라 할 만하다. 그러나 전체적으로 문항 제작에 대한 이론 부족, 교과내 평가협회의 구성의 어려움, 평가에 대한 의사소통 결여 등이 평가 관련 문제의 원인이라 할 수 있다.

가. 선다형 문항 제작 절차와 분석에 대한 지식 및 경험의 부재

■ 문항 수준, 유형, 완성도, 출제근거

...(중략)... 단순한 글의 성격을 묻는 문제 또는 글의 종류를 묻는 문제가 매 단위 한 문제 이상 나오거나 7차 교육과정 이전의 유형들이 대부분이었다. 학습목표와의 연관성도 전혀 고려되지 않았고 정답과 오답의 차이가 심

5) 다음 사례들은 대구광역시 중등연수원에서 실시한 연수(지역 단위 연수 프로그램)에서 나온 ‘비형식 자료’이다. 정식 ‘강의자료집’과 연수 결과 ‘분임토의보고서’가 별도로 제작되었지만, 이 사례모음은 국어과 평가 연수(‘국어과 평가문항제작 연수’) 실시 전에 연수 참가자의 관심과 요구를 조사하기 위해 선행 과제로 내도록 한 것이며, ‘학교에서의 평가 문제점과 수행평가의 어려움’에 대해 서술 형식으로 자유롭게 기술하도록 하였다. 이 연수는 자율 연수로서, 연수 주제에 대한 수요를 조사하고 선정된 것이며, 연수 정원을 선착순으로 제한할 만큼 모두 평가에 대한 관심과 학구열이 높은 상태였다. 아래 예문은 다듬지 않고 그대로 옮겨왔다.

하게 많이 나거나 정답이라고 하기에 애매한 문제들이 많았다. 그리고 문제집에서 베껴서 그대로 내거나 바꿀 경우 이상하게 바꾸어서 그대로 사용할 수 없는 문제들이 많았다.(M고)

■ 체계적인 평가 문항 제작 과정

‘대학수학능력시험’과 ‘학업성취도평가’(교육과정평가원 주관)와 같은 국가수준 평가에서 평가문항(평가도구)을 제작하는 과정을 직접 체험해 볼 수 있다면 현장 국어교사에게는 좋은 경험이 되겠다. ‘성취기준-지문선정-성취수준(문항곤란도)-발문-선택지-검증’의 과정을 거치는 것으로 알고 있다. 각 단계별로 진행되는 과정, 노하우 등을 전수 받을 수 있기를 바란다. 문항제작과 관련된 논문이나 자료집, 기타 참고자료의 목록을 제공받을 수 있으면 좋겠다. 또한 제작한 문항이 적절한지를 평가할 수 있는 문항 검토 원리에 대해서도 다루어야 할 것이다.(M고)

■ 이원목적분류표의 실효성과 활용

우선 이원목적분류표를 작성하고 그에 맞게 문항지를 작성해야 옳으나, 이원목적분류표의 지식, 이해, 적용 문항에 적당히 문항을 나누고 그와 상관 없이 문항을 제작한다. 아직까지 문항내용을 지식, 이해, 적용 등으로 나누어 출제하시는 선생님을 보지 못했다.(K고)

■ 평가 결과에 대한 분석법

학교에서 평가를 마치고 NEIS에서 채점을 완료하면 문항에 대한 정답률이 나온다. 이것을 잘 이용하면 제작한 문항(평가도구)이 잘 만들어졌는지를 평가할 수 있을 것 같은데, 어떤 방법으로 해야 할지 모르겠다. 각 문항에 대한 정답률을 이용해서 평가 문항(도구)의 적절성을 판단하는 기준이나 방법이 있다면 이번 연수에서 배울 수 있었으면 좋겠다. 또한 평균, 표준편차 등을 이용해서 개별 문항뿐 아니라 당해 시험 전체 문제의 곤란도와 변별력에 대해 검증하는 방법도 알 수 있었으면 좋겠다.(M고)

■ 매력적 오답 1

잘 못 만들면 오답 시비에 휘말리고, 뻔하게 만들면 문제가 너무 쉽고 유치해진다. 뭔가 좋은 수가 없을까? 문항을 만들 때만다 머리를 쥐어뜯는 것이 이 문제다. 평가문항 제작 방법을 익힌다 해도 여전히 이 문제는 남을 것 같은 불길한 생각이 든다. 묘수가 있다면 나눌 수 있기를 바란다.(S고)

■ 매력적 오답 2

지필평가 문항을 제작하며 가장 어려움을 느끼는 것 중의 하나는 바로 매력적인 오답을 만드는 것이다. 매력적인 오답이라고 만들었으나 동료교사와 협의 과정에서 논란의 가능성이 있다고 수정해야 하는 경우도 많았다. 일부러 문제를 비트는 것이 아니라, 정말 매력적인 오답을 만들려면 어떻게 해야 하는지를 이번 연수에서 꼭 배우고 싶다.(K고)

■ 문항 전체 난이도 조절

각 선생님이 시험범위 안에서 만든 문제를 수합하고 조정해서 최종 출제한다. 각 선생님들이 신중하게 출제하시지만 선생들 간의 문제의 난이도에 대한 의견 조정이 문제가 된다. 그래서 결국은 학생들의 1차 지필고사 성적이 기준이 되는 일이 발생한다. 즉, 1차 지필고사의 학생 평균이 높으면 2차 지필고사는 좀 어렵게, 1차 지필고사의 학생 성적이 낮으면 좀 쉽게 출제하는 방식으로 조정한다. 이 때문에 문제가 어려워 혹은 쉬워서 학생들이 불만을 갖는 경우가 발생하는 경우도 있고, 출제자의 의도와 전혀 다른 결과가 나올 때도 있다.(Y고)

■ 단답형 제작

단답형의 문제를 출제하시는 분이 의외로 많다. 단순히 고전소설의 유형을 묻는다거나, 글의 성격을 묻는 등의 문제들은 5~6차 교육과정 당시 우리들이 시험에서 접했던 문제들이다. 이는 교사들의 연령과도 상관성이 있는 듯하다.(D고)

이러한 문제점을 살펴보면, 먼저 측정과 검사에 대한 지식이 부족함⁶⁾을 지적할 수 있다. 측정 중에는 직접 측정이 불가능하여 간접측정을 해야 하는 대상이 있는데, 이때 사용되는 간접측정 도구의 하나가 검사이다.

6) 측정(measurement)은 사물의 성질을 구체화하기 위해서 수를 부여하는 절차이다. 측정의 대상을 직접측정이 가능한 것과 직접측정이 불가능하여 간접 측정만이 가능한 대상으로 구분하는데, 예컨대 가시적인 길이, 무게 등은 직접 측정이 가능하지만 인간이 지니고 있는 잠재적 특성은 직접 측정이 불가능하다. 따라서 인간 내면에 지니고 있는 능력은 ‘검사(test)’라는 도구를 사용하여 측정한다. 예컨대 지능, 성격, 흥미, 태도, 자아개념, 학업성취도 등이며, 보통 지필검사(paper and pencil test)를 사용한다(성태제, 2001 : 30).

이에는 지능검사(Binet 검사와 Wechsler 검사), 학업적성검사—미래의 학업을 얼마나 성공적으로 잘 할 수 있는지 예언하는 검사로 SAT I, TOEFL, GRE, 대학수학능력시험 등—, 학업성취도 검사—현재 가르친 내용을 얼마나 알고 있는지 성취 정도를 측정하는 검사로 학교에서 치러지는 대부분의 검사이다. SAT II, ACT, 1994년 이전의 대학입학학력고사 등—가 있다.

특히 대학수능시험과 같은 학업적성검사(Scholastic Aptitude Test)는 개인이 미래의 학습을 얼마나 잘 할 수 있는지를 예견하기 위한 검사이므로 예측타당도가 중시되고, 미래 학습을 위한 잠재적 능력을 측정하므로 축적된 학습 내용을 측정한다. 반면, 학업성취도검사(Academic Achievement Test)는 개인의 지식, 기술, 성취의 현재수준을 측정하기 위한 검사로서 가르치고 배운 내용을 얼마만큼 알고 있는지를 측정하는 검사이므로 내용타당도가 중시되고, 검사 목적이 지식이나 기능의 현재 수준을 측정하므로 예전부터 축적된 학습내용보다는 최근에 배운 학습 내용을 측정한다는 점에서 학교 학습과 밀접한 관련을 가진다. 따라서 검사난이도나, 출제범위 면에서 학업적성검사가 상대적으로 더 어렵고 더 폭이 넓다고 할 수 있다.

이 둘을 구분하는 절대적 기준은 없지만 대개 고등학교에서 실시하는 검사는 학업성취도 검사로서, 학업적성검사인 수능시험에 비해, 교수학습 목표에 명시된, 즉 가르치고 배운 내용을 측정해야 하며, 보편적이고 적절한 문항유형을 사용하고 검사 결과를 어떻게 사용할 것인지를 고려해야 하는 평가 원칙에 더 주의를 기울여야 한다.

두 번째는 평가 절차에 대한 지식이 부족하다는 것이다. 사실 이러한 형식과 절차를 반드시 따르는 것이 중요하다기보다, 절차에 따른 개념이 교사에게 내재화 되어 있다면 평가 협의 과정에서 의사소통이 원활하게 이루어질 수 있다. 문항 제작은 일반적으로 평가 영역과 내용 선정, 평가 목적 설정, 평가 방법 및 도구 결정—기타, 이원분류표, 채점 방법까지 포함—, 문항 제작의 네 단계를 거친다. 그러나 피험자들이 획득한 원점수에 대한 상대적 서열을 알려주어야 하는 규준지향(norm-referenced) 표준화검사의 경우에는 문항응답자료 분석 등의 문항 분석 및 결과 평가 단계를 더

거친다.⁷⁾ 대부분의 교사들은 문제와 답을 문항 출제의 전부라고 생각하여 사후 조정을 하지 않는다. 그래서 위와 같은 문항 제작상의 문제에 부딪힌다.

세 번째는 이와 함께 문항 자체에 대한 평가적 용어들에 대한 개념을 이해해야 한다. 출제된 문항이 좋은 문항인지 아니면 나쁜 문항인지를 평가하기 위한 문항분석 방법⁸⁾은 복잡한 통계를 수반하기 때문에, 교사가 일일이 문항분석을 할 수 없는 상황에 있다 하더라도 이에 대한 기초 지식을 갖고 있는 경우, 문항을 보다 잘 제작하고 이해할 수 있게 된다.

이외에 문항난이도, 문항변별도, 문항추측도,⁹⁾ 타당도, 신뢰도 등과 특히 선택형 문항에서 답지의 매력¹⁰⁾을 분석하여 문항 난이도를 조절하는 방법 등을 아는 것이 중요하다. 이들은 대부분이 문항에 대한 반응 분석을 통해 알 수 있다는 점에서 결과론적이라는 한계가 있지만, 다음 문항 제작에 참고할 수 있으며 무엇보다도 이를 통해 출제에 대한 감각을 키울 수 있다는 점에서 필요한 과정이라 할 수 있다.

다음 절에서 자세히 언급하겠지만, 문항 제작을 해보지 않은 국어과 교사는 없을 것이다. 그러나 단순 지식을 확인하는 것이 아닌, 언어 사고력을 평가하는 문항을 만들기 위해선 좋은 문항에 대한 일종의 감식안

7) 성태제(1996)외에, 국어과의 경우, 최지현(2000), 류수열(2004)을 참조.

8) 검사 이론으로서 고전검사이론과 문항반응이론이 있는데, 후자(items response theory)는 최근 국내, 외국어 능력 시험 표준화 정책에 영향을 받아 국어과에서도 관심이 높아졌다. 이는 검사총점으로 문항분석을 하는 것이 아니라, 문항은 하나하나의 불변하는 고유한 속성을 지니고 있다는 데 기반하여, 문항의 개별적 속성을 (문항특성곡선을 써서) 분석하는 이론이다.

최근 NEIS(교육행정정보시스템)에는 이를 프로그램화 하여 제시하고 있다.

9) 간략히, 문항난이도는 문항이 쉽고 어려운 정도이고, 문항변별도는 문항이 피험자의 능력을 변별하는 데 적효하느냐 하는 판단으로 능력이 높은 피험자가 문항의 답을 맞히고 능력이 낮은 피험자가 문항을 맞히지 못했다면 이는 변별도가 높은 것이다. 문항추측도는 능력이 전혀 없어도 문항의 답을 맞히는 확률을 말한다.

10) 이를 ‘오답지 매력도’라고 하는데, 교사들이 관심을 많이 가지는 이유는 선다형 문항에서 답지 작성이 문항의 질을 좌우할 뿐 아니라, 고등사고력 측정에까지 영향을 주기 때문이다. 오답지들이 그럴듯할 때 문항의 난이도도 높아지고 비교, 분석, 종합 등의 고등정신기능을 측정할 수 있지만, 반대로 매력도가 낮은 경우는 5지선다라 할지라도, 4지, 3지 선다형, 진위형 문항으로 변환 수 있다.

(connoisseurship)을 갖추는 필요가 있다. 평가자들이 이러한 일정 수준의 지식을 동일하게 갖추고 있을 때, 교과내 평가자간 협의가 효과적으로 이뤄지고, 결국엔 평가를 대비한 학습 지도에도 도움이 된다.

나. 출제자 간 또는 평가자와 피평가자 간 평가에 대한 인식 차이

교사 간에 교수·학습 내용의 차이로 인해 평가 결과의 차이가 생기는 것은 빈번하고, 평가관의 차이로 평가 결과가 달라지기도 한다. 또한 평가자와 피험자인 학습자 간에도 평가에 대한 태도와 인식의 차이가 크게 존재하기도 한다. 위의 ‘가’절이 평가 문항 자체와 관련되었다면, 이것은 평가의 상황 맥락과 관련된 문제들이라 할 수 있다.

■출제자 간 교수·학습 내용의 차이

대단위 인문계 고등학교의 교과 수업 운영은 한 명의 선생님이 여러 반으로 쪼개어 수업하거나 학급 단위로 선생님이 수업함으로써 교과 수업 내용의 차이가 발생할 수밖에 없다. 이러한 차이는 공정한 평가의 걸림돌이 된다. 즉, 교과와 통일된 교안이나 학습 요소에 대한 협의의 부재가 평가에서 고스란히 나타나는 문제가 발생한다.(Y고)

■문항수와 유형 배분 결정

평가 결과가 학급별로 큰 차이가 났을 경우에 출제를 한 교사가 불편한 의심을 사기도 한다. 문항을 출제할 때 좀 더 깊이 있는 논쟁과 협의가 따라야 함에도 나눠먹기식으로 문항수를 분배하여 각자 몇 문제씩 출제한다든가, 그렇게 출제된 문항에 대하여 서로가 토를 달지 않는 것—전문교과에 대한 자존심 때문인지—이 불문율처럼 되는 경우도 있어, 생기지 않을 수 있는 민원(이의제기)을 야기하기도 한다.(C고)

■교과들 간, 교사·학부형 간의 갈등

문제의 수준이나 발문, 지문 선택, 선택지 등과 관련해서 침범할 수 있는 여지가 많지 않고 조심스럽게 이의를 제기하거나 의문점을 드러내도 개인의 견해로 밀고 나가는 경우 문제의 방향을 수정하기가 힘들다.(K2고)

■교사와 학생/학부형 간의 갈등

교사의 문항 출제에 대한 이의제기가 빈번하게 일어나고 있음. 학생 개인의 사소한 이의제기도 있지만 학부모가 직접 찾아와 문항에 대한 이의를 제기하는 경우도 찾아지고 있음. 학부모들의 학력수준이 높아진 원인도 있겠으나 학교 내신 성적에 대해 대단히 민감하여 지필고사 문제뿐만 아니라 수행평가의 문제를 모두 검토하여 구체적이고 전문적으로 문제를 제기하는 경우도 있음. 자녀의 성적에만 집착하여 때때로 억지에 가까운 주장을 하는 경우도 있어 현장에서 문항 출제와 관련하여 굉장히 주의를 기울이고 있음.(C2고)

■문항 출제 기피

교사간에는 문항출제와 관련한 갈등이 가장 많다. 교사라면 누구나 수업을 하고 평가를 해야 하는 의무와 권리가 있으나 가장 기본이 되는 이 일이 가장 부담스러운 일이기도 하기 때문이다. 요즘엔 출제가 거의 컴퓨터 워드 문서로 작성되기 때문에 이에 능력이 부족한 선배교사들은 애초부터 문항출제에서 제외되기를 기대한다. 그러나 이는 평계에 불과하며 결국엔 문항출제가 가장 부담스러운 것이다.(C고)

■교사·학생 간의 갈등

정기 교사 실시 후 정답 시비가 있는 경우, 교사 간의 수업 내용의 차이로 말하거나, 혹은 수업 시간에 다 이야기하지 않았느냐는 대답으로 일관함, 수업 장면과 상관없이 보편타당하게 누구나 정답으로 인정할 수 있는 문항 개발의 어려움.(K2고)

■학교 밖과의 갈등

작년에 시험을 쳐놓고 나서 한 문제의 답이 잘못 되었다고 학원 강사 한 분이 전화를 걸어왔다. 한참 이래저래 설왕설래를 한 다음에야 내 논리에 수긍을 하고 마무리 된 적 있다.(D고)

정답과 관련하여 대학 수준의 전문적인 지식을 가진 학부형이 시비를 걸어와 복수정답으로 수정하기도 함.(K2고)

이들 문제점은 결국 평가에 대한 태도와 인식 차이에서 비롯된 것이라고 할 수 있는데, 이는 평가자간 협의를 통해 해결할 수밖에 없다. 평가

자 중에는 평가에 대한 전문적인 지식을 갖고 있는 교사가 있는가 하면 그렇지 못한 경우도 있다. 그래서 사실 이들 사례는 평가관의 인식차로 볼 수도 있지만, 평가 전문성이 부족하여 잘못된 평가 문항을 제작한 결과 생기는 건전한 문제제기로 볼 수도 있다.

이를 상세히 살펴보면, 첫째, 평가자간 인식차를 지적할 수 있다. 학교에서의 평가는 교과내 협의 하에 여러 명의 교사가 공동 출제하다보니 같은 국어과 내에서도 교사마다 수업 내용의 차이나 강조점이 달라, 평가 내용을 선정하고 그에 따라 평가 문항을 제작했다손 치더라도, 다른 교사에게는 가르치지 않은 내용이거나 강조되지 않은 것일 수 있어서 출제자에 따라 평가 결과에 상당한 차이를 보인다. 이러한 편차를 막기 위해서 시험 문항을 만들고 난 뒤 시험 실시 전에 ‘총정리’ 식으로 문제를 암시하기도 한다.

또한 평가 시행과 관련하여 평가 방법 및 평가 도구의 결정, 평가 안내 등에서 차이가 있다. 선택형 평가의 경우는 이러한 문제가 발생할 경우가 적지만, 그 외 수행평가의 경우는 평가 태도와 시행에 관한 차이로 혼란이 생겨날 수 있다.

따라서 이러한 소모적인 논쟁을 막기 위해서는 어떤 부분을 어떻게 평가할 것인지에 대한 결정 과정을 공유해야 한다. 예컨대, 최소한 이 단원에서는 무엇이 중요한지를 요약화¹¹⁾하거나 명시적으로 기술한 평가 지침이 마련되거나 비형식적 평가가 아니라면 평가에 대해 어떻게 상세히 안내할 것인지, 평가 내용에 맞는 평가 방법과 도구를 어떻게 다양하게 사용할 것인지 등을 협의하여 결정할 수 있다면 문제를 줄일 수 있다.

둘째, 평가 협의 중 문제가 제기되었더라도 이를 받아들일 만한 분위기가 조성되지 않는 것도 문제라고 할 수 있다. 언급했듯이, 평가 협의체 내의 의사소통을 원활하게 하기 위해, 견해차에 대한 설득의 근거가 될만

11) 평가의 근거로서 학습 내용을 요약화하는 것은 편리한 측면도 있지만, 교육과정이 지나치게 상세하게 됨으로써 교육과정 해석에 있어서의 융통성과 자율성을 저해하고, 획일화될 우려가 있다. 평가의 원리에 토대를 두고, 평가 내용을 선정하는 것부터 평가 도구 제작까지 평가 상황에 적절하게 구성되는 것이 더 바람직하지 않은가 판단된다.

한 평가 지침이 마련되어 있어야 할 것이다. 그러나 무엇보다 중요한 것은 사전 평가협의의를 통해 평가 계획을 협의하고 공유하는 과정이 필요하다는 것이다.

다. 평가의 내용적 측면에 대한 심의 부족 : 평가 내용, 평가 기준, 평가 방법, 채점 방식, 평가 수준의 적절성

■ 평가 내용 선정

교육과정의 성취기준이 추상적이고 포괄적으로 되어 있다보니, 교육과정을 그대로 평가와 연결하려면 막연하기만 하다. 이것은 평가 이전에 교수학습계획을 세우기 위해 교육과정을 펼쳐 볼 때도 마찬가지이다. 이런 문제에 봉착했을 때 대부분의 교사들은 교육과정 상의 성취기준보다 교과서의 학습 목표나 제재 자체에 중점을 두어 평가 문항을 만들게 되고, 결국 시중에 나온 문제집의 문제 유형을 모방하여 출제하는 길로 빠지기 쉽다.

올바른 평가를 위해서는 먼저 교수학습계획 단계에서부터 교육과정 상의 영역별 성취기준에 대한 ‘세부학습(평가)요소’의 설정이 필요하다. 듣기, 말하기, 읽기, 쓰기, 문학, 국어지식 각 영역에서 교육과정의 성취 기준을 달성하는 데 반드시 필요한 하위 항목을 분명하게 정립할 필요가 있다. 예를 들면, 시에서 ‘화자, 이미지, 운율, 은유, 상징’ 등은 반드시 학습하고, 평가해야 할 요소가 될 수 있을 것이다.

문학이나 국어지식의 경우 이러한 세부학습요소를 설정하기가 비교적 수월한 편이다. 영역의 특성상 학습하고 평가해야할 항목이 문학이론, 국어학 이론에서 중요하게 다루고 있는 내용과 비교적 일치하기 때문이다. 그런데, 듣기, 말하기, 읽기, 쓰기의 경우 세부학습요소를 설정하기가 쉽지 않다. 읽기 영역의 경우, 보통은 수능 시점 문제 유형을 흉내 내어서, 1) 전체 주제를 묻는 문제, 2) 세부 내용을 묻는 문제, 3) 형식·표현 방식을 묻는 문제, 4) 글 내용 비판 혹은 평가 하는 문제, 이런 식으로 출제하는 경우가 대부분이다. 문제 출제를 끝냈을 때 무엇을 평가하기 위한 문제인가에 대한 고민은 이미 멀리 사라지고 없다.(M고)

■ 영역별 평가 내용과 방법 : 듣기, 말하기, 쓰기 영역의 선택형 평가문항은 어떻게 만들어야 하나?

현장에서 평가는—평가뿐만 아니라 실제 수업도 그렇지만—문학과 읽기

중심으로 이루어지고 있다. 그리고 듣기, 말하기, 쓰기 수업은 하지만, 실제로 이 영역의 특성을 고려한 평가문항(특히 선택형)을 어떻게 만들어야 할지에 대해서는 내 경우에는 아는 것이 하나도 없다. 물론 이 세 영역은 수행평가가 위주로 되기 때문에 수행평가를 하지만 분명히 선택형 평가도 가능할 것 같다. 할 줄 모르니까 선택형 평가를 못하는 것이다. 방법을 알고 싶다. 각 영역별 특성을 반영한 유형이 있을 것 같다.(C고)

이를 통해 평가 문항 자체의 형식적 요건도 중요하지만, 평가의 내용적 측면, 즉 평가 목표와 내용·방법, 항목별 평가 기준 및 채점 방식 등이 선정, 기술되는 것도 평가에서 협의해야 할 중요한 문제임을 알 수 있다. 선다형 평가의 경우, 평가 방법이나 채점 방식, 평가 기준 등이 잠정 결정되어 있으므로, 상대적으로 평가 내용의 선정이 더 중요하게 되었다. 개정 교육과정에서는 성취기준과 학습 요소를 구분, 기술하고 있어서 평가 요소를 선정하는 것이 이전에 비해 용이해졌다.¹²⁾

둘째는 국어사용 영역의 경우, 내용 요소로부터 평가 내용 요소를 추출하고, 이에 적합한 문항을 만드는 것이 다른 영역에 비해 더 어렵다는 점이다. 예컨대, 10학년 쓰기 영역에서 ‘그림이나 사진, 그래프나 도표 등

- 12) 기존의 교육과정에서는 학년별 ‘교육 내용’을 제시하고, 이에 대한 ‘기준’과 ‘심화’ 수준별로 해설하는 식이라면, 2007년 개정 교육과정에서는 ‘성취기준’을 제시하고, 학습 내용 요소를 항목화하여 제시했다. 즉, 7차 교육과정이 ‘내용—기본/심화’의 2단계 진술이라면, 개정교육과정은 ‘성취기준—내용요소-텍스트’의 3단계 상세화 진술이라고 할 수 있다. 개정 교육과정의 기술 방식은 교사가 학습 내용 요소와 평가 내용 요소를 연계시키는 데 상대적으로 유리하다고 할 수 있다. 다음은 고등학교 [10-읽-5]의 예이다.

7차 교육과정	개정 교육과정
[10-읽-(5)] 읽은 내용의 신뢰성과 타당성을 평가한다.	[성취 기준] 인기 도서를 읽고 책의 가치와 인기를 얻게 된 원인을 비판적으로 평가한다.
[기본] 하나의 주제에 대하여 여러 가지 글을 비교하며 읽고, 읽은 내용의 신뢰성과 타당성에 대해 토의한다. [심화] 읽은 내용의 신뢰성과 타당성을 판단하는 기준을 알아본다.	[내용 요소의 예] <ul style="list-style-type: none"> • 최근 인기를 끌고 있는 책들의 특성과 분야 파악하기 • 많은 사람들에게 호응을 얻게 된 원인을 책의 내용에서 추론하기 • 신문, 방송 등 책의 인기와 작용한 여러 요인을 파악하기 • 인기 도서에 반영된 당대의 문화적 기호를 비판적으로 평가하기 [글의 수준과 범위] 대중적인 인기를 끌고 있는 도서

의 자료를 해석하는 글을 쓴다[10-쓰2]'에 적절한 선다형 문항을 제작하는 것은 상당히 어려울 것이다. 이는 표현·이해 영역의 경우, 전략에 대한 앎을 묻기보다는 수행적 성격을 지닌 전략 사용 능력을 평가해야 하기 때문에, 문항을 품으로써 바로 '할 줄 아는지'가 드러나도록 한다는 것이 어렵기 때문이다.

영역에 따른 평가 방법도 단순 나누기식으로 출제되고 있는데, 김진곤·김혜영(2005)의 연구 결과를 살펴보면, 영역별 경향성을 발견할 수 있다. 우선, '문학'과 '읽기' 영역이 주로 선다형으로 출제되고, '듣기, 말하기, 쓰기'는 통합형 수행평가로 실시하는 경향이 있다. 다음으로는 이와 반대로 '문학'과 '읽기' 영역은 선다형은 물론, 정례적인 독서이력철 평가와 같이, 수행평가도 다양하게 활용되는데 반해, '듣기, 말하기, 쓰기'는 직접 평가를 최소 한 차례 이상 실시하여 이를 관찰·누가 기록해야 하기 때문에 시간과 노력은 물론이고 평가 및 채점 계획을 세워야 하는 부담 때문에, 오히려 선다형 평가를 선호하는 경향이 있다. 이 둘은 서로 상반된 경향이지만, 전술한 바와 같이, '듣기, 말하기, 쓰기' 영역을 어떻게 고차적인 국어적 사고력을 묻는 수행형 문항으로 제작할 수 있는지에 대한 어려움이 클 것이다.¹³⁾

그러나 현재, 국어과 선다형 평가와 관련해서는 교사의 '경험'에만 의존하고 있을 뿐, 국어과 문항 제작에 대한 이론적 소개가 거의 없이, 과학적 접근이 이뤄지지 못하고 있다.

그러나 위의 사례의 밑줄 친 경우에서 구분되어야 할 점도 있다. 평가 내용 영역을 정하고 평가 요소를 추출해도 이를 구체적인 문항으로 제작하는 것은 별개의 것으로, 여전히 어려울 수 있다는 점, 또한 '말하기·듣기·읽기·쓰기'와 같은 국어사용영역의 경우는 기능 중심이라서 평가 요소를 설정하기 어렵다고 보는 것과 지식('본질')이 아닌 '기능'은 평가 요소가 될 수 없다고 보는 것은 구분되어야 한다는 점 등이다. 그리하여 위 사

13) 반면, '문법'은 단순 암기형 지식의 지양을 염두해서인지 평가 영역으로 비중있게 다루이지 않는다. 이는 탐구학습을 바탕으로 한 다양한 문항 개발의 필요성이 제기되는 부분이다.

례의 ‘1)~4)’와 같은 평가 내용을 평가 유형으로 오인하는 것, 또는 이를 수능형 문항이라고 생각하는 것은 잘못된 지식이다. 그러므로 평가 내용과 방법에 관한 올바른 지식이 요구된다.

■ 평가 수준 : 일관성의 시행착오

첫 발령부터 8년을 그렇게 열악하다는 ○○지구에서 근무하였습니다. 대학을 갓 졸업하고 교단에 서서 넘치는 의욕으로 다들 수준이 낮다는 아이들을 수준 높게 가르쳤습니다. 그리고 그에 걸맞은 문제를 냈지요. 교육평가 시간에 배운 지식은 남아있지 않고 발령선배들의 도움을 받아 이원목적분류표를 깡깡대며 작성하여 출제를 한 결과 평균이 50점대이더군요. 문제가 어렵다는 지적을 받고 문제를 쉽게 내고자 노력했더니 4년 쯤 지나자 평균을 70점 정도로 올릴 수 있었습니다. 학생들의 수준이 높아진 것이 아니라 순전히 문제 수준만 낮아진 것이었지요.

그렇게 8년이 지나고 저는 동부교육청 근처에 있는 학교로 옮기게 되었습니다. 거기서 처음 정기고사를 치게 되어 출제를 했는데 동료 교사들이 은근히 비웃는 것이었습니다. 문제가 너무 쉽다고……. 제가 아이들에게 수준을 맞추는 사이에 저는 슬프게도 어려운 문제, 수준 있는 문제를 낼 수 없게 되었습니다.(S고)

■ 평가 기준 : 학교 밖의 표준 지향 1

고등학교 3학년의 경우 교육과정평가원이 주관하는 모의고사를 6월, 9월에 2회 실시하고 서울, 경기도, 인천광역시 교육청에서 주관하는 전국연합학력평가(모의고사)가 4회에 걸쳐 실시됨. 모두 총 6회(1, 2학년은 4회)에 걸쳐 모의고사가 실시되나 상대적으로 수준이 높은 재수생과 비교를 하기 위해 학부모들은 사설모의(대성, 중앙, 종로 등)를 치르기를 요구하고 있음. 교육청에서는 강력하게 사설모의고사를 치르지 못하게 하고 있음. 현장에서는 모의고사 실시 자체에 대해 어정쩡한 태도를 취할 수밖에 없음.¹⁴⁾(C2고)

14) 이 사례보고에 이어진 내용은 다음과 같다. : **시행과정과 시행 후**—모의고사 감독 자체가 느슨한 것도 문제임. 실시 후에는 교과별 문항분석, 정답률 검토, 문제풀이(피드백), 담임교사의 성적 분석을 토대로 한 진로상담 등이 잘 이루어지지 않음. 이는 교사 간 갈등의 원인이 되기도 하고 사설학원과 비교되어 공교육의 신뢰도 하락에도 영향을 주기도 함.

■ 평가 기준 : 학교 밖의 기준 지향 2

학생들이 정기고사 문제지 여분을 모아서 사설학원에 가져다주면 학생들에게 문제지 당 몇 천 원씩을 주는 학원도 있으며 학원은 그 문제지들을 모아 일명 ‘기출문제집(죽집게문제)’으로 만들어 학원영업에 활용하기도 함. 이 과정에서 학생들과 학부모들이 학원강사(심지어 대학교수까지 동원)로부터 문제의 수준과 적절성을 평가받아 학교에 와서 교사에게 따지는 경우가 많아 교사로서 전문성을 갖추는 문제와 함께 교권에 대한 심각한 위협을 느끼고 있음.(C고)

교사라면 다음과 같은 의문을 가질 것이다 : 평가 수준의 항상성을 유지하는 기준이나 평가 수준의 절대 지표는 무엇인가? 시험을 치르기 전에 문항의 난이도와 타당도를 미리 알 수 있는 방법은 없는가?

요즘 같이 교육이 지역적 영향을 많이 받는 때가 없었다는 점을 생각하면 학군에 따라, 피험자의 수준에 따라 문항의 난이도와 타당도가 바뀌는 현상을 어떻게 설명해야 할지 난감하다. 매년 수학능력시험을 치르고 나면, 언론에서는 문항분석을 근거로 하여 수학능력시험이 쉬웠다고 보도하는데, 이 쉽다는 정도의 판단은 과연 교육과정 내용 수준에 근거한 것인지, 50%에 육박하는 수도권 응시자 인구에 근거한 것인지, 우수 학생의 독점에만 관심이 있는 상위권 대학들이 보기에 우수 학생의 층이 두터워 선발의 변별력이 떨어진다는 점에 근거한 것인지 알 수 없다. 아무리 서열화가 목적인 시험에서는 어쩔 수 없는 상황이라 하더라도, 평가 수준에 대한 일관된 지표, 예컨대, 학년별 성취수준을 따르지 않고 없이 지나치게 학습자나 교육구 수준에 좌우되는 것은 문제이다.

또한 교사는 모의수학능력시험과 학업성취도 평가는 그 외형이 일명 ‘수능형’으로 유사하다 할지라도 평가의 목적은 구분된다는 점을 인식하고 이를 적극적으로 설득할 필요가 있다. 이러한 목적이 혼란된 지나친 시험 시행은 고등학교 내에서 출제하는 평가 문항의 수준과 난이도에 영향을 미치고, 교실 수업에도 영향을 미치기 때문이다.

요컨대, 평가의 준거(criterion)가 일관성이 없게 되면, 평가 문항의 수준이나 난이도도 흔들리게 된다. 위의 사례와 같이, 학군에 따라 평가 문항

의 수준이 바뀌기도 하고, 학교 밖의 학부모 요구에 따라 문항의 적절성이 비판의 대상이 되기도 하고, 학교 밖의 사설학원(재수생을 포함하여)의 교육을 비교 기준으로 삼기도 하기 때문이다.

라. 선택형과 기타 지필평가 방법과의 배율 결정

다음은 국어과이기 때문에, 서답형 문항의 배율에 대해 상대적으로 더 논쟁거리가 된 경우이다. 이는 전술한 평가자간 평가관의 인식차에서 오는 문제이기도 하다.

■ 객관식 대 주관식

지금까지 지필평가 문항을 객관식으로만 출제한 적은 없었습니다. 국어과 특성상 오히려 주관식을 객관식보다 더 중요하게 생각하는 편이었습니다. 함께 수업을 했던 동료 선생님들도 생각이 같았기 때문에 별다른 고민을 하지 않았습니

다. 그런데 학교를 옮기고 문제가 생겼습니다. 새 학교에서는 주관식 출제에 대해서 선생님들께서 부정적인 입장이었습니다. 국어니까 당연히 주관식 시험을 통해서 쓰는 연습이라도 해야 하지 않느냐는 저의 생각과는 달리 선생님들께서는 단답형 주관식이 무슨 의미가 있느냐? 그것은 진정한 주관식이 아니지 않느냐? 라는 말씀을 하셨습니다. 그렇다면 서술형으로 내면 되지 않느냐?는 의견에는 채점이 너무 어렵고 객관적이지 못하다는 의견을 주시더군요. 그리고 대학 수능 시험에 주관식이 나오는 경우는 없지 않느냐? 그리고 수행평가를 통해서 충분히 서술적인 부분을 평가하고 있으니 주관식 문제는 의미가 없다는 말씀이었습니다.

국어뿐만이 아니라 다른 교과목들도 객관식으로 출제하고 있었기 때문에 정기고사 후에 성적 결과가 곧바로 나오곤 했습니다. 그러니 어쩌다가 주관식을 내는 국어과는 결과가 늦게 나올 수밖에 없었고, 시험 결과를 궁금하게 생각하는 담임 교사들에게 독촉을 받는 상황이 되었습니다.

그때부터 평가에 대해서 지금까지와는 다른 고민을 하게 되었습니다. 내가 정말 객관적인 평가를 하지 않고 있는 것인가? 초임교사들도 당연히 생각해 가지고 있는 객관식만으로 출제하는 것에 대해서 나만 뭘 모르고 고집을 부리고 있는 것인가? 시험 문제를 교사들이 나누어서 출제를 하는데, 주관식

을 내는 사람은 저밖에 없는 경우가 생기게 되었고, 다소 힘든 1년이 그렇게 지나갔습니다.

그런데 해가 바뀌고 교사 이동이 있는 후 실무자가 달라지자 주관식 출제에 대한 문제가 논의 대상이 되었습니다. 주관식을 왜 내지 않느냐에 대한 문제였습니다. 그분들의 생각은 기존의 내 생각과 비슷했기에 많이 놀랐습니다. 결국 평가라는 것이 어떤 한 평가자의 생각에 의해 좌지우지되었다는 것이라는 것을 알게 되었으니까요.(M2고)

국어과에서 수행평가가 실시된 이후, 지필평가에서 서답형¹⁵⁾은 출제 비율이 낮아졌다고 한다. 이는 수행평가가 서답형 문항의 역할을 일부 담당한다고 보기 때문이다. 수행평가의 과제 문항이 무엇인가에 따라 사실상 그렇기도 하고 아니기도 하다. 그러나 원칙적으로는 지필평가에 서답형이 포함되어야 하는 것이 균형있는 평가도구라는 점에서 더 바람직하다 할 것이다.

학교 단위별 세부 평가 방식과 유형은 시도 교육청별 ‘중/고등학교 학업성적관리시행지침’ 규정에 근거해서, 학교 단위에서 최종 결정된다. 그러나 ‘2007학년도 중등 장학 계획’에 의하면, 교과학습평가를 크게 ‘지필평가 및 수행평가’로 구분하고 ‘서술·논술형 평가’의 반영 비율을 50% 이상으로 권유하고 있어, 서술·논술형 평가를 지필평가에 포함시키라는 의미인지, 수행평가 중에 서술형과 논술형을 포함시키라는 뜻인지, 또는 별개로 인식하라는 의미인지, 이것만 봐서는 다양한 해석이 가능하다.

그러나 무엇보다도 수행평가가 전통적인 평가의 대안으로 등장하게 된 궁극적인 쟁점은 평가도구의 형식보다는 평가의 내용이 무엇인가, 곧 무엇을 평가하는가에 있다(김경희, 2007 : 110)는 점을 상기해볼 필요가 있다. 따라서 서술 논술형의 평가 형태가 학습자에게 무엇을 요구하는가에 따라 전통적인 평가가 될 수도 있고 대안적인 평가가 될 수도 있다는 관점에서 평가 방법을 결정하는 것이 바람직할 것이다.

15) 논술형, 단답형, 괄호형과 완성형의 3가지 유형을 포괄한다.

(2) 대안적 평가,¹⁶⁾ 수행평가

수행평가란 학습과제, 수행과정이나 결과를 직접 관찰하고 관찰 결과를 전문적으로 판단하는 평가 방식으로, 유사한 용어로는 대안적 평가(alternative assessment), 실제적 평가(authentic assessment), 역동적 평가(dynamic assessment), 직접 평가 등이 있고, 명칭이 보여주는 대로 각 강조점이 부각되는 정도로 이해할 수 있다.

가. 수행평가에 대한 인식 부족

지필평가 위주의 단일 평가 방식이 전부라고 생각하던 국어교육계에 수행평가의 도입으로 인해 평가에 대한 패러다임의 전환이 이뤄졌다. 그러나 아직까지도 수행평가에 대한 긍정적인 면에 대한 이해보다도 시행상의 어려움을 내세워 부정적으로 보는 시각이 팽배하다.¹⁷⁾ 그러나 평가의 결과 타당도(consequential validity)¹⁸⁾ 측면에서 대안 평가로서 수행평가의 시행 목적과, 이러한 평가가 시도됨으로써 발생한 긍정적·부정적 측면에 대해 모두 주목할 필요가 있다.

즉 변화된 평가가 측정하는 것이 무엇인가에 대한 반성이 이뤄질 수 있다. 연구 결과에 따르면, 수행평가의 도입 결과 학생들은 학습에 대한 개념은 물론, 학습하는 방식의 변화를 겪게 되었으며, 단순 반복 암기식

16) 이 용어는 어느 특정한 평가체제나 방식을 의미하는 것이 아니라 한 시대의 주류 평가가 아닌, 다른 평가 체제를 의미한다는 점에서 대용적(代用的)으로 사용되었다. 이에 따르면 객관식 선택형 평가방식의 대안 평가로서 수행평가가 있을 수 있다.

17) 예체능 교과에서는 실기평가와 비슷한 개념으로 비교적 쉽게 정착된 반면, 인문·수학·과학 계열의 교과에서 수행평가는 어떤 의미로 받아들여지는 것일까? 1999년 교육개혁의 과제로 시행되어 10년이 다 되어가는 이 시점까지 평가자에게 수행평가는 여전히 낯설게 다가온다. 게다가 현 정부의 교육시책으로, 전(全)집단에 대한 국가수준의 성취도 검사를 실시하겠다는 보도가 나간 이후, 수행평가에 대한 관심은 줄고 선다형 성취도 평가는 극대화되었다.

18) 1990년 초부터 평가의 도입으로 발생하는 결과까지도 평가 타당도로 포함하여 연구하기 시작했는데, 이러한 연구를 평가의 결과타당도라고 한다(McNamara, 2001 : 65).

지식에 대한 회상적 반응보다는 복잡한 프로젝트나 과제물에 대한 지속적인 평가 중심으로 바뀌면서, 복잡한 과제에 대한 분석, 종합, 평가 등의 고등정신 기능을 요하는 구성적 반응을 하게 됨으로써 학습의 질적 변화를 유도하게 되었다.

그러나 반면, 수행평가를 하면서, 만약 학생들이 다른 사람들의 도움을 받을 수 있는 경우,¹⁹⁾ 평가가 무엇을 측정하고 있는지에 대한 타당도의 문제를 일으킨다. 그래서 처음에는 평가의 개혁으로 보이던 것이 학생들 간의 능력을 올바르게 변별하는 능력만 감소시키는, 본래 의도하지 않은 부작용을 낳게 된다. 결과타당도를 떨어뜨리는 이 현상은 평가의 공정성 및 신뢰성에도 흠집을 내게 된다.

이와 관련하여, 수행평가 적용 후 교사의 수업방식에는 여전히 변화가 없다(김진곤·김혜영, 2005 : 141-142)는 것도 수행평가에 대한 편견을 굳히게 하는 요인이 된다. 교육과정평가원의 조사(박정 외, 2006)에 의하면 교과 수업에서 설명식 수업이 차지하는 비율은 국어가 50.7%로 교과 중에 가장 높았다. 이것은 3등을 대상으로 한 경우지만, 고등학교도 비슷할 것으로 추측된다. 현재의 강의와 설명 위주의 수업 방식을 고수하면서 수행평가가 추구하는 능력이나 기술을 계발하고 향상시킬 수 있는지에 대한 검토도 필요하지만, 현재 그런 방법으로 학습하고 있는 학생들에게 고차적인 사고력을 평가하는 것은 학습 내용에 대한 평가 방법의 적합성에 의문(김경희, 2007 : 112)을 제기한다.

■ 수행평가의 내적 타당도 : 과연 학생의 수행능력을 평가해주는 타당한 평가가 맞나?

수행평가의 경우 국어과는 상당히 항목이 많습니다. 그 중에는 방학 중에 집에서 꾸준히 해 봐야 하는 항목(독서 신문)이 있었는데, 조건은 ‘학생 혼자 힘으로 해 오기’입니다. 그런데 부모님이 해 주는 경우가 더러 있었어요. 학생은 주는 점수에 대해 수긍을 했는데, 부모님이 본인의 실력이 다른 중학

19) 예를 들어 부유한 몇몇 학생들만이 개인 교습을 받을 수 있다거나 또는 학력이 높은 부모를 둔 학생들만이 도움을 받는 현상은 외국에도 이미 논문으로 보고된 보편적 사례이다. 실제로 우리나라의 경우도 수행평가 전문학원이 우후죽순으로 생겨나 수행평가 과제를 전문가 수준으로 해결해 준다.

생들과 비교를 해서 못하다는 것을 못 받아들이겠다고 한 적이 있었습니다. 그분은 다른 교과목에 대해서도 같은 말씀을 하셨기 때문에 더욱 당황스러웠습니다. 그래서 올해는 집에서 해 오는 수행은 거의 실시하지 않으려고 모든 것을 수업 중에 진행하다 보니, 시간에 쫓기는 경우가 많았습니다. 내년에는 좀더 깔끔하고 객관적인 평가를 할 수 있기를 바라고 있습니다.(M3고)

교사뿐만 아니라, 학생들도 수행평가에 대해 무조건 오해하기도 한다. 예컨대, 공부를 잘하는 사람에게 수행평가가 불리하다는 말은 공부를 잘 한다는 것과 수행평가의 결과와는 별개라는 생각, 학업성적은 지필평가 결과에 근거해야 한다는 전통적인 학업관에 바탕한 말이다. 게다가 왜 어째서 수행평가를 쳐야 하는지 매일매일 생각할 정도로 수행평가가 심정적으로 싫거나 이해되지 않는 학생들도 있다.²⁰⁾ 이러한 불신은 비단 물론 국어과의 경우만은 아닐 것이다.

결국 어느 평가든지 단일 평가에 의해서 인간을 평가하는 것은 문제가 있다는 원론은 중요하다. 그렇기 때문에 다면적으로 평가할 수 있는 복수 평가 도구를 사용한다. 이때 각 평가도구의 장단점과 특성을 이해하고 평가 내용에 맞게 적절히 활용할 수 있어야 한다.

수행평가가 보완적 평가로서 장점이 있고 이를 시행해야 한다면, 교수·학습과 별개의 창의성을 요하는 어렵고 낮은 것으로 보이기보다는, 수행평가에서 필요한 수행 과정이 국어과 교수·학습의 단계에서 학습되거나 시범 보여질 수 있도록 배려하는 것이 중요하다. 이는 수행평가에 대한 부정적 인식을 변화시키는 데도 필요하다.²¹⁾

20) 인터넷 블로그에 ‘수행평가 필요성?(네이버, 2007.06.18)’이라는 제목의 고등학생 질문이 다음과 같이 떠있다. 이 질문은 답답한 심정의 토로에 가깝다. 이에 상당히 많은 댓글이 붙어 공감대가 형성되어 있음을 보여주었다. 사교육을 받은 학생의 경우 수행평가 적용률 30% 수준에서도 다른 학생들에 비해 유리하다는 점 때문에 평가의 공정성과 타당성에 의문이 제기된다.

21) 그럼에도 여전히, 각기 다른 평가 도구를 어느 정도로 반영할 것인지에 대한 반영률은 문제로 남아있다. 일반적으로 교수·학습 상황에서 강조되었거나 그 평가의 과정이 학습된 지식인가 혹은 학습자에게서 순전히 처음으로 구성된 것인가에 따라, 전자는 배점을 높이고, 후자는 상대적으로 낮게 한다. 이때, 구성적이라는 점에서 이전에 학습된 지식이 아닌 창조적인 지식이란 뜻이지, 지식의 구성 과정은 학습된 것이라 할 수 있다.

나. 수행평가 도구 개발에 대한 지식 부족

선다형 평가와 마찬가지로, 수행평가 과제를 설계하고 검사하기 위한 구체적인 절차와 방법에 대한 이론이 부족하다. 예컨대 포트폴리오 평가는 ‘과정 활동철’과 ‘결과 활동철’로 나뉘듯이, 수행의 결과뿐만 아니라 수행의 과정도 드러난다. 즉 교수·학습 과정과 결과의 모습을 누가적으로 보여준다는 의미에서 하나의 교수·학습 과정안이라 할 수 있다. 이를 평가 자료로 활용하기 위한 목적으로 평가 계획을 구상한다면, 하나의 중장기적인 교수·학습 ‘프로그램’을 설계하는 것과 같다. 특히 전문적인 평가의 내용 및 요소, 평가 방법 및 평가 도구, 평가 과제별 평가 기준, 채점 방식 등을 체계적으로 결정, 기술해야 한다.

■수행형(특히 서술형) 평가문항의 채점기준표 제작

수행형 평가 문항의 경우 가장 큰 고민거리는 채점기준표를 만드는 것이다. 특히 서술형 문항인 경우 학생들이 작성해내는 답이 상당히 다양해서 채점할 때 난처한 경우가 많다. 만약 부분점수를 부여하려면 타당한 채점기준을 제시해야 하는데, 사전에 채점기준을 만드는 것이 힘들 뿐 아니라, 만들어 놓고도 내가 과연 타당하고 정확하게 만들었는지 확신이 서지 않는다. 만약 10점 만점인 서술형 문제에서 10점짜리 답부터 1점짜리 답까지 채점기준표를 만든다면, 어떻게 해야 하는지 타당한 방법을 배우고, 연습해 보았으면 좋겠다.(M고)

수행평가는 반드시 교사 간에 정해진 수행평가의 하위 방법과 구체적인 채점 방식을 공유해야 한다. 즉 채점자의 동질 자질 관리를 위해서도 수행평가의 내용에 대해 협의를 통해 충분히 의사소통해야 한다. 다음은 이와 관련된 사례이다.

■수행평가의 객관도와 타당도 : 제고 방법

수행형 평가가 도입될 때부터 가장 많이 많았던 부분이 평가의 타당도이다. 동일한 학생의 답안에 대해 A, B, C 교사가 각각 채점했을 때 평가점수가 다르게 나올 수 있다는 것이다. 그런데 보통 교과담당 교사가 한 반을 전

담하는 경우가 많기 때문에 해당 교사의 스타일에 따라 유리한 학생과 불리한 학생이 생길 수 있다. 이러한 문제를 극복할 수 있는 방법은 없을까?

학교 현장에서 위와 같은 고민으로 학생이나 학부모의 원성을 사거나 민원을 살 일은 사실상 별로 없다. 적어도 내가 근무하는 학교에서는.

그런데 학교나 교육청에서 독후감 등 각종대회가 있어서 공동 심사를 할 경우에 채점자에 따른 채점결과의 차이가 상당히 크다는 점을 발견하게 된다. 어떤 학생의 글에 대해 나는 80점을 주었는데, 다른 3명의 심사위원은 95~100점을 준다든가, 반대로 나는 100점을 주었는데 다른 심사위원들은 최하점을 주는 경우도 가끔씩 있다. 대회의 경우 최하점이나 최고점의 점수를 빼고 중간점수를 평균 내는 수월한 방법이 있긴 하지만, 이런 경우에 불현듯 학교에서 하는 수행평가가 떠오른다. 혹시 내가 채점해서, 100점 받은 학생이 80점을 받은 것은 아닌지.(M2고)

■수행평가의 평가기준 : 평가 기준과 채점 기준

특히 학생들에게 완결된 한 편의 글을 쓰게 할 때 많이 느끼는 점이다. 계획하기에서 시작해서 고쳐쓰기를 거쳐 한 편의 글을 완성하게 하면서, 피드백을 하고 계속 자신의 글쓰기 과정을 조절해 나가는 과정을 연습시킨다. 그런데 학생이 40명이라면 40명의 반응과 중간 과정, 결과물의 수준이 모두 다르다. 만약 결과물만 가지고 평가한다면 40명 학생이 수준에 따라 변별이 되겠지만, 중간에 거친 과정을 계량화해서 평가하기는 쉽지 않다. 주로 학생들이 교사의 조언을 듣고 수정해 나가는 모습을 보고 점수를 주기는 하지만 과연 그렇게 주는 점수가 타당하고 정확한 것인지 모르겠다. 늘 하고 있으면서도 항상 의문이다. 잘하고 있는가?(M고)

이들은 모두 평가의 표준이 부재한 상황에서 생기는 경우들이다. 수행평가의 경우, 예컨대 토론이나 구술면접과 같이, 실제 언어 수행 장면을 평가하는 경우, 평가 내용이 제대로 명시되어 있지 않으면, 실제로 평가해야 할 능력과 평가하지 않을 부분의 혼동이 생겨 평가하지 않을 내용 때문에 평가가 영향을 받는 경우가 생긴다. 이때 채점상에 개입되는 많은 변인들과 평가 구인들을 구분, 경계하지 않으면 안 된다. 평가 구인이란 평가에 의해 측정되는 잠재능력 또는 자질로서, 일반적으로 평가에는 측정하고자 하는 능력과는 관련이 없는 많은 요인들이 개입되기 때문이

다.²²⁾

뿐만 아니라, ‘평가 내용’이 설정되어 있더라도, ‘평가 등급(levels)’에 따른 평가 기준(standards)이 기술되어 있지 않으면 역시 주관 평가를 하게 된다. 위의 예에서처럼 ‘내용, 조직 또는 구성, 표현’의 평가 영역을 나누고, 각 평가 영역마다 하위 평가 요소들을 설정한 뒤, 채점을 위한 등급을 설정하고, 각 등급마다 평가 기준을 기술한다. 예컨대, 영역별로 3단계 등급을 설정했다고 한다면 작문의 ‘내용’면에서 level1~level3은 각각 어떤 능력을 보여야 하는 것인지에 대한, 평가 준거를 포함하는 표준안을 기술하고 그에 따라 채점해야만 공정성을 확보할 수 있다.

또한 평가 내용이나 방법이 목표하는 수행을 이끌어내기에 적합해보일지라도 평가 절차의 다른 요소들이 수험자의 능력에 대한 추론을 그르치게 할 수 있다. 채점과정에는 위와 같은 많은 변인들이 개입된다. 그래서 읽기와 쓰기 같은 수행평가에서는 채점에 대한 연구가 타당도 검증의 일부로 되어 있다(McNamara, 2001 : 64). 일반적으로 수행의 상황이 복잡할수록 채점에 관한 타당도를 위협하는 요소들이 많아지게 된다.²³⁾

요컨대, 평가자들은 측정을 통해 추론하려는 수행능력에 대해 개념 정의되어 있어야 하며, 각 평가 영역의 하위 영역마다 잘 수행한 것, 예컨대 잘 쓴 글과 못 쓴 글에 대한 판단 기준으로서 평가 기준을 공유하고 있어야 하며, 이를 통해 엉뚱한 다른 능력에 현혹되거나 주관적 인상 평가에 빠지지 않도록 구인무관변이에 대비해야 한다.²⁴⁾

다. 채점자의 공정성

채점자들 간 공정성은 채점자 간의 점수 일치도를 살펴보는 것이다.

22) 이를 구인 무관 변이(construct irrelevant variance)라고 한다.

23) 1950년대와 1960년대부터 Lado는 이 부분을 강조하여 수행평가에 대해 특히 주의할 것을 지적했다(McNamara, 2001 : 15-25).

24) 이 외에도 너무나 적은 수행을 통해 수험자의 능력이 평가되는 경우, ‘구인의 과소 표현(construct under-representation)’도 주의해야 하는데, 이 역시 피험자의 능력에 대한 원하는 만큼의 정확한 정보를 얻을 수 없다는 데서 수행평가를 변화시킨다.

채점자간 점수 일치도를 ‘채점자간 신뢰도’라고 하는데, 이는 상관계수(0~1 사이), 즉 채점자간 상관성이 높은 경우 신뢰도가 높은 것이다.

예컨대, 두 명의 채점자가 작문 논술에 대해 전혀 상반된 점수를 줬다고 하면, 한 사람의 채점 점수가 다른 사람의 채점 점수와 아무런 상관성이 없다.²⁵⁾ 즉 한 채점자의 점수로부터 다른 채점자의 점수를 예측할 수 없다. 이것은 채점 과정은 객관적으로 공정하게 이뤄졌을지언정, 채점 그 자체는 공정한지 아닌지 알 수 없게 된 경우이다. 즉 신뢰도가 낮은 평가이다. 반대로, 이상적인 채점의 경우는 한 채점자의 점수를 보면 다른 채점자의 점수도 대략 짐작할 수 있는 경우이다. 두 사람의 채점 점수가 똑같은 경우, 채점자간 일치도, 두 점수의 상관계수가 높아 채점자간 신뢰도는 아주 높게 된다. 우리는 그러한 채점에 대해 이상적인 채점이면서 공정하다고 생각한다.

최근 국어과의 경우, ‘말하기, 듣기, 읽기, 쓰기’가 학교 안팎에서 경시대회 형태로 수행 평가가 이뤄지고 있다. 이로 인해 평가자간 신뢰도가 매우 중요한 요건이 되었다. 물론 평가자 간의 일치를 위한 사전 협의의 과정을 거쳐 채점자간의 신뢰도를 높이는 것이 가장 중요하다. 이를 위해서는 언급한 평가 내용 영역, 평가 내용의 요소화, 평가 등급별 평가 기준의 설정 등이 선행되거나 협의되어야 한다. 평가자간 어느 정도 일치가 되면 복잡한 수행평가에서는 협의 평가도 가능하다.

라. 다양한 평가 방법의 개발

특히 고등학교로 갈수록 수행평가 시행율은 낮고 방법도 제한적이다.²⁶⁾ 평가의 신뢰도나 타당도 등이 문제 되지 않는다 하더라도, 고등학교

25) 이 경우, 일치하지 않는 부분은 어떻게 처리하는가가 관건이다. 예컨대, 논술의 경우도 채조의 경우처럼 여럿이 평가할 경우에는 최고점수와 최소점수를 뺀 나머지 점수의 평균으로 할 수도 있다.

26) 이에 대해서는 김진곤·김혜영(2005)의 고등학교 수행평가 실태조사를 참조할 수 있다. 실태조사와 관련된 연구들이 주로 중학교를 대상으로 하고 고등학교를 대상으로 한 것이 많지 않은 이유는 고등학교에서는 다양한 반응을 조사할 만큼 수행평가가 활

라는 ‘특수’ 상황에서 학생들에게 더 어렵고 낯선 과제를 내주는 것은 학습 부담을 가중시키는 것이고, 과제가 복잡할수록 교사도 채점에 상당히 많은 시간이 소요된다는 점 때문에 수행평가는 다소 익숙하면서도 단순한 것, 평가 과제를 평이하게 할 수 있으면서도 융통성 있게 변형 가능하도록 포괄적인 것으로 설정되는 경향이 있다.

김진곤·김혜영의 연구 결과(2005 : 130)에 의하면, 연간 수행평가 계획서를 근거로 할 때, 고등학교에서 가장 많이 하는 수행평가 방법은 ‘독서 노트, 포트폴리오, 서답형, 논술형, 관찰법’ 등등이다. 또 서민원(2001)과 박순경 외(2003)의 연구에서도 국어과의 수행평가 방법은 ‘서술·논술형, 포트폴리오, 구술, 토론·발표’ 등이다. 또한 학교에서 제시하고 있는 ‘학교별 평가 연간 계획표’는 대부분 평가 방법과 도구 간의 위계나, 수준을 고려하지 않고 때로는 지엽적으로, 때로는 포괄적으로 제시하고 있다. 예컨대 ‘독서노트’, ‘독서일기’, ‘문학감상기록장’을 제시하는가 하면, ‘포트폴리오’만 제시하기도 한다.

이러한 수행평가가 제대로 이뤄지기 위해서는 수행평가에 대한 해석으로서의 교과 수업에 맞는 적용안이 개발되어야 할 것이다. 즉, 국어과 수업 상황과 평가 상황을 연계하여 실시할 수 있는 다양한 수행평가 방법이 상세화 되고, 김진곤·김혜영(2005 : 116)의 연구에서 지적했듯이, 피평가자인 학습자에게 수행평가를 소개하고, 수행평가에 적응할 수 있도록 수행평가 모형과 자료를 마련하는 작업이 선행된다면 수행평가도 훨씬 안정적으로 시행될 수 있을 것으로 기대된다.

발히 적용되지 않기 때문이기도 하다.

실제로 수행평가 반영 비율은 학교마다 자율적으로 정하도록 하고 있는데, 국어과의 경우, 지필평가(중간·기말고사) 대 수행평가의 비율이 20%에서 60%까지, 평균 30~40% 정도 상당히 높게 적용된다

3. 맺음말

평가의 중요성에 비해 평가의 실재에 대해 거의 이해되고 있는 것이 없다. 평가에 관한 많은 논문들이 서두에서 평가에 대한 인식이 변화되어야 한다고 역설한다. 평가는 인간 이해를 목적으로 하기 때문에, 학생이 남겨 놓은 낙서 한 줄, 일기 한 토막, 대화 한 마디가 모두 그를 이해하는 단서가 될 수 있다는 점에서 교육평가의 자료와 대상, 시간은 거의 무제한적이라고(김종서 외, 1997 : 356-359, 류수열, 2004 : 59-69 재인용)한다. 맞는 말이다. 이의를 제기할 수 없을 만큼 원론적이고 그래서 지나치게 낭만적이다. 그러나 인간 이해를 목적으로 주변 친구를 평가하지는 않는다. 또는 친구를 평가하면서 이해했다고 하지 않는다. 평가를 통해 대상을 이해할 수도 있지만, 평가는 이해의 한 방법일 뿐, 평가는 제도적 통제 장치라는 속성이 내재한다는 점에서 보다 비판적으로 이해되어야 한다. 즉, 어떤 것이든 평가의 자료와 대상이 될 수 있지만, 대상의 무엇을 어떤 기준에 따라 바라볼 것인가가 결정되지 않으면 이해 또는 경우에 따라서는 오해는 될지 몰라도 평가는 아니다.

그렇기 때문에 평가는 체계적으로 설계되고 엄격하게 적용되어야 한다. 이는 형식적 평가나 객관식 평가를 지지한다는 의미는 전혀 아니다. 지필평가는 대안적 평가로서 수행평가는 그것이 평가로서 작용하기 위해서는 기준이 명확해야 한다. 특히 기존의 평가 관행을 바꿀 목적이 있다면, 새 평가 제도가 뿌리내리도록 시행초기에는 원론에 그치지 않는 상세하고 친절한 평가의 표본(standards)이 제시되었어야 했다.

지난 10년간 수행평가를 비롯한 다양한 대안적 평가방법들을 소개하면서 일관되게 평가에 대한 관점의 변화를 얘기하고 있다. 그것이 단지 학문적 유행이나 외국의 모방에 따른 것이 아니라 우리의 미래사회가 요구하는 인재들의 핵심 역량을 판단할 수 있는 준거로서²⁷⁾ 적절한 것이라

27) 교육과정평가원 연구보고(RRC 2007-1)에 의하면, 미래사회에 필요한 한국인의 핵심역량은 ‘갈등조정능력, 문제해결능력, 의사소통능력, 정보처리능력, 시민의식, 창의력, 다

면, 그리고 그러한 교육의 질적 변화를 이끌어갈 견인차로서 필요한 것이 라면 진정으로 받아들일 준비를 해야 할 것이다. 선택의 문제가 아니라, 당위의 문제라면 각급학교 수준에서 응용과 재창조가 가능해질 때까지 책임을 가지고 평가 방법에 대한 지식이 자세하고 친절하게 안내되어야 할 것이다. 우리는 평가의 질적 변화를 왜 받아들이지 않으면 안 되는지에 대한 철학적 논의의 진지성이 결여되어 있다.

동시에 선다형에 대한 무조건적인 매도, 수행평가에 대한 과장된 기대, 특히 고등학교의 경우 현실 논리에 맞지 않는 무리하고 성급한 적용 등을 되짚어볼 때다.*

문화이해능력, 자기주도적학습능력, 삶의 향유능력' 등으로 설정하였고, 이는 향후 교육과정 설계에 비전으로 제시될 것이라고 했다.

* 본 논문은 2008. 6. 29. 투고되었으며, 2008. 7. 6. 심사가 시작되어 2008. 7. 27. 심사가 종료되었음.

■ 참고문헌

- 대구광역시 중등연수원(2008), 평가연수 자료집, 중등연수원.
- 김경희(2007), 수행평가 질 관리의 실제적 이슈와 과제, 교육과정 운영에서의 수행평가 적용실태와 개선 방안, 연구자료 ORM 2007-8, 한국교육과정평가원.
- 김진곤·김혜영(2005), 고등학교 국어과 수행평가 실태조사 연구, 교과교육연구, 26권 1호, 조선대학교.
- 강성우 외 역(2001), 언어평가(Tim McNamara, Language Testing), 박이정.
- 대구광역시교육연수원(2008), 2007 중등 국어과 평가도구 개발능력 향상 직무연수 분임토의보고서, 대구광역시교육연수원.
- 류수열(2004), 중등학교 국어시험의 현황과 개선 방안, 국어교육학연구 20집.
- 박인기 외(1999), 국어과 수행평가, 삼지원.
- 백순근 외(1999), 고등학교 국어과 수행평가의 이론과 실제, 연구보고서 RRE99-1-2, 한국교육과정평가원.
- 성태제(1996), 문항 제작 및 분석의 이론과 실제, 학지사.
- 윤현진 외(2007), 미래 한국인의 핵심역량 증진을 위한 초·중등학교 교육과정 비전 연구(1)－핵심역량 준거와 영역 설정을 중심으로, 연구보고서 RRC 2007-1, 한국교육과정평가원.
- 이도영(2001), 국어과 교육 평가 모형, 『국어교육학연구』, 제13집, 국어교육학회.
- 정구향(2005), 21세기 국어과 평가의 발전 방향, 새국어교육 71호.
- 천경록(1999), 국어과 수행평가와 포트폴리오, 교육과학사.
- 최지현(2000), 선택형 지필평가의 한계와 가능성, 국어교육 103호.
- 최미숙(2000), 국어교육평가의 원리와 실제, 국어국문학 126호.

<초록>

고등학교 국어과 평가의 문제점과 개선 방안
—평가 사례를 중심으로—

김혜정

평가는 교육과정의 내적 구조 속에서만 이해되지 않는다. 사회적 측면에서 그 어는 것보다 제도적이고 정치적인 특성을 여실히 드러낸다.

본 연구는 현행 고등학교 교육의 특수성을 주지하고, 현재의 사회문화적 맥락에서 국어과 평가가 어떻게 이뤄지고 있으며, 개선의 여지가 있다면 무엇인지 살피보고자 한다.

‘수행평가’와 ‘국어과 평가의 문제점’이라는 두 가지 주제에 대해서는 이미 많은 연구가 수행되어왔다. 원론적 수준이든 아니든 문제점에서 개선방안으로 나가는 방식은 더 이상 의미가 없을 것이다. 그럼에도 불구하고, ‘고등학교’ 상황에 보다 더 초점을 둔, 교사들의 목소리를 통해 이미 우리가 문헌으로 알고, 분위기로 짐작하고 있었던 문제들을 확인해보고, 또 각 경우마다 개선안을 붙여보았다.

선다형 평가에 대해서는 문항 개발과 관련된 지식과 경험이 부족하고, 수행평가에 대해서는 평가 내용 및 방법에 대한 지식이 부족하였다. 이를 해결하기 위해서는 우선, 평가자간 사전 평가협의 과정을 통해 평가 계획을 철저히 수립하고 평가에 대해 공유하는 것이 필요하다. 또한 수행평가를 그 자체로 완결된 평가로 보기보다는 선다형 지필평가에 대한 보완적 평가로서, 상보적 역할을 할 수 있도록 구성하는 것이 중요하다.

【핵심어】 수행평가, 선다형평가, 보완적 평가, 평가자간 신뢰도, 타당도, 사전 협의 과정, 평가 공유

<Abstract>

Problems and Their Improvements of Korean Language
Evaluation in High School

—‘Opinion of Evaluation Case’ Study—

Kim, Hye-jeung

Evaluation is considered as systemic and politic aspects in socio-cultural aspects as well as internal structure of curriculum.

This paper's aims are to survey the ways how Korean language evaluation is performed in this Korean social cultural context and how the problems of evaluation is improved.

So far, two themes of 'Performance Assessment' and 'Problems and their Improvements of Korean Language Evaluation in high school' are studied much. And so its study may not have significance, whatever it is principle or progressive.

However, its findings show us that lack of expert-knowledges and experiences in development of multi choice test items, and that lack of expert-knowledge of evaluation in various performance assessment making.

To solve these problems, first, pre-conference of assessors is planned and harmonious communications is accomplished. And there, plans of evaluation is achieved and viewpoints of evaluation is mediated. Performance assessment and multiple choice test are operated complementary.

[Key words] performance assessment, multi-choice test, alternative assessment, inter-rater reliability, validity, prior negotiation process, communication of test-context