

초등학교 국어과 시험의 문제점과 개선 방안

서 혁*

〈 차 례 〉

1. 머리말
2. 초등학교 국어과 평가 실태
3. 초등학교 국어과 평가문항 분석
4. 초등학교 국어과 평가체제 개선 방안
5. 맺음말

1. 머리말

최근 국어과 평가와 관련하여 수행평가를 비롯한 다양한 방법들이 소개되고 적용됨으로써, 학교 현장에서의 국어과 평가도 매우 정교화 되고 체계화 되어 온 것이 사실이다. 비록 대학 입학을 위한 입시 준비와 사교육에 매달리고 있는 것이 우리의 교육 현실인 것은 사실이지만, 교육 이념과 정책, 기능의 변화에 따라 '시험' 제도 역시 큰 변화를 거쳐 왔다. 시험에 대한 교육 담당자들의 인식은 물론 학부모와 일반인들의 인식도 크게 변하고 있다.

교육열이 강한 우리 사회에서는, 시험 제도가 어떻게 변하든지 간에, 교육에 대한 관심보다는 '시험'에 대한 관심이 더 컸으며, 이는 현재도 변하지 않는 사실인 듯하다. 최소한 학부모의 처지에서는 그러한 것이 사실이고, 교육계 역시 이러한 사회적 현상과 관심을 간과할 수 없었

* 전주교육대학교 국어교육과

다.

그러나 국어 '시험'에 대한 관심에 비해 국어교육학적 체계로서의 '평가론'에 대한 관심과 노력은 그리 크지 못했던 것이 사실이다.¹⁾ 국어 교육평가론의 부재에 대한 문제 제기는 이도영(1999)에서, 국어교육 메타평가의 필요성과 국어교육평가론(학)의 필요성과 가능성에 대한 논의는 김창원(1999) 등에서 이미 제기된 바 있다. 특히 국어과 수행평가와 관련된 활발한 논의들 역시 국어과 평가 방법과 평가론의 심화에 기여한 바가 크다고 생각된다.

국어과 평가의 개념과 유형은 평가의 대상, 목적, 내용, 주체, 평가가 이루어지는 공간, 방법과 매체 등에 따라 매우 다양하다. 흔히 말하는 '시험'이란 피시험자의 학습 성취 정도나 능력, 수준을 확인하기 위해 이루어지는 다양한 유형의 의사결정 활동으로서 평가의 하위 개념이다.²⁾ 좁게는 수업 중에 교사가 학생들을 대상으로 국어과 학습 진단 혹은 형성 평가를 위한 쪽지 시험이나 받아쓰기는 물론, 국어교육 프로그램 개선을 위한 (광의의) 국어과 교육과정 평가 등도 포함된다. 제도적인 학교 교육 차원에서 이루어지는 '국어 시험'은 학습자들의 국어 학습 성취 정도와 국어(사용)능력을 확인하기 위한 의사결정 행위 혹은 그 과정을 가리킨다.

이 글에서 다루고자 하는 초등학교 국어 시험과 관련한 논의는 전술한 국어교육 평가의 큰 틀 속에서 매우 작은 한 부분이다. 넓은 의미의

1) 물론 최근에 국어과 수행평가에 대한 활발한 논의(박인기(1997), 천경록(2000), 최미숙(1999), 진영란(2000) 등)와 함께 국어교육의 메타평가에 대한 논의들(이삼형(1999), 김창원(1999), 이도영(1999) 등)이 있었다. 이러한 논의들은 국어교육평가학의 중요한 출발점이 될 수 있으리라 생각된다.

2) 이 때의 시험은 일반적으로는 'test'에 가깝다고 볼 수 있다. 흔히 말하는 'measurement, evaluation, assessment, test'의 구분은 평가 현상에 일대일로 대응되지 않는 경우가 많다. 이러한 개념 자체 구분 자체가 평가 현상에서 출발했다기보다는 평가에 대한 관점, 가치관을 반영하고 있기 때문이다. 실제로 구체적인 평가 현상의 장면을 분석하자면 이들 용어들은 서로 중첩되어 나타나기 일쑤이다. 황정규(1987)의 구분에서도 볼 수 있듯이 이들은 '측정관, 평가관, 총평관'과 같이 모두 '(평가)관'으로 번역되고 있다.

‘평가’ 속에서 ‘시험’이 차지하고 있는 비중은 미미하지만, 현실적으로 ‘시험’이 끼치는 영향력은 대단히 큰 것도 사실이다.

요컨대, 국어교육 평가의 의미와 공간은 매우 광대하나, 그 출발점은 ‘국어 시험’에서 시작하지 않을 수 없다. ‘국어 시험’ 자체가 매우 광범위하고도 중요한 국어교육 현상일 뿐 아니라, 국어교육 평가 논의 역시 국어교육 현장이라는 토대 위에서 이루어질 필요가 있다는 것이다.³⁾

이러한 관점에서, 초등학교 국어 시험과 관련한 현황과 문제점을 설문조사, 면담, 시험지(평가 문항)를 중심으로 분석하고, 그 개선 방안을 제시해 보고자 한다.

2. 초등학교 국어과 평가 실태⁴⁾

초등학교에서의 국어과 평가 현황 분석을 위한 구체적 방법으로 설문조사, 면담, 시험지 분석의 방법 등을 이용하였다. 설문조사는 2003년 3월초를 전후하여 실시하였으며, 주요 대상은 초등학교 교사(총 189명), 초등학생(총 413명), 초등학생 학부모(317명)의 세 집단을 중심으로 총 900여 명에 이른다. 면담은 2003년 12월부터 2003년 2월에 걸쳐 주로 초등학교 현장의 교사와 장학사 등 약 15명을 대상으로 수시로 진행하였다. 시험지 분석은 서울(1), 경기(2), 전북(9) 지역의 총 12개 교를 대상으로 하여 각 학교에서 2001-2003년 기간에 치러진 각종 학력평가, 월말고사, 중간고사, 기말고사, 성취도 평가, 수행평가 문항들을 중심으로 살펴보았다.

3) 이와 관련하여 이도영(1999)에서는 ‘평가 연구의 속성과 현장 교사들의 요구에 비추어 볼 때 평가 연구에서 현장 연구는 필수적이나, 대부분의 국어과 평가 연구들은 현장과 밀착된 평가 연구를 거의 도외시키고 있다.’는 점을 지적하고 있다.

4) 국어과 평가(시험) 실태 분석과 관련한 구체적인 설문조사 결과는 국어교육학회 제27회 학술발표대회 자료집을 참조하기 바람. 여기에서는 지면관계상 주요 내용만을 발췌함.

3장에서 주로 살펴볼 시험지(평가 문항) 분석 대상 학교는 모두 13개 교로 서울 S초등학교(42학급), 경기도 용인 S초등학교(30학급), 경기도 시흥 O초등학교(12학급), 전북 지역의 M초등학교(6학급), G초등학교(12학급), 익산의 I초등학교(30학급), 전주 B초등학교(30학급), J초등학교(6학급), K초등학교(36학급), Y초등학교(40학급), 남원의 A초등학교(6학급) 등이다. 대체로 학교 규모를 다양화하고자 시도했다. 학교 규모에 따라서 평가 운영 실태가 다를 수 있다고 판단했기 때문이다. 그러나 지역적 안배는 이루어지지 못했다.

설문조사의 주요 내용을 요약하면 다음과 같다.

먼저 교사 대상 설문 조사 내용은 총 15개 문항으로 근무 경력, 학교 규모(전체 학급수), 근무 학교의 소속 행정 단위, 운영 학급에서 실시된 지필 평가 유형, 중간고사나 기말고사 평가문항의 구안 방법, 자주 사용하는 수행평가 유형, 수행평가 실시 횟수, 수행평가 문항 구안 방법, 수행평가 실시상의 어려운 요인, 지필평가와 수행평가의 적절한 반영 비율, 전반적으로 국어 능력 평가의 어려운 요인, 평가 결과에 대한 송환 방법, 생활통지표 상의 국어과 영역 기록 내용, 생활통지표 개선 요망 내용, 국어 능력 평가 관련 개선 필요 사항이 주요 내용이다. 설문 조사 관련 내용에 대한 적용 기준 시점은 2003년도 재직 당시를 기준으로 하도록 했다.

학생 대상 설문 조사 내용은 모두 8개 항목으로, 학년, 생활통지표 일람 여부, 생활통지표상의 국어과목 내용 상세화 정도 판단, 생활통지표 개선 요망 사항, 국어 시험 빈도에 대한 판단, 경험한 국어 시험 유형, 가장 부담이 되는 국어 시험이나 과제 유형, 기타로 구성되었다.

학부모 대상 설문 조사 내용은 모두 10개 문항으로, 연령, 거주 지역 행정 단위, 자녀 학년, 자녀의 생활통지표 일람 여부, 생활통지표의 국어과 관련 내용 상세화 정도 판단, 생활통지표 개선 요망 사항, 자녀의 과제 해결 도움 지원 정도, 자녀의 국어 능력에 대한 판단, 자녀의 국어 능력 판단에 대한 근거, 기타로 구성되었다.

학급 규모는 평가 문항 제작, 평가 횟수, 유형, 방법 등과 밀접한 관

련을 지낸다. 설문 조사에 포함된 교사들이 재직한 학교의 규모는 31학급 이상이 절반을 차지했는데, 이는 교사들의 경력과도 밀접한 관련을 갖는 것으로 추측된다. 그밖에 25-30학급이 28%, 6학급 이하 7%, 7-12학급 6%, 19-24학급 4%, 13-18학급이 3%이다.⁵⁾

2.1. 주요 지필평가 유형 및 평가문항 구안 방법

피설문 교사들이 교육한 학급에서 실시되었던 지필평가 유형으로는 기말고사, 시도 학력평가, 중간고사, 학년말고사, 받아쓰기, 전국 단위 성취도 검사, 월말고사, 쪽지시험 등으로 매우 다양하게 나타났다. 그 중에서 기말고사, 시도학력평가, 중간고사가 모두 15% 이상으로 가장 높게 나왔는데, 전북지역의 대다수의 초등학교에서는 기말고사 혹은 중간고사를 도학력평가로 대체하고 있음을 면접 과정에서 확인할 수 있었다.

교사들의 평가 문항 구안 방법은 동료교사 간의 분담 제작이 44%

5) 여기에서는 지면 관계상 교사 대상 설문 결과를 중심으로 주요 내용만을 소개하기로 한다. 구체적인 내용은 국어교육학회 제27회 학술발표대회 자료집을 참고할 수 있다. 국어교육학회 홈페이지(<http://www.koredu.or.kr>)에서 자료를 확인할 수 있다.

로 가장 높게 나왔고, 인터넷 자료 활용(16%), 시도 학력평가로 대체(14%), 사설 교육기관이나 출판사에서 구입한 평가지로 평가(13%), 직접 제작(12%) 순이 뒤를 이었다. 교사들이 평가 문항을 독자적으로 제작하는 경우는 전반적으로 드물었고(12%), 6학급 이하 규모의 학교에서는 거의 대부분 시도학력평가나 출판사를 통해 구입한 평가지들을 중간고사나 기말고사에 활용하였다. 분담 제작하는 경우는 최소 10학급 이상 특히 20학급 이상에서 이루어지고 있음을 확인할 수 있었다.

여기에서 한 가지 유의할 것은 기본적으로 초등학교 단계에서의 평가 문항 개발의 주역은 현장 교사들이라는 점이다. 즉, 직접 제작이나 분담 제작의 방법을 따르지 않더라도, 시도 학력평가나 출판사에서 개발한 문항들 역시 거의 전적으로 교사들이 중심이 되어 개발되고 있으며, 장학사나 일부 교감들이 출제에 참여하기도 한다. 인터넷에 올려진 자료들 역시 개발 주체는 현장 교사들이라고 보아야 할 것이다.

2.2. 자주 사용하는 국어과 수행평가 유형

국어과 수행평가와 관련하여 교사들이 자주 사용하는 평가 방법은 서술형·논술형(19%), 관찰평가(16%), 실기평가(16%), 누가기록철(14%), 일기장(12%), 조사연구보고서(6%), 구술시험(4%), 면접법, 토론법(각각 2%) 순으로 매우 다양하게 나타났다.

사실 일반적으로 알려진 수행평가 방법과 국어과의 고유한 특성을 살린 수행평가 방식에는 상당한 차이가 존재한다. 국어과 수행평가와 관련한 좀더 구체적인 합의(컨센서스)와 교사들의 수행평가 개념 인식에 대한 면밀한 연구가 필요하다고 생각되지만, 이번 조사에서는 포함시키지 못했다. 설문 항목 구성상의 차이가 조금 있기는 하지만, 진영란(2000)의 연구와 비교할 때 교사들이 선호하는 국어과 수행평가 방법에 대한 응답률의 차이는 상당한 시사점을 제공한다. 즉, 진영란(2000)에서는 단답형 서술형의 방식이 55%, 실기평가 16%, 관찰평가 10% 순으로 나타났다. 이러한 결과 비교에 따르자면 최근 몇 년 사이에 초등학교 현장에서 국어과 수행평가가 어느 정도 정착되고 있으며 교사에 따라 다양한 방법이 적용되고 있음을 알 수 있다.

2.3. 국어과 수행평가 실시 횟수

매월 1-2회가 63%로 가장 높게 나왔고, 학기당 1-2회가 26%, 매주 1-2회 6%, 거의 실시하지 않는다. 2%, 거의 매시간 1%로 나타났다. 여기에서 말하는 수행평가의 개념은 학생들의 공식적인 국어 능력 평가를 위한 것으로, 선택형 지필고사 형식을 가능한 한 배제한 실제 국어 활동에 초점을 두고 있다. 그러나 수행평가에 대한 교사들의 개념 차이가 존재하는 것이 사실이다. 이와 관련한 좀더 정밀한 조사연구가 필요하다.

2.4. 수행평가 문항 제작 방법

인터넷 검색 자료 활용이 43%로 가장 높았고, 다음으로 교육청 등 기관 제공 자료 활용(20%), 직접 제작(14%), 동료교사 사례 활용(13%), 사설 교육기관이나 문제집 자료 활용(9%) 순으로 나타났다.

2.5. 국어과 수행평가시 가장 어렵다고 느끼는 점

평가 기준 설정이 35%로 가장 높았고, 문항 제작(25%), 채점 작업(18%), 평가 결과 정리(17%)가 뒤를 이었다. '평가 결과에 대한 학생이나 학부모의 수긍에' 대한 반응은 2%로 연구자의 예상과 달리 낮게 나왔다. 이는 초등학교 단계에서는 수행평가와 관련하여 교사와 학생 및 학부모 간의 미묘한 갈등이 심각한 현상은 아님을 보여준다.

2.6. 지필평가 대 수행평가의 적절한 반영 비율

피설문 교사들에게 적절하다고 생각되는 비율을 직접 기입하게 하여 평균값을 구하는 방식으로 이루어졌다. 대체로 초등학교 교사들은 국어과에서 수행평가 대 지필평가의 반영 비율을 53% 대 43%로 수행 평가에 조금 더 비중을 두었다. 이는 현재 초등학교 상황에 따라 수행평가의 비중을 30%에서 70% 정도를 반영하고 있는 것을 감안할 때, 흥미 있는 결과라고 할 수 있다. 연구자는 소규모 학교, 소인수 학급일수록 수행평가 비중이 높게 나올 것으로 예상했으나, 분명한 상관관계를 보

여주지는 못했다. 학교 규모나 학급인원보다는 교사들의 국어교육 철학과 가치관의 영향에 따른 개인차를 반영하는 것으로 보인다. 어쨌든 초등 국어과 평가에서 대체로 교사들은 수행평가에 더 비중을 두고 있는 것으로 나타났으며, 이는 수행평가가 어느 정도 정착되고 있음을 보여주는 것으로 해석할 수 있다.

2.7. 국어 능력 평가시 가장 어려운 점

교사들은 국어 능력 평가와 관련하여 평가 문항 개발이 가장 어렵다고 느끼고 있으며(33%), 그 다음으로 평가 기준 설정의 어려움(27%)을 들고 있다. 이는 어느 정도 예상했던 반응이다. 평가자료 확보(21%), 평가 결과 정리 작업(17%)이 뒤를 잇고 있다.

2.8. 평가 결과 송환 방법

평가 결과 송환 방법으로는 시험지나 결과표를 나눠주는 방법(44%)이 가장 높았고, 학기말 생활통지표(26%), 구두로 알려줌(14%), 시험지나 결과표를 나눠주고 학부모 확인을 받게 하는 방법(13%) 등으로 나타났다. 기타 의견으로는 학생들에게 개별적으로 시간을 내어 피드백해 준다는 의견이 일부 있었다. 이는 연구자가 평가결과 송환의 형식적인 측면에 초점을 맞춘 면이 있기는 하지만, 대체로 평가 결과의 송환이 적극적으로 이루어지지 못하고 있음을 보여준다. 국어과 수행평가의 송환 작용과 관련하여, 진영란(2000)의 조사 결과에 따르면, 60% 이상의 학생들이 자신의 수행 결과를 확인하지 않고 있으며, 자신의 국어 능력을 알지 못한다는 반응에서도 드러나고 있다.

초등학교에서 평가결과 송환과 관련하여 가장 공식적인 방법은 생활통지표이다. 56%의 교사들이 국어과 6개 영역 중 특히 우수한 영역이나 기능을 중심으로 기술한다고 답하고 있다. 특정 영역의 우수한 면과 부진한 면을 모두 기술(15%), 전체 영역 중 우수한 영역을 중심으로(15%), 전체 영역에 걸쳐 우수한 영역과 부진한 영역을 모두 기술(13%)하는 경우가 뒤를 잇고 있다.

이러한 결과는 실제 현행 생활통지표상의 교과영역 발달 사항 기술과 관련하여 대부분의 학교들이 학습자의 우수한 성취 내용 중심으로 기술하고 있는 경향을 잘 보여준다. 이러한 경향은 석차와 성취 결과 중심에 관심을 갖는 우리 사회의 교육 문제 해소 방안이라는 점에서는 긍정적인 면이 없지 않으나, 반드시 '교육적'인 방법이라고만은 할 수 없다고 본다. 좀더 상세한 내용은 뒤의 4장의 평가 결과 송환 문제에서 언급하기로 한다.

2.9. 생활통지표 개선 필요 사항

생활통지표 개선 사항 필요 사항과 관련하여 36%의 교사들이 성취

가 뛰어난 부분과 부진한 부분을 모두 진술할 필요가 있다는 반응을 나타내고 있다. 항목별 체크리스트 방식 도입 필요(24%), 간편한 프로그램 개선 보급 필요(22%)가 영역별 상세화 필요(16%)를 앞서고 있는 것은 교사의 과중한 업무 부담과 그 경감책이 필요하다는 점에서 자연스런 결과라 할 수 있다.

궁극적으로는 교사의 업무 부담을 경감하는 방안을 마련하되, 평가 결과에 대한 상세화는 오히려 강조될 필요가 있다고 본다. 이에 대한 구체적인 논의 역시 후술하기로 한다.

2.10. 국어능력 평가에서 가장 시급한 개선점

애초에 연구자는 '평가문항 제작 능력 제고'와 '교사 전문성 신장', '성취 기준 상세화'에 대한 반응이 가장 높을 것으로 예상했다. 그러나 이들은 각각 8%, 19%, 13%로 나타나, 문제 은행 도입(30%)이나 평가 결과 정리 프로그램 개발(19%)의 뒤를 이었다.⁶⁾ 이러한 결과 역시 초등학교 교사들의 과중한 업무 부담을 반영하고 있는 것으로 해석된다.

기타 의견으로는 평가 결과 송환 방법과 관련하여 학습계획서와 평가서를 작성하게 하는 방법이 제안되기도 했다. 또 생활통지표의 (영역) 통합적 기술 방식이 필요하며, 평가 결과 작업의 체계화, 피드백의 구체적인 형식 마련이 필요하다는 의견이 제시되었다.

2.11. 초등학생들의 최근 1-2년간 생활통지표 일람 여부

초등학생들을 대상으로 한 설문조사와 관련하여, 최근에 생활통지표를 본 적이 있느냐는 질문에 놀랍게도 36%에 이르는 학생들이 '잘 모르겠다'고 반응했다.⁷⁾ 반면에 95%의 학부모들이 자녀의 생활통지표를 본 적이 있다고 반응했고, 본 적이 없다(4%), 잘 모르겠다(1%)로 나타났다. 자녀의 학습 성취도에 대한 학부모들의 높은 관심도를 나타내 준

6) 전문적인 국어교육 연구 과정을 밟고 있다고 생각되는 한국교원대 대학원 과정의 교사들을 비롯해서 몇몇 학교에서는 '교사의 전문성 신장'에, 근소한 차이기는 하지만, 가장 큰 반응을 보이기도 했다.

7) 초등학생 대상의 설문조사에서 피설문자 학년 구성은 5학년(38%), 6학년(32%)가 주류를 이루고 있다. 2학년(18%), 4학년(8%), 3학년(4%)이 그 뒤를 잇고 있다. 1학년은 설문의 특성상 조사 대상에서 제외하였다.

다. 이는 피설문 학생들이 설문의 내용을 제대로 이해하지 못했거나, 자신의 생활통지표에 대해 무관심한 결과를 반영한 것으로 해석될 수 있다. 그런데 질문의 내용이 특별히 어렵다고 판단되지는 않기 때문에, 무관심에 기인한 것으로 생각된다. 이러한 해석의 근거는 뒤의 학부모 대상 설문에서는 95% 이상의 학부모들이 '본 적이 있다'고 답하고 있는 것과도 관련된다. 최근의 생활통지표 기술 방식이나 내용은 학습자들이나 학부모에게 적절한 피드백 기능을 하지 못하고 있고, 이는 학생들로부터 큰 관심을 끌지 못하는 것으로 보인다.

2.12. (국어)생활통지표의 상세화 정도에 대한 초등학교 들의 반응

대부분의 학생들은 보통이다(54%)는 반응을 보이고 있으며, 자세하다(32%), 매우 자세하다(7%), 자세하지 못하다(6%), 전혀 자세하지 못하다(1%)가 뒤를 잇고 있다.

생활통지표에서 더 자세하게 알려주었으면 하는 내용은 무엇인가라는 질문에 성적 상승 정도(34%), 석차(31%), 과목 적성(21%), 특별활

동이나 재량활동 적성(11%)의 순의 반응을 보였다. 학생들은 대체로 과목 적성보다는 성적에 관심을 더 나타냈는데, 자신의 성취도 변화에 대한 관심이 높다는 것이 다행이기는 하나 석차에 대한 관심이 그에 못지 않다는 점도 눈에 띈다. 수도권 지역 특히 서울 지역 학생들의 경우 석차에 대한 관심이 타 지역 학생들에 비해 현저히 높게 나타났다.

앞의 설문조사 결과 분석 부분에서도 잠시 언급했듯이 초등학교 단계에서 국어과 평가 문항 개발의 핵심은 교사이다. 즉, 비록 각 학교 차원에서는 모든 교사들이 독자적으로 국어과 평가 문항을 개발하지는 않더라도, 전국단위의 성취도 평가나 도학력평가, 여러 출판사의 문항 개발의 주체가 기본적으로 교사들로 구성되기 때문이다. 여기에 장학사나 일부 교감들이 참여하게 된다. 2003년도의 전국단위 성취도 검사 문항 개발에 있어서도 초등 교사들이 출제하고, 일부 교수들이 검토 작업에 참여한 것으로 알려져 있다.

그렇더라도 실제로는 중간고사나 기말고사의 경우 출판사 등의 문항을 학교 차원에서 구입하여 평가를 실시하는 경우가 대부분이다. 특히 학급 규모가 작은 지역의 경우 이는 일반적인 현상으로 판단된다. 교사들이 독자적으로 혹은 분담하여 자체 제작하는 사례는 매우 드문데, 30학급 내외 규모의 학교에서는 자체 경시대회나 월말고사 형태로 자체 제작하는 사례가 상당수 있는 것으로 밝혀졌다. 그러나 이 경우에도 대부분 시중의 각종 문제집 등을 부분적으로 변형하거나 재사용하는 재편집 수준이 대부분인 것으로 밝혀졌다.

이들 평가 도구의 경우 문항 개발의 기술적, 내용적 차원에서 여러 가지 문제점들이 발견되고 있다. 즉, 평가 문항 혹은 채점과 관련한 타당도, 신뢰도, 객관도와 관련된 문제점이다. 전국 단위의 성취도 평가나 시도 단위의 학력 평가의 경우 심각성이 덜 한 편이기는 하지만, 문제점이 없는 것은 아니다.

초등학교 국어과 평가 문항에 대한 평가와 관련하여 초점은 일차적으로 교사에서 출발하지 않을 수 없다. 교사들은 실질적인 평가의 주체

이자, 평가에 대한 평가자이어야 하기 때문이다. 그리고 평가 결과를 피드백하고, 교육과정에 반영하는 실질적 주체이기 때문이다.

특히 현재 초등 수준의 평가 문항 개발 주체의 핵심은 어떠한 형태건 간에 교사가 그 중심에 놓여 있기 때문이다. 즉, 소규모 학교의 경우 비록 출판사에서 개발 제공하는 평가지들을 구입하여 평가를 치른다 하더라도, 그러한 평가 문항 개발의 주체는 또 다른 교사인 경우가 대부분이기 때문이다.

여기에서 우리는 교사의 평가 도구 개발 능력, 평가 도구에 대한 판단 능력, 수행평가 도구와 관련한 성취수준 판단 능력 (혹은 채점 기준 설정 능력 등)이 절대적으로 필요하고 중요하다는 점에 유의할 필요가 있다. 일차적으로 교사들의 평가 도구 판단 능력과 수행평가 문항들에 대한 성취수준 판단 능력에 대한 연수가 필요하다. 더 나아가서 초등교사들의 독자적 평가문항 제작 능력을 향상시켜 나가는 전문적인 연수 프로그램들이 개발될 필요가 있다. 특히 국어과의 평가 도구들은 국어사용과 관련되고, 또한 국어로 이루어진다는 점에서 평가 도구 제작과 관련하여 매우 정밀하게 다듬어져야만 하는 고도의 전문성을 바탕으로 한다. 대체로 각 학교에는 평가 담당 교사가 있는데, 이들에게는 필수적으로 국어과 평가문항 개발이나 메타평가에 대한 연수를 지원할 필요가 있다. 국어 평가 문항의 개발은 타 교과 문항 개발에도 기본적인 도움을 주기 때문이다.

3. 국어과 평가 문항 분석

초등학교 국어 시험지의 평가 문항 분석은 약 15개 교의 월말고사, 교내 경시대회, 중간고사, 기말고사, 도(기본)학력평가, 전국단위기초학력평가, 전국단위성취도평가 등의 문항을 바탕으로 하였다. 전반적으로 평가 문항의 타당도, 신뢰도, 객관도 면에서 매우 창의적이고 완성도가

높은 문항들이 주류를 이루고 있다고 판단되었다. 특히 도학력평가나 전국단위 평가 문항의 경우 학생들의 국어적 사고력을 측정하기 위한 창의적이고 완성도가 높은 문항들이 많았다.

그럼에도 불구하고 전반적으로 평가목표의 타당성, 지시문, 문두, 불충분한 지문과 정보, 답지 구성의 문제점, 오답 시비의 문제점, 조건화와 평가 기준의 미비, 성취 기준 설정의 문제점 등 여러 가지 문제점을 보여 주고 있다.⁸⁾ 이러한 문제점들은 평가 문항의 타당도, 신뢰도, 객관도를 떨어뜨리는 직간접적 원인이 된다. 구체적인 문항 예와 함께 살펴보면 다음과 같다.

3.1. 평가목표의 타당성

평가 목표나 의도가 국어과 교육과정과 거리가 먼 경우의 예이다. 다음의 문항 예 1-3은 국어과 교육과정이나 학년에 비추어 볼 때 적절한 평가 목표를 지닌다고 보기 어렵다.

〈문항 예 1〉 2002년에는 제17회 '월드컵 축구대회'가 우리나라와 일본에서 열렸습니다. 이 대회에 참가한 축구선수를 두 명만 쓰세요. (Y초, 2학년 국어 경시대회, 2003.)

〈문항 예 2〉 2002년 월드컵 축구 경기는 어느 나라에서 열렸나? (Y초, 2학년 중간고사, 2003.)

〈문항 예 3〉 내가 먹어 본 아이스크림 중에서 가장 맛있었던 것에 대하여 적어 봅시다. (O초, 2학년 중간고사, 2003.)

3.2. 문두의 화계가 불일치한 경우

§ 문두의 화계(語階, speech level)가 일정치 않다.

8) 문항에 대한 기술적 용어와 특징에 대해서는 김창원(1999)을 참조할 수 있다.

(예) 무엇인가?, 무엇일까?, 무엇인가요?, 무엇일까요?, 무엇입니까?

(예) 쓰시오, 써라, 쓰라,

대체로 저학년의 경우 아주높임이나 예사높임체를, 고학년의 경우 예사낮춤이나 아주낮춤체가 사용되고 있다. 그러나 서로 다른 평가지간에는 물론이고, 동일한 평가지 내에서도 화계가 서로 일치하지 않는 예가 많다. 이는 교사들의 분담 제작과 단순 재편집에 기인한 경우가 대부분이다.

3.3. 학습된 지문이나 지문 내용의 기억을 전제로 한 문항

〈문항 예 4〉와 같이 수업을 통해서 학습된 기억을 확인하는 문항들도 빈번하게 나타난다. 또 〈문항 예 5〉에서처럼 교과서 지문의 일부만을 평가 지문으로 출제하는 경향이 많아서, 학생들이 상상한 내용을 쓰기보다는 실제 교과서에서 제시되고 있는 내용을 쓸 가능성이 매우 높다. 교사의 채점 기준 역시 이에 따르기 쉽다.

〈문항 예 4〉 “내가 한 명 더 있었으면”에서 꽃담이는 눈을 감고 상상한 것은 무엇인가요? (C초, 2학년 중간고사, 2003.)

〈문항 예 5〉 선생님께서 들려주시는 이야기를 듣고 앞 부분을 상상하여 써 넣어라. (M초, 4학년 학력평가, 2001.)

위 〈문항 예 5〉의 경우, 교과서에서 학습한 글의 앞부분을 생략하고 출제됨으로써, 학생들의 반응이나 평가 기준이 학습된 내용에 기초할 가능성이 매우 높다.

〈문항 예 6〉 * 다음 글을 읽고, 물음에 답하여라.

“어째서 이 망고 열매를 따 먹으면 안 되는 것입니까, 단장님?”

“이 나무는 마을 가까이 있는 과일 나무일세. 많은 가지가 더위잡기 좋게 늘어져 있고, 탐스런 망고 열매가 주렁주렁 열려 있네. 그런데도 열매 하나 따 먹은 흔적이 없네. 이파리 하나 건드리지 않았어. 이 망고 열매는 마을 사람들이 먹을 수 없는 것 아니겠는가?”

“독약을 넣었다는 말씀이십니까?”

“그렇지. 우리 상인들만 하나씩 따 먹으라고 둔 걸세.”

“그럼, 저 마을은 도둑 마을이겠네요?”

“눈치채지 못하게 조용히 이 곳을 지나가세.”

§ 도둑 마을 사람들은 상인들의 수레를 빼앗기 위해서 어떤 일을 하였는지 써라. (S초 6학년 국어 경시대회, 2003.)

위 문항에서 요구하는 추론의 근거를 지문을 통해서만 확인할 방법이 없다. 이는 출제자의 실수이거나 학습한 지식을 전제로 출제된 문항이라고 볼 수밖에 없다.

3.4. 단순 지식을 물어보는 문항

국어과 교육과정상의 목표와 직접적인 관련이 없는 단순 지식들을 물어보는 문항들이다.

〈문항 예 7〉 이 시에 주로 쓰인 표현 방법은 어느 것입니까? (J초, 5학년 기말고사, 2003)

- ① 직유법 ② 반복법 ③ 의인법 ④ 대구법 ⑤ 은유법

〈문항 예 8〉 다음을 맞으면 ○, 틀리면 ×로 표시하시오.

§ 단오는 우리나라 명절이다. ()

3.5. 불충분한 지문과 정보

가장 빈번하게 나타나는 오류의 하나이다. 제시된 지문의 내용과 정보만으로는 문항에서 요구하는 과제를 해결할 수 없는 경우에 해당한다.

〈문항 예 9〉 밑줄 친 ㉠은 무슨 뜻입니까? (M초, 6학년 학력평가, 2003)

㉠ 방구 아저씨가 떠났습니다. 봄비가 부슬부슬 처량하게 내리던 날이었습니다. 이 날은 방구 아저씨의 귀빠진 날이기도 하였습니다. 그리고 일본이 덜컥 하와이의 진주만을 기습해서 태평양 전쟁을 일으킨 지 일년 너 달하고 스무 하루가 된 날이었습니다.

이 문항 역시 학생들이 이미 수업을 통해서 학습한 것을 전제로 하지 않고서는, 여기에 제시된 지문과 정보만으로는 너무 많이 열려있는 평가 문항이라 할 수 있다. 이 상태에서 문항의 완성도를 갖추려면, 특정의 한 문장만을 대상으로 하는 것이 아니라, 여기에 제시된 텍스트 전체를 대상으로 학습 독자들이 읽기 중 과정에서 추론하면서 제기할 수 있는 질문이나 추론 내용으로 답지를 구성함으로써 선택형으로 처리하는 방식을 들 수 있을 것이다.

3.6. 도드라진 답지

※ 다음 시를 읽고 물음에 답하시오. (O초, 3-2학기말고사, 2003. 자체 제작)

한 학교/ 옆 교실/ 공부시간
 자주빛 새 옷 입은/ 짧은 머리 / 어깨너머 옆얼굴
 창문 사이로 보고 싶어/ 망설이다 망설이다/ 몸을 웅크리며 손을 든다

“선, 생, 남……, 화장실에……”

“요새 자주 아프구나.”

선생님도 모르는 피병

〈문항 예 10〉 이 시에 나오는 인물의 마음은 어떠할까 고르시오.

- ① 외롭다 ② 쓸쓸하다 ③ 기쁘다 ④ 옆 교실 친구를 보고 싶다

위의 두 문항 모두 의도된 된 답지의 길이가 현격하게 길다. 또한 두 문항 사이에 답지의 간섭이 심하다. 이러한 답지 간섭의 문제점은 자주 목격되는 유형의 하나인데, 이는 평가의 정확성을 크게 떨어뜨리는 문제점을 갖는다.

3.7. 복수답, 오답시비

〈문항 예 11〉 다음의 여러 말 중 겹받침의 발음이 다르게 나는 것은 어느 것인가?

- ① 굶다 ② 삶고 ③ 닭이 ④ 맑고 ⑤ 삶다가

위 문항은 답지 ③, ④가 모두 답이 될 수도 있어 복수답 혹은 오답시비와 관련된다.

〈문항 예 12〉 겪은 일에서 글감을 찾으려고 할 때 가장 알맞은 내용은 무엇인가요? (M초, 3학년 학력평가, 2001.)

- ① 운동회 날 부모님과 함께 경기한 것 ② 책에서 읽은 별자리에 대하여
③ 선생님께 들었던 옛날이야기 ④ 텔레비전에서 본 만화영화

〈문항 예 13〉 ‘새와 부리’라는 제목으로 글을 쓴다면 어떤 내용이 좋을까? (J초, 4학년 기말고사, 2003)

- ① 나무에 구멍을 파는 모습 ② 새끼를 돌보는 어미새
③ 알을 품고 있는 모습 ④ 여러 가지 새의 종류

12, 13번 문항은 의도된 답은 분명한 듯하지만, 엄밀하게 따지자면 모두 답이 될 수 있다.

〈문항 예 14〉 다음 중 4학년 학생들의 이야기거리로 적절한 것은 어느 것인가? (경기 S초, 4-1학기말고사, 2003. 자체 제작)

- ① 우리 나라의 국제적 역할
- ② 우리 학급의 어려운 친구를 도와 주는 방법
- ③ 우리 나라의 경제 수준
- ④ 평화 통일을 이한 정부의 역할
- ⑤ 우리 나라 스포츠가 발전하기 위한 방법

이 문항은 출제자의 주관적 판단이 크게 작용함으로 인해 오답시비와 평가의 신뢰도와 관련한 문제점이 제기될 수 있다. 학생 대화 상황을 제시하고, 이 중에서 청자(같은 반 친구들 등)를 고려하고 있는 것을 선택하도록 하는 등의 변형이 요구된다. 또한 답지 배열도 조정이 필요하다.

3.8. 조건화, 구체화의 필요성

〈문항 예 15〉 다음 글을 읽고 떠오르는 생각이나 느낌을 글로 써보시오.
(J초 기말고사, 2003)

풀과 나무가 흙의 품 속에서 생명의 싹을 틔우듯, 벌과 나비가 향기를 맡고 힘을 얻듯, 스님은 아이가 그렇게 자라기를 바랐습니다.

〈문항 예 16〉 다음 글을 읽고 떠오르는 생각이나 느낌을 글로 써보시오.

풀과 나무가 흙의 품 속에서 생명의 싹을 틔우듯,
벌과 나비가 향기를 맡고 힘을 얻듯,
스님은 아이가 그렇게 자라기를 바랐습니다.

이 문항은 학생들의 텍스트 이해와 감상 능력을 평가할 수 있는 좋은 수행평가 문항의 예라고 할 수 있다. 다만 공식적인 평가 장면에서는 특정 부분에 초점을 맞추어 —예컨대, ‘그렇게’에 담긴 스님의 마음을 중심으로— 서술할 수 있도록 어느 정도 제한해 줄 필요가 있다.

〈문항 예 17〉 다음 장면에 나오는 인물이 되어 주고받은 말을 상상하여 써 보시오. (J도, 3학년 수행평가, 2003)

파트라세:

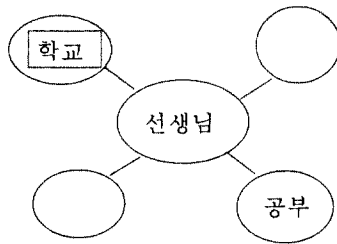
네로:

〈그림〉

이 문항은 유명한 동화 〈플랜더스의 개〉를 바탕으로 하고 있다. 학생들은 이미 수업이나 개인적 독서 체험을 바탕으로, 어두운 환경 속에서도 꿈을 잃지 않고 살아가는 네로와 그의 믿음직스러운 개 파트라세에 대해 잘 알고 있을 것으로 판단할 수 있다. 그런 점에서 이 문항은 한 장면의 삽화를 통해 학생들의 상상력과 표현력을 유도하는 참신한 문항이라고 할 수 있다. 그러나 공식적인 평가 상황에서는 좀더 구체적인 상황 설명이나 조건화를 해 주지 않는 이상, 학생들의 반응이 어떠한지 간에 심지어는 네로와 파트라세의 역할이 바뀌더라도 잘못 되었다고 지적하기 어렵다. 최소한 문두나 보기에서 차리라 작품명을 밝혀 주는 것이 좋을 것이다.

다음 문항 18, 19번 역시 좀더 구체적인 조건을 제시하지 않으면 채점상의 어려움이 따른다. 18번의 경우 생각과 함께 그 이유를 쓰도록 하면 문제점이 어느 정도 해소된다.

〈문항 예 18〉 떠오르는 생각을 정리하여 써 넣어라. (J초 3년 중간고사, 2003)



〈문항 예 19〉 다음 시를 읽고 생각이나 느낌을 자유롭게 표현하여라. (B초, 6학년 기말고사, 2003)

‘-하고, 떠오르는 생각을 써라.’ 형태의 문항들은 학습자들의 국어적 사고 능력을 평가할 수 있는 좋은 수행평가 문항이기는 하지만, 구체적인 평가 기준을 설정해 두지 않는다면 채점 시 어려움에 부딪히게 된다. 수업중이나 학습 과정에서의 개방형 질문은 권장되지만, 공식적 평가의 의미를 갖는 지필평가 형태에서는 문항 제작 시 좀더 구체적인 조건화를 통해 과제 수행과 반응의 방향을 어느 정도 제한해 줄 필요가 있다.

3.9. 차등 배점의 필요성

〈문항 예 20〉 이 글을 읽고 여러분은 어떻게 할지 생각하고 그 이유도 써 보세요. (M초, 3학년 학력평가, 2001.)

- 나의 생각:
- 그 이유:

자신의 생각을 구체적 이유와 함께 서술해야 하는 이러한 쓰기 수행 평가 문항의 경우 다른 문항보다는 배점을 높일 필요가 있다. 그러나 대부분의 학교 자체 평가에서 동일한 배점을 유지하고 있다. 한편 도단위 혹은 전국단위의 성취도 평가에서는 차등 배점제를 잘 활용하고 있다.

3.10. 평가의 텍스트 층위

읽기, 쓰기의 기본 학력 평가가 단어 중심으로 이루어지고 있는 문제점을 지적할 수 있다(도 학력 평가). 단어 차원의 평가는 평가의 용이성 측면에서만뿐만 아니라 교육과정상으로도 필요한 사항이겠으나, 최소한 문장 차원이 중심이 되고, 더 나아가서는 한 단락 정도의 텍스트 차원의 읽기, 쓰기가 반영되어야 할 것이다.

3.11. 평가 기준 상세화 필요성

〈문항 예 21〉 글씨를 바르게 써 봅시다.

사금파리를 깨어 하늘에 흩뿌린 듯한 별들이 싸움을 하자고 대들었다.

대부분의 국어 학력평가, 특히 도(道) 단위 학력평가지의 마지막 문제는, 제시된 문장을 바르게 써 보는 문항으로 이루어진다. 이는 학생들이 바르게 글자를 쓰도록 하는 수행평가 문항으로 의의가 있다고 하겠다. 그런데 채고의 여지가 있는 것은 거의 모든 이원분류표상에서 평가 기준이 제공되지 않는다는 점이다. 교사들의 자율적 판단에 맡길 수밖에 없는 것이다. 어쩌면 이는 가장 정확하고도 권장될 만한 사항이기도 하다. 교사가 직접 학생들의 글자쓰기를 확인하고 평가할 수 있기 때문이다. 다만 문제는 도 단위 평가와 같이 평가 결과와 관련하여 학교 간 비교가 되는 경우 채점상의 신뢰도와 타당도가 떨어질 수밖에 없다는 점이다. 즉, 바르게 쓰기의 정도에 관계없이 옮겨 쓰기만 하면 맞았다고 처리하기 쉬운 것이다. 이는 실제로 현장 교사들과의 인터뷰 과정에서 확인된 것인데, 실제로 거의 대부분의 교사들은 '(학생들이) 쓰기만 하면 모두 맞았다고 처리한다'는 것이다.

따라서 '글씨를 바르게 써 보자.'⁹⁾는 국어과 평가와 관련하여 두 가

9) 이와 유사한 평가 문항으로 '이 시(또는 글)의 주제를 써라', '떠오르는 생각을 써라' 등도 포함될 수 있다. 이러한 평가 문항 역시 학교 현장의 평가지에 자주 등장하는 한 예

지 시사점을 던져준다. 하나는 순수한 국어과 평가적 측면에서 수행평가적 측면에서 평가 기준의 상세화 문제이며, 다른 하나는 평가의 피드백과 활용의 문제이다. 특히 후자의 경우, 교육청이나 학교장 등에 의한 학교 간 우열을 판단하는 비교대상 자료로 활용되어서는 안 될 것이다. 최소한의 교육행정적 근거로 활용된다 하더라도 교사 평가의 기준이 된 다거나 하는 것은 무리라는 점이다. 오히려 지역간 혹은 계층간 성취도 차이를 확인하고 이를 극복, 지원할 수 있도록 하는 포괄적 의미의 국어과 교육과정 계획 수립에 반영될 수 있도록 하는 것이 이러한 교육 평가의 본래적 목표라 하겠다.

3.12. 성취 기준 설정의 문제점

2003년 국가수준 기초학력 진단평가 (읽기, 쓰기)의 경우 평가 문항의 난이도와 관련한 문제점을 지적하지 않을 수 없다. 이 시험은 비록 기초학력을 진단하기 위한 것이라고는 하지만, 학년 간 난이도에 대한 기본적 고려가 미흡하다고 판단된다. 일부 평가 문항의 경우 3학년을 대상으로 한 것이 오히려 6학년에 비해 어렵다고 생각되는 문항들이 보인다.

이러한 예는 도표 이해와 관련된 다음 두 문항의 예에서도 볼 수 있는데, 3학년 수준의 문항의 난이도가 오히려 5학년 수준보다 높다고 하겠다. 위 두 문항은 각각 2003년 10월과 3월에 치러진 것이고, 주관은 각각 교육과정평가원과 전라북도 교육청으로 알려져 있다. 시기나 주체가 다르다고는 하나 이러한 성취 수준과 기준 설정 문제는 매우 큰 차이를 보여준다.

〈문항 예 22〉 비가 올 확률이 가장 높은 지역은 어느 지역입니까? (2003학년도 국가 수준 기초 학력 진단 평가, 3학년 16-17번 문항)

- ① 서울 ② 대전 ③ 광주 ④ 강릉

〈문항 예 23〉 ‘오늘의 날씨’와 ‘비 올 확률’을 보고, 바르게 말한 것은 어느 것입니까?

- ① 서울 지역에서는 등산을 하는 사람들이 많을 거야.
② 부산 지역에서는 외출할 때에 우산이 필요해.
③ 광주 지역에서는 운동회를 하는 학교가 많을 거야.
④ 제주 지역에서는 빨래를 바깥에 말리는 게 좋겠어.

〈문항 예 24〉 다음 표에서 인구 100명당 자전거를 타는 사람 수가 가장 많은 나라는 어느 나라입니까? (2003학년도 국가 수준 기초 학력 진단 평가, 6학년 12번)

이러한 학년 간 수준차 변별의 문제점은 쓰기 영역의 평가 문항에서

도 발견된다.

〈문항 예 25〉 우리 반에서 있었던 일 중에서 기억에 남는 일을 한 가지 골라 글로 쓰시오. (3학년 쓰기 영역 25번)

— <글에 들어갈 내용> —

- 기억에 남는 일을 씁니다(예: 현장 학습 갔던 일, 수업 시간에 있었던 일, 친구가 전학 온 일 등)
- 언제, 어떤 일이 일어났는지 자세하게 씁니다.
- 생각이나 느낌이 잘 드러나게 씁니다.

〈문항 예 26〉 내가 겪은 일 중에서 기억에 남는 일을 한 가지 골라 글로 쓰시오. (6학년 쓰기)

〈글을 쓸 때 주의할 점〉

- 즐거웠던 일, 기뻐했던 일, 화났던 일 중에서 한 가지만 골라 씁니다.
- 언제, 어떤 일이 일어났는지 자세하게 씁니다.
- 생각이나 느낌이 잘 드러나게 씁니다.

우선 5학년과 6학년의 쓰기 수행평가 문항이 동일하게 제시되고 있다는 문제점을 알 수 있다. 더 나아가서 3학년과 5·6학년의 문항 역시 큰 차이가 없다. 얼핏 보기에는 3학년의 과제가 학습자들에게 좀더 친숙한 학급 상황이라는 내용을 제시하고 있는 듯 보이지만, 엄밀히 따져 보면 5·6학년 문항의 '글을 쓸 때 주의할 점'에서 제시되고 있는 주제에 비해, 오히려 주제 한정적이라는 어려움을 가중시키고 있는 면도 부인할 수 없다.

물론 동일한 주제의 쓰기 과제를 서로 다른 학년에 적용해서는 안 된다고 단언하기는 어려운 부분도 있다. 피시험자들의 반응의 수준에 따라서 평가의 결과는 다르게 측정되고 해석될 수도 있기 때문이다. 그러나 이들에 대한 '채점 기준'이나 '채점상의 유의점'이 크게 다르지 않다

는 데에 또 다른 문제점이 있다.

여기에서 우리는 쓰기 능력 평가의 수준 문제와 관련하여 좀더 세밀한 고려가 필요하다는 점을 알 수 있다. 예컨대, 예상되는 텍스트의 주제와 내용, 과제 수행 과정에서 요구되는 어휘력, 주어진 제한된 시간 내의 수행에 따른 분량¹⁰⁾의 문제 등을 좀더 정밀하게 고려하여야 할 것이다. 궁극적으로 이러한 문제점은 우리의 국어과 교육과정 '쓰기' 영역에서 텍스트의 수준과 특성들에 대한 다양성을 반영하고 있지 못하다는 점과 관련된다.

요컨대 평가 문항 제작 시에는 평가 문항의 타당성, 객관성, 신뢰성을 충분히 고려해야 하는데, 이를 위해 평가 목표와 학습 목표의 관련성, 구체적 과제와 조건의 제시, 교과서 지문 재활용 문제점, 문두·답지 등과 관련한 정확하고 간결한 표현, 구체적인 평가 기준, 학년 간 성취수준 차이 등에 유의해야 한다.

4. 초등학교 국어과 평가체제의 문제점과 개선 방향

국어교육 메타 평가의 논의는 국어교육 '현장'에서 출발하지 않을 수 없다. 기본적으로 초등학교 국어 시험의 문제점들을 개선하기 위해서는 현장 평가 실태, 평가 문항 오류 등에 대한 분석을 중심으로 하는 메타 평가의 시각에서 접근할 필요가 있다. 앞에서 살펴본 국어 시험과 관련된 여러 가지 문제점들을 극복하고 개선하기 위해서는 근본적으로 국어 교육 평가의 시스템을 개선해야 한다. 이는 궁극적으로 국어교육 환경

10) 실제로 채점 결과의 몇몇 유형에서 나타나는 문제점으로, 분량에 대한 규정이 정해져 있지 않기 때문에, 상세화의 정도 문제는 고려될 여지가 별로 없다는 것을 지적할 수 있다. 특히 국가수준의 성취도 평가는 해당 학교 차원에서는 부진아 교육을 책임져야 하는 부담이 따르기 때문에, 어떤 식으로든 학생들이 어느 정도 적절한 반응만 하면 높은 점수를 줄 가능성이 높다. 따라서 일관된 평가의 문제, 즉 평가의 신뢰도 문제도 주요한 고려 사항이 될 수 있다.

과 내용의 개선을 의미한다.

평가 시스템의 개선은 국어교육평가학과 관련되는 학문적 연구, 교육 주체(연구자, 행정가, 교사, 학생, 학부모) 등의 인식 전환, 국어교육 평가 자료 제공이나 교사 연수 등 국어교육 평가 활동을 위한 행·재정적 서비스 제공, 평가결과 송환방식(생활통지표 등)의 개선과 교사의 수업 부담 경감 등 제도적 개선과 지원 차원에서 접근할 필요가 있다. 이러한 맥락에서 좀더 구체적인 개선 방안들을 제안하면 다음과 같다.

4.1. 메타평가 체계의 확립

국어과 평가에 대한 지속적이고도 체계적인 메타평가 활동이 이루어져야 하며, 이를 위해서는 좀더 구체적인 메타평가 모형의 수립이 요구된다. 더 나아가서는 국어교육평가학의 체계화가 구체적으로 이루어져야 할 것이다.

4.2. 국어교육 평가 연수의 체계화와 지속적 운영

평가 문항이나 평가 기준을 교사들이 스스로 구안하고 제작할 수 있도록 국어교육평가 연수 프로그램이 체계적으로 마련되어야 할 것이다. 또 교사들이 직접 문항 제작을 하는 경우보다는 다양한 자료와 문항들을 재활용하거나 변형 적용하는 경우도 많기 때문에, 문항 개발 이외에도 양질의 문항에 대한 감식안을 길러줄 필요가 있다. 아울러 각 학교에는 대체로 평가 담당 교사가 지정되어 있는데, 이들에게 기본적으로 국어과평가 연수를 받을 수 있도록 지원해 주어야 할 것이다. 모든 교과들에서 언어를 통해 평가 문항이 제작되기 때문에, 평가 담당 교사들의 국어과 평가 연수 기회는 큰 도움이 될 수 있다.

4.3. 문제은행 설립과 평가 자료 제공

초등 국어과 평가 문항 개발의 실질적 주체는 초등교사들이지만, 모든 학교에서 모든 교사들이 완성도가 높은 문항을 개발해 낸다는 것은 현실적 교육 여건에 비추어 볼 때 쉬운 일이 아니다. 따라서 초등 국어과 문제 은행의 설립과 문항 검토 시스템, 다양한 평가 자료의 제공이 이루어져야 할 것이다. 이를 위해서는 국어과 교육 연구자들도 일정 부분 적극적으로 보조할 필요가 있다.

4.4. 국어과 수행평가의 재개념화와 인식 전환

국어과 수행평가는 교과 일반의 수행평가와는 다른 고유한 특성과 방법들을 지니고 있다. 따라서 국어과 수행평가에 대한 개념과 인식 차원에서 어느 정도의 합의(컨센서스)가 필요하며, 교육관계자들의 인식 전환도 요구된다. 수행평가의 개념이 일반화 되면서, 개념과 현상적 포섭성도 비대해짐으로써, 기존의 지필평가의 개념과 구별이 안 될 정도로, 수행평가의 본질이 크게 훼손된 감도 없지 않다. 이와 관련하여 다음과 같은 초등학교 현장 교사의 고백에 유의할 필요가 있다.

“교육 현장에서 국어과 수행 평가는 제도적으로는 그 틀을 잡아가고 있지만, 수행평가의 시행상황은 아직도 목표에서 시행절차와 방법, 그리고 문항개발 등에 있어서 많은 혼란을 겪고 있다. 대부분의 교사들은 국어과 수행평가를 성적표 작성의 근거자료 확보의 수단으로 인식하고 있기 때문에 국어과 수행평가가 통합적이고 총체적이며, 지속적인 평가가 되지 못하고 단편적이고 일회적인 평가에 머물고 있다. 평가 문항을 제작함에 있어서는 지도서나 상업용 교육 자료를 가장 많이 참고하고, 이미 만들어진 문항을 발췌하거나 수정하여 사용한다. 실제 평가는 평가 계획 중 일부를 투입하거나, 평가계획과는 별도의 문항을 적용하는 경우가 많다. 다양한 평가 방법이 있으나 현장에서 실시되는 국어과 수행평가는 정답을 답지 중에서 선택하게 하는 선택형 문항에서 답을 직접 작성하게 하는 단답형 문항으로 변화되었을 뿐, 평가 방법이

다양하지 않다. 그리고 교과서의 문제 상황과 지문을 반복하여 제시함으로써 학습한 내용을 반복하여 평가한다. 이는 실제 수행하는 활용능력을 측정하고자 하는 수행평가의 근본취지와 거리가 먼 것이라 할 수 있다. (……) 학기 중에는 평가를 방치하고 있다가 학기말에 무리하게 평가를 실시한다거나, 그 동안 학생의 국어과 수행 능력에 대한 결과 정리와 기술을 학기말에 집중적으로 실시하기 때문에 업무의 과정과 평가의 비현실성을 토로하게 된다.”¹¹⁾

요컨대 수행평가 운영의 개선을 위해서는 교사들의 국어과 수행평가에 대한 재인식이 요구되며, 교사들간에 실제로 문항을 제작하고 토론하는 적극적 의미의 활동적 수행평가 연수가 요구된다.

4.5. 국어과 준교과전담제 필요

국어과 평가체계의 확립과 상세화는 곧 현장에서의 국어교육 체계화와 내실화를 의미한다. 이를 위해서는 무엇보다도 교사의 수업 부담과 행정업무 부담을 경감시켜줄 필요가 있다. 예컨대 후술하는 텍사스 주 락프렐리 (공리)초등학교의 예에서처럼 매칭(matching) 수업을 통한 준교과전담제, 학년전담제의 활성화를 적극적으로 도입하고, 장기적으로는 학급당 학생수의 감축 등 재정적 지원을 늘려 나가야 할 것이다. 국어 시험, 평가 체제의 개선은 결코 문항 개발 수준에서 완수될 수 있는 성질이 아니며, 궁극적으로 평가 시스템 더 나아가서는 교육시스템의 개선을 필요로 한다.

4.6. 평가결과 송환체계 개선

평가결과 송환 방법의 대표적 양식인 초등학교 생활통지표와 관련하여 교과 발달 상황의 획일화는 여러 가지 문제점을 안고 있다. 각 교과별 특성이 드러날 수 있게 구성되어야 하며, 특히 국어과의 특성이 드러

11) 전주교육대학교 대학원 김미연 교사의 보고서에서 인용.

날 수 있도록 재구성할 필요가 있다. 현행처럼 모든 교과 통일 체계는 불합리하다. 기술 내용에 있어서도 최소한 국어 영역별 혹은 활동별로 학습자의 발달 과정이나 장단점을 고려 기술할 수 있어야 한다. 학습자가 부진한 면이 있다면 무엇인지 알려주어야 하고, 학부모도 이러한 내용을 알 권리와 의무가 있다.

평가 결과 송환 양식의 대표적인 장치인 생활통지표 기록 방식이 대폭 개선될 필요가 있다. 가능한 한 상세하게 자주 학생과 학부모에게 피드백해 주는 것이 좋지만, 우리 교육 여건상 교사들의 업무 가중 등 어려움이 많기 때문에 현행 기술 방식을 간소화하면서 체크리스트 방식을 가미하거나, 좀더 간편한 전산 처리 방법들을 지속적으로 개발할 필요가 있다. 참고로 우리나라와 외국(미국 텍사스 주 락프렐리 초등학교, 호주 켄싱턴 초등학교)의 통지표 기술 양식의 의 통지표 예를 간단히 비교대조하여 살펴보면 다음과 같다.

먼저 우리나라에서 최근에 새롭게 개발한 국가교육정보시스템(NEIS)에 따른 평가결과 기록과 송환방식은 다음의 자료 1, 2와 같다. 자료 2는 각 가정으로 송환된 결과물이다.

자료 1 NEIS의 성적표 기록 방식

자료 2 NEIS에 따른 생활통지표(예)

NEIS는 비록 우리의 선진적인 ICT를 기반으로 체계적으로 만들어졌는지는 몰라도, 최소한 국어과 교과발달 상황의 기록 양식을 보면 전혀 만족스럽지 못하다. 수업 부담 등으로 격무에 시달리고 있는 교사들 역시 상당수가 달가워하지 않는 눈치다. 가정으로 1년에 한두 번 배달

되고 마는 이른바 생활통지표상의 학생 성취 결과에 대한 정보는 '잘 했다'는 한 마디 이상의 의미를 찾기가 쉽지 않다. 물론 교사에 따라서 비교적 상세한 정보를 제공해 주는 경우도 있겠지만, 이는 흔치 않은 일이며 잘해야 1년에 한두 번이다. 앞의 설문 조사 분석 결과에서도 나왔지만, 최소한 학습자의 강약점과 성취 도달의 경과 등에 대한 정보는 가능한 한 자주, 상세하게 이루어질 필요가 있다. 이를 이는 것은 학생과 학부모의 권리이자 의무이기도 하다. 요컨대 장점 중심의 기술식 피드백은 성취 결과에 대한 무관심 이상의 의미를 주기 어려운 경우가 많다.

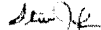
교사의 업무를 경감시키고, 학습자에 대한 학업 성취 도달 과정에 대한 정보를 상세화 해 주는 방법으로 텍스트 랙프레리 초등학교의 예¹²⁾처럼, 체크리스트 형식과 평가 결과 고지 형태에 교사와 학부모의 소통란을 결들인 다소 전통적인 양식이 차라리 도움이 될 수도 있다. 호주의 켄싱턴 초등학교 같은 경우에는 학생의 자기 평가란도 마련되어 있어 눈길을 끈다.

자료 3 텍사스 랙프레리 초등학교 통지표(예1-1)

COLLEGE STATION INDEPENDENT
SCHOOL DISTRICT

ELEMENTARY
PROGRESS REPORT

Report cards are a way of reporting the progress your child is making in school. In order to serve the total needs of our students, we encourage parents to have periodic conferences with teachers and counselors to discuss student progress. Conferences may be scheduled by calling the school where your child attends. We appreciate having you child in the College Station Independent School District.



12) 자료를 제공해 주신 전주교대 이경한 교수님께 감사 드린다.

160 국어교육학연구 제20집 (2004. 8)

자료 4 텍사스 락프렐리 초등학교 통지표(예1-2)

자료 5 텍사스 락프렐리 초등학교 월통지표(예1-3)

5. 맺음말

국어교육 평가의 의미와 공간은 매우 광대하나, 그 출발점은 '국어 시험'에서 시작하지 않을 수 없다. 이러한 관점에서 이 논문은 초등학교 국어 시험과 관련한 현황과 문제점을 설문조사, 면담, 시험지(평가 문항)을 중심으로 분석하고, 그 개선 방안을 제시해 보고자 하였다.

설문조사의 주요 대상은 초등학교 교사(총 189명), 초등학생(총 413명), 초등학생 학부모(317명)의 세 집단을 중심으로 총 900여 명에 이른다. 시험지 분석은 서울(1), 경기(2), 전북(9) 지역의 총 12개교를 대상으로 하여 각 학교에서 2001-2003년 기간에 치러진 각종 학력평가, 월말고사, 중간고사, 기말고사, 성취도 평가, 수행평가 문항들을 중심으로 하였다.

설문조사 결과 초등학교 현장에서는 중간고사, 기말고사, 시도학력평가, 월말고사, 학력평가 등 다양한 지필평가가 이루어지고 있었으나, 소규모 학교일수록 학교장의 재량에 따라 학생들이 시험을 많이 치르지 않았다. 평가 문항 구안은 동료교사와 분담제작하거나 인터넷 자료 활용, 시도학력평가로 대체 등이 많았다. 자주 사용하는 국어과 수행평가 유형으로는 서술형·논술형 평가, 실기평가, 관찰평가, 누가기록철 등을 활용하는 방법이었다. 국어과 수행평가는 대체로 매달 1-2회 정도로 실시하고 있는 것으로 나타났으며, 수행평가지 어려운 점으로는 평가기준 설정과 문항 제작을 꼽았다. 국어과 평가 결과 기록은 학습자가 뛰어난 성취를 보인 특정 영역이나 기능을 중심으로 서술한다는 의견이 대다수였으며, 개선책으로는 학습자가 미진한 부분도 같이 기술할 필요가 있고, 체크리스트 방식이나 기록 관리에 간편한 프로그램, 문제은행 개발 등을 들었다.

초등학생들은 학부모에 비해 생활통지표에 대한 관심이 현저하게 낮게 나타났으며, 학기말고사의 부담이 가장 크다고 반응했다. 학생과 학부모 모두 현행 생활통지표의 정보 상세화 정도는 대체로 보통 수준이

라는 반응을 보였다. 학생들은 국어 시험 빈도가 보통이거나 자주 보지 않는다는 편이라고 반응했다. 학부모들은 자신의 자녀의 국어능력이 대체로 보통이거나 우수하다고 생각하였고, 그 판단의 근거는 평소의 관찰결과라는 반응이 학교시험이나 생활통지표보다 훨씬 높았다.

초등학교에서 치러진 국어과 평가문항들은 전반적으로 양호한 편이었으나, 평가목표의 타당성, 지시문, 문두, 불충분한 지문과 정보, 답지 구성의 문제점, 오답시비의 문제점, 조건화와 평가 기준의 미비, 성취 기준 설정의 문제점 등 여러 가지 문제점을 보여 주는 사례가 상당수 발견되었다. 이러한 문제점들은 평가 문항의 타당도, 신뢰도, 객관도를 떨어뜨리는 직간접적 원인이 된다.

국어시험 개선은 평가 시스템의 개선에서 접근해야 한다. 이를 위해서는 국어교육평가학과 관련되는 학문적 연구, 교육 주체(연구자, 행정가, 교사, 학생, 학부모) 등의 인식 전환, 국어교육 평가 자료 제공이나 교사 연수 등 국어교육 평가 활동을 위한 행·재정적 서비스 제공, 평가결과 송환방식(생활통지표 등)의 개선과 교사의 수업 부담 경감 등 제도적 개선과 지원 차원에서 접근할 필요가 있다. 생활통지표 기술방식과 관련하여서는 텍사스 락프레리 초등학교의 예를 우리의 NEIS 체제와 비교 대조하여 보았다. 평가결과 소환의 목적이 동기유발이나 사기 진작뿐만 아니라, 학습자의 정확한 발달 상태에 대한 정보를 공유하고 개선해 나가는 기초 자료가 된다는 점에서 현행 통지표 기술 방식은 문제점이 있음을 지적하고 그 개선 방안을 제시해 보았다. 국어과 메타평가와 관련 좀더 정밀하고도 구체적인 연구 성과들이 뒤따라야 할 것으로 생각된다.*

* 본 논문은 2004. 6. 21. 투고되었으며, 2004. 7. 12. 심사가 시작되어 2004. 7. 26. 심사가 종료되었음.

■ 참고문헌

- 김동일(1999), "국어교육 평가에 대한 메타 인식", 『국어교육 평가의 일반 원리』, 99학술대회 자료집.
- 김미숙(1998), "국어교육에서의 평가", 『국어교육연구』 제5집, 서울대 국어교육연구소.
- 김재춘, 조덕주, 노국향(1998), 『초·중등 교육과정 평가 방안 -제7차 교육과정 평가 실행에 대비한 기초 연구』, 연구보고 RRC 98-2, 한국교육과정평가원.
- 김창원(1999), "국어교육 평가의 구조와 원리", 『국어교육 평가의 일반 원리』, 99학술대회 자료집.
- 노명완(1990), "국어 교육에서의 평가", 『국어교육 개선 방안 연구』, 서울대학교 국어교육과.
- 박갑수 외(2000), 『국어 표현·이해 교육론』, 집문당.
- 박영목(1993), 『교육 과정과 국어 교육 평가』, 서울대학교 국어교육연구소 자료집.
- 박인기 외(1999), 『국어와 수행평가』, 삼지원.
- 박인기(1997), "국어과 교육 평가의 개선과 교육의 질적 도약", 『21세기 국어과 교육의 지향과 수준별 교육과정』, 한국교육개발원.
- 서혁(1999), "듣기 능력 평가의 개선 방안 연구", 『선청어문』 제27집, 서울대 국어교육과.
- 서혁(2000), "국어 이해 능력 평가 방안", 『국어표현이해론』, 집문당.
- 손영애 외(1992), 『교육의 본질 추구를 위한 국어 교육 평가 체계 연구3: 예시 평가 도구 개발 및 현장 적용』, 한국교육개발원.
- 이도영(1999), "국어과 교육 평가 연구의 문제점과 개선 방안", 『국어교육학회 발표대회 자료집』, 1999.
- 이삼형(1999), "국어교육 평가의 관점", 『국어교육 평가의 일반 원리』, 99학술대회 자료집.
- 임천택(2002), 『학습자 중심의 국어과 평가』, 박이정.
- 진영란(2000), 『국어과 수행평가의 문제점과 개선 방안』, 전주교육대학교 석사논문.
- 천경록(1999), "국어과 학습평가의 원리와 방법", 학술발표대회자료집, 한국국어교육연구회.
- 최미숙(1999), "국어교육 평가의 원리와 실제", 『국어교육 평가의 일반 원리』, 99학술대회 자료집.
- 최미숙·양정실(1998), 『국가 교육과정에 근거한 평가 기준 및 도구 개발 연구-고등학교 국어』, 연구보고 PRE98-3-3, 한국교육과정평가원.
- 최영환 외 (2001), 『국어과 수행중심 평가』, 학문출판(주).
- 한국어문교육학회(2003), "7차 교육과정과 국어과 평가", 한국어문교육학회 전국학술발표대회.
- 한철우·이인제(1990), 『교육의 본질 추구를 위한 국어교육 평가 체계 연구 (1)』, 한국교육개발원.
- 황정규(1998), 『학교학습과 교육평가』, 서울: 교육과학사.

- Bachman L. F.(1989). Fundamental Considerations in Language Testing, London: Oxford University Press.
- Bloom, B. S.(1970), Toward a theory of testing which include measurement, evaluation, and assessment, in M. C. Wittrock & D. E. Willey(Eds), The Evaluation of Instruction, New York: Holt, Reinhart Winston.
- Carroll, J. B.(1968), 'The psychology of language testing' in A. Davis(ed.): Language Testing Symposium: A Psycholinguistic Approach, London: Oxford University Press.

〈초록〉

초등학교 국어과 시험의 문제점과 개선 방안

서 혁

이 논문은 초등학교 국어 시험과 관련한 현황과 문제점을 설문조사, 면담, 시험지(평가 문항)를 중심으로 분석하고, 그 개선 방안을 제시해 보고자 하였다.

설문조사 결과 초등학교 현장에서는 중간고사, 기말고사, 시도학력평가, 월말고사, 학력평가 등 다양한 지필평가가 이루어지고 있었으나, 소규모 학교일수록 학교장의 재량에 따라 학생들이 시험을 많이 치르지 않았다. 국어과 수행평가는 대체로 매달 1-2회 정도로 실시하고 있는 것으로 나타났으며, 수행평가 시 어려운 점으로는 평가 기준 설정과 문항 제작을 꼽았다.

초등학교에서 치러진 국어과 평가문항들은 전반적으로 양호한 편이었으나, 평가 목표의 타당성, 지시문, 문두, 불충분한 지문과 정보, 답지 구성의 문제점, 오답시비의 문제점, 조건화와 평가 기준의 미비, 성취 기준 설정의 문제점 등 여러 가지 문제점을 보여 주는 사례가 상당수 발견되었다. 이러한 문제점들은 평가 문항의 타당도, 신뢰도, 객관도를 떨어뜨리는 직간접적 원인이 된다.

국어 시험 개선은 평가 시스템의 개선에서 접근해야 한다. 이를 위해서는 국어교육평가학과 관련되는 학문적 연구, 교육 주체(연구자, 행정가, 교사, 학생, 학부모) 등의 인식 전환, 국어교육 평가 자료 제공이나 교사 연수 등 국어교육 평가 활동을 위한 행·재정적 서비스 제공, 평가결과 송환방식(생활통지표 등)의 개선과 교사의 수업 부담 경감 등 제도적 개선과 지원 차원에서 접근할 필요가 있다.

【핵심어】 국어교육, 국어 시험, 국어과 평가, 수행평가, 평가문항, 평가 기준, 성취 기준

〈Abstract〉

Item Analysis and Suggestions
to Improve an Evaluation System in KLE

Suh, Hyuk

A research on evaluation system in Korean language education (hereafter KLE) are increasing recently, but we could find out only a few thesis regarding items and tools analysis. In this research, I presented several suggestions to improve some problems concerning evaluation items and tools made for the KLE in primary school, which based on surveys, interviews and examinations. All 900 people were included for the survey including teachers, student, parents and some researchers. Many kinds of evaluations are holding in the schools such as mid-term exam, final exam, monthly exam, year-end exam, quiz test and diagnostic or achievement exam of national and provincial level. There were specific contrastive features in numbers of exam-times as the scale of schools, i.e. the school which has more than 40 classes took exams much more frequently than the smaller schools. The performance-oriented evaluations were held generally 1-2 time every month. The most difficulties teachers felt is making items and criteria for evaluations.

Generally speaking, the evaluation items were successful, but there were still some problems such as validity, reliability, question sentences, insufficient text and information to solve the item and standards of scoring, etc.. To improve the problems concerning evaluations in KLE, we need to approach to evaluation system.

【key words】 Korean language education, evaluation in KLE, item analysis, performance-oriented evaluation, criteria of evaluation.