

우리나라 수행평가 정책의 특징과 발전 방향 연구*

—국어과 수행평가의 실행 양상의 원인에 대한 정책적 반추를 중심으로—

신명선**

< 차례 >

1. 서론을 대신하여 : 국어과 수행평가의 실행 양상
2. 수행평가 정책의 통시적 고찰
3. 수행평가 정책의 공시적 고찰
4. 수행평가 정책의 개발 논리에 대한 재고
5. 결론

1. 서론 : 국어과 수행평가의 실행 양상

그 동안 우리나라에서 수행평가에 관한 교육적 연구는 매우 풍성하게 이루어져 왔다. 1990년대 후반부터 현재까지 수행평가를 주제로 한 각종 학위 논문과 소논문, 보고서 등은 거의 5,000여 편에 이를 것으로 추정되는데, 단일 주제에 관해 이렇게 많은 학문적 관심이 주어진 예도 흔치 않을 것이다. 국어과에서도 수행평가에 관한 여러 연구가 있었다. 국어과에서 수행평가에 관한 연구는 주로 기존의 선다형 평가가 갖는 문제점을 해결할 수 있는 교육적 대안으로서 논의되었다. 특히 초기 연구에서는 현장

* 이 논문은 한국교육과정평가원에서 이루어진 ‘학교 교육 내실화를 위한 수행평가 개선 연구’를 수행하면서 이루어진 것으로, 서지영 · 신명선 · 김소영(2008)의 논의와 일정 부분 문제 의식을 공유하고 있음을 밝힌다.

** 인하대

의 잘못된 평가 관행을 변화시킬 수 있는 이론적 대안으로서의 수행평가에 대한 학문적 기대와 관심이 논문을 통해 집중적으로 펼쳐지는 경향이 있었다(천정록, 2001 ; 박인기, 2000 ; 최미숙, 1999 등). 한편 수행평가의 시행 실태를 살펴보는 연구 경향 또한 교육대학원 석사 논문¹⁾뿐만 아니라 개별 연구자들의 소논문 등을 통해 초기부터 지금까지 꾸준히 이루어져 왔는데 비교적 최근에 이루어진 것만 해도 김혜정(2008), 이수진(2008), 김진곤·김혜영(2005) 등을 들 수 있다. 최근 김혜정(2008 : 99)에서는 ‘수행평가’와 ‘국어과 평가의 문제점’이라는 두 가지 주제에 대해서는 이미 많은 연구가 수행되어 와서, 원론적 수준이든 아니든 문제점에서 개선 방안으로 나가는 논문의 구성 방식은 더 이상 의미가 없을 것이라는 진단을 내리기도 하였다.

수행평가가 전국의 학교 현장으로 도입(1999년)된 지 이제 꼭 10년이 되었다. 그간 수행평가의 실시와 더불어 교사의 평가에 대한 인식이 일정 부분 변화된 것도 사실이다. 특히 다양한 평가를 시행할 필요가 있다는 것에 대한 폭넓은 동의가 교사들 사이에 이루어진 점은 높이 평가할 만하다.²⁾ 그러나 지금 단위 학교의 평가가 도입 당시의 기대를 충족시키지 못하고 있음 또한 분명하다. 수행평가의 시행 실태에 대한 그간의 논의를 정리해 보면 학교 현장에서 수행평가는 다음과 같은 방식으로 이루어지고 있다.

<국어과 수행평가의 시행 실태>

1. 평가 방법

- 작년도 수행평가 실행 계획서의 무비판적 수용에 따른 평가 방법의 혼란³⁾

1) 김슬아(2007), 구민경(2006), 이희숙(2004) 등.

2) 이에 대해서는 다음과 같은 지적도 이루어진 바 있다. “수행평가의 관심으로부터 시작된 평가의 변화는 교수·학습의 내용에 따라 서술형 평가, 토론형 평가, 보고서 평가, 포트폴리오 평가 등 다양한 평가 방식을 활용하는 것으로 나타나고 있다. 이제 국어과 교육의 성격에 맞는 평가를 해야 한다는 것은 광범위한 이해를 얻었다고 볼 수 있을 것이다.”(최미숙, 2006 : 83~84)

3) 학교에서는 학년 초, 각 학년별로 교육과정에 맞는 수행평가 계획과 문항을 제작하여

- 교과서(예컨대 중학교 ‘생활 국어’) 검사
- 국어 수업 시간에 떠들거나 졸지 않고 수업에 열심히 참여했는지, 그리고 수업 준비물을 잊지 않고 챙겨 왔는지 등의 검사
- 국어 수업 시간에 교사의 설명 내용을 필기한 노트 검사
- 기존의 선다형 문항을 제작한 후 쪽지 시험 형태로 시험을 보고 그 결과를 수행평가 점수로 부여하는 방식(타학교와의 비교를 우려해 문항 출제의 난이도를 의도적으로 낮게 해 좋은 점수를 부여하는 풍조 만연)

2. 평가 시기

- 특정 시기에 수행평가를 집중 실시해 학생과 교사에게 과중한 부담을 준.
- 수행평가 과제(예 : 한 학기에 10편의 책을 읽고 독후감 쓰기 등)를 부과한 후 학기말에 채점. 이 경우 학부모가 대신 해 주거나 학원의 도움을 받는 경우가 존재함.

3. 결과 처리 방식

- 학생 간 점수 편차를 최소화함. 심지어 수행평가 과제를 내지 않거나 그 수업에 불참하더라도 기본 점수⁴⁾를 부여해 최고점과 큰 차이가 안 나게 함.
- 학급 간 수행평가 점수 차이가 날 경우 교사들끼리 의논하여 학급 간 점수 차이가 거의 나지 않게 조정함.
- 초등의 경우 수행평가 결과를 통지표에 기록할 때 긍정적 진술로만 일관함. 교사의 주관적 기술이나 학생의 학업 부진 상황은 가능하면 기록하지 않음.
- 모둠별 평가가 이루어질 경우 모둠 활동에 열심히 참여한 학생과 그렇지 않은 학생이 모두 동일한 점수를 받으므로 학생과 학부모의 불만이 있음.

수행평가야말로 학습자의 국어 능력(고등 사고 능력)을 제대로 측정할 수 있는 대안적 평가가 될 수 있을 것이라는 학문적 기대와는 달리 실제

제출하고 있다. 교사들은 시간이 부족하여 기존의 자료들을 별 비판없이 그대로 제출하고 있는 실정이다(김대조, 2004 : 42). 그러다보면 학기 중에 계획서와 다른 수행평가를 실시하는 일도 발생하게 된다.

4) 여러 사정으로 수행평가에 참여하지 못한 학생들에게 주는 점수.

교육 현장에서 수행평가는 단순히 ‘중간·기말 고사가 아닌 기타 다른 종류의 평가’ 정도에 불과하다. 국어 교과서나 노트를 걷어서 점수를 매기고 수업 시간에 조용히 했는지 여부를 체크하는 것이 진정한 의미에서의 수행평가가 될 수 없다는 데에는 많은 이들이 동의할 것이다. 그런데 현장 교사들은 수행평가 때문에 업무가 늘어났다는 불평을 하면서도 수행평가 실시를 적극 찬성하고 있는데 그 이유 중의 하나는 수행평가가 수업 분위기를 좋게 하는 데 기여하기 때문이다.⁵⁾

그렇다면 왜 이런 결과가 나타나게 됐을까. 이수진(2008 : 33~36)에서는 이에 대해 ‘이는 체제의 문제가 아니라 인식의 문제가 아닐까 생각된다. 즉 수행평가 자체가 지닌 문제라기보다는 그 실행 과정에서 문제가 발생하기 때문일 것’이라고 진단한 바 있다. 수행평가가 아무리 교육적으로 타당한 개념이라고 해도 학교 현장에서 제대로 구현될 수 없다면 사실상 무의미할 수도 있다. 따라서 학교 현장의 여건이나 교사의 인식 등의 변수를 고려하여 수행평가 실행 과정의 문제를 꼼꼼히 분석할 필요가 있을 것이다. 그러나 그간 이에 대한 구체적인 연구는 이루어지지 못했다.

본고에서는 국어과 수행평가의 실행 양상에서 드러난 위와 같은 문제점에 대한 원인을 정책을 중심으로 검토해 보고자 한다. 수행평가 정책 도입의 초기 과정에 대한 통시적 고찰과 더불어 현재의 수행평가 정책의 특징을 국가 단위, 도 단위, 학교 단위로 나누어 살펴보고자 한다. 검토 자료는 정책의 시행을 담고 있는 각종 문서를 대상으로 하도록 한다. 정책 분석의 결과는 올바른 국어과 수행 평가 실행의 방향을 탐색하는 데 도움을 줄 것이라 기대된다.

5) 실제로 필자가 만난 10여 명의 교사들도 모두 이구동성으로 ‘수행평가가 없으면 수업 시간에 떠들거나 딴 짓하는 아이들을 통제할 방법이 없다’고 말했다. 어떤 교사는 지금 교사들에게 수행평가는 일종의 ‘무기’라고 말했다.

2. 수행평가 정책의 통시적 고찰

우리나라에서 수행평가는 제7차 교육과정의 연차별 적용 첫 해인 1999년부터 전국적으로 실시되었다. 수행평가의 전국적 실시를 앞두고 1998년 10월 제시된 ‘2002학년도 대학입시제도 개선안’이나 ‘교육비전 2002: 새 학교문화 창조’의 어디에도 수행평가의 개념이 명료하게 제시되지 않았다. 비교적 자세하게 언급된 후자, 즉 ‘교육비전 2002: 새 학교문화 창조’에는 ‘다양한 평가 기법의 도입, 학습 과정과 결과의 능가적 기록, 수행평가의 기준 마련, 학생 상호평가 방법 도입, 집단의 수행 결과 포함’ 등과 같은 지침만 나열되어 있다. 학교 현장의 교사들 입장에서는 해야 할 일은 있으나 그것이 무엇인지 어떻게 해야 하는지 등에 대해서는 알 수 없는 상황이었다고 할 수 있다. 그런데 각 시·도교육청은 1999년도 6월과 7월에 걸쳐 교육부로부터 시·도교육청 평가를 받게 되어 있었고, 평가의 항목 속에 수행평가의 실시 정도가 포함되어 있었다. 따라서 각 시·도교육청은 1999년도 학년 초에 관내의 모든 일선 학교에 수행평가 실시를 강력하게 요구하였다.⁶⁾

당시 학교 교사들은 수행평가를 지나치게 ‘이상적’(허경철 외, 1999: 6)으로 생각하고 있었기 때문에 학교 여건상 수행평가 실시는 불가능하다고 생각하고 있었다. 물론 많은 교사들은 수행평가가 무엇인지 그 개념을 명료하게 이해하지 못한 상태였다. 이러한 상황을 인식하고 교육부는 수행평가 홍보를 위한 다양한 노력을 펼쳤다.⁷⁾ 그러나 그것들은 실제적으로

6) 초등학교의 경우는 50점 만점 중 “평가의 다양화를 위한 노력”이라는 기준 이름으로 4점이 부여되어 있고, 중학교의 경우 역시 50점 만점에서 “평가 방법의 다양화, 전문화를 위한 노력”이라는 기준에 4점이 배당되었다. 그리고 고등학교는 60점 만점에서 중학교와 같은 기준 하에 4점이 배당되어 총 12점이 배당되었다. 그러므로 각 시·도교육청 입장에서 보면 수행평가 점수는 전체 600점 중 12점이 되어 전체 점수의 2%가 된다고 한다.

7) 그 구체적인 내역은 다음에 제시되는 <표 1> 참조. 그 외 기타 연수 내용은 다음과 같다. : “새 학교문화창조 핵심교원 연수”(교원대, 801담당 4,500명), “평가전문과정 연수”(국가전문행정연수원, 177명), “목표지향평가 핵심교원 연수”(교사, 장학사 830명)

학교 현장에 큰 영향을 미치지 못하였는데, 단순한 자료의 성격을 띠었기 때문이다. 그 홍보 자료 내용이 학교 현장에 직접적인 영향을 끼친 것은 1999년 4월 29일 공포된 <훈령 587호(학교생활기록부 전산 처리 및 관리 지침)>⁸⁾이다. 그런데 <훈령 587호>는 ‘수행평가는 지나치게 이상적이어서 학교 현장에서 실행하기 어렵다.’는 현장의 목소리를 의식했기 때문인지 수행평가의 개념을 매우 폭 넓게 설정하였다. 그 실상은 <훈령 587호>의 기반이 되었던 다음과 같은 문서에서도 잘 드러난다.⁹⁾

수행평가란 이전에는 하지 않았던 어떤 거창한 것이 아닙니다. 그동안 우리들이 늘 해 오던 독후 감상문 평가도 수행평가이며, 서술형, 논술형 평가도 모두 수행평가입니다. 수행평가란 학생들이 글을 쓴 것을 보면서, 노래하는 것을 들으면서, 토론하는 것을 직접 관찰하면서 평가하는 것이기 때문입니다.(수행평가 이렇게 한다, 1999 : 1)

<훈령 587호>의 개념 설정 방식을 허경철 외(1999)에서는 긍정적으로 평가하였으나, 수행평가 개념에 ‘신축성 및 융통성을 부여’(허경철 외, 1999) 하고자 한 의도는 결국 객관식 선다형 시험만 아니면 모두 다 수행평가라는 인식, 바꿔 말하면 ‘수행평가란 일종의 평가 기법에 불과하다’는 인식을 퍼트리는 계기가 되었다고 판단된다. 즉 학교에서 중간·기말 고사 때에 주로 이용하는 객관식 4지 선다형(선택형) 문항만 아니면 모두 수행평가라는 인식이 팽배할 수 있는 계기가 되었던 것이다. 이후 2001년 개정된 <훈령 616(학교생활기록부전산 처리 및 관리 지침)>에서도 수행평가에 대한

등의 수행평가 관련 연수, 교육부 자료실을 통한 수행평가 자료 제공(19,500여 명 조회), 시·도교육청의 상담창구 운영(5월 1달간 3,243건 상담) 등

- 8) 학교 생활 기록 작성 및 관리 지침은 1955년 1월 25일 문교부 훈령 제10호 이래 지속적으로 제·개정되어 왔다. 수행평가의 실시와 더불어 종래의 ‘학업 성적 관리 지침’의 ‘교과 성적’란에 수행평가 성적을 반영할 수 있도록 개정한 것은 ‘학교생활기록 작성 및 관리 지침(훈령 제587호)’이다. 이후에도 수준별 교육과정의 도입이나 교육정보시스템의 도입 등으로 인해 기존의 훈령들은 꾸준히 제·개정되어 왔다.
- 9) <훈령 587호>는 이 문서의 내용을 바탕으로 수행평가 개념을 규정하여 이 문서 내용과 동일하게 수행평가 개념을 설정함.

이러한 개념 규정은 계속 유지된다.

백남진(2007)에서 현재 학교 현장에서 수행평가가 어떻게 실시되고 있는지를 분석한 결과에 따르면, 위의 <훈령 587호>와 <훈령 616호>의 지침에 따른 16개 시·도 교육청 지침 및 단위 학교의 수행평가 관련 지침에서는 수행평가를 선다형 시험을 제외한 기타 다른 평가로 인식하고 있음이 잘 드러난다. 선택형 시험 외의 기타 다른 시험이 수행평가가 될 수 있다손치더라도 수행평가 실시의 교육적 의도나 가치를 고려하지 않은 채 단순히 선택형 시험만 아니면 모두 다 수행평가라는 인식은 타당하지 않다. 수행평가를 기법 내지 방법적 차원의 개념으로만 규정짓기는 어렵기 때문이다.

흥미로운 것은 정책 담당자들도 이와 같은 문제점을 인식하여 2008년 현재 학교 현장에 적용되고 있는 <훈령 719호>에서는 <훈령 587호>에 있었던 수행평가의 주요 용어 해설과 평가 방법의 예시를 삭제하였고, 수행평가의 개념을 규정할 때에 <훈령 587호>에 있었던 ‘즉, 선택형 지필 평가 이외 다른 모든 평가 방법이라고 할 수 있다.’라는 말도 없었다.¹⁰⁾

위의 논의를 바탕으로, 수행평가 정책의 변화에 따른 수행평가 개념에 대한 현장의 이해 변화 과정을 정리하면 다음과 같다.

<표 1> 정책 변화에 따른 수행평가에 대한 개념 이해 과정의 변화

수행평가 정책의 변화			수행평가 개념에 대한 이해
정책 시행 초기(1998년 10월)	교육비전 2002 : 새 학교 문화 창조		불분명, 이상적 개념으로 이해
정책 시행 후 홍보 노력과 정책(1999년)	‘수행평가의 이해’(1998년 12월)	한국교육과정평가원 개발, 15,000부 배포	
	‘수행평가 이렇게 한다’(1999년)	학부모용 홍보 자료, 15,000여 부 배포	

10) 2007년 한국교육과정평가원에서 주최한 ‘교육과정 운영에서의 수행평가의 적용 실태와 개선 방안’ 세미나에서 충청남도 교육청에 근무하시는 한 장학사님께서도 위의 예시가 삭제되었음을 강조한 바 있는데 이는 정책 담당자들 사이에서도 위의 예시가 갖고 있는 문제점을 인식하고 있음을 잘 보여 준다.

수행평가 정책의 변화			수행평가 개념에 대한 이해
정책 시행 후 홍보 노력과 정책(1999년)	‘수행평가의 문제점과 개선점’(1999년)	교육전문직 연수 자료, 4000여 부를 16개 시도 교육청에 배포	정책 시행 후 교육부의 홍보 노력 속에서 수행평가 개념 다소 변화 : 신축성, 융통성 있는 개념으로 변화
	훈령 587호(1999)	교육부 훈령 587호로 공포한 ‘학교 생활기록부 전산처리 및 관리 지침’에서 위 책자 내용 반영	
	훈령 616호(2001)	수행평가의 개념, 용어 설명, 평가 방법의 예, 학업 성적 결과 처리 방법 등을 상세하게 제시	
현재(2008)	훈령 719호	수행평가 방법의 예 및 학업 성적 결과 처리 방법 등을 삭제	수행평가를 평가 방법 내지 기법 차원으로 인식 : 수행평가는 객관식 선다형 지필 평가를 제외한 기타 평가 방법

그간 이와 같은 수행평가 정책의 도입 및 추진 과정과 관련하여 다양한 비판이 제기되어 왔다. 여러 연구자들이 공통적으로 지적한 가장 중요한 문제점은 역시 준비 부족이다(남명호, 2007 등). 정책 시행 초기에 수행평가에 대한 개념이 정립되어 있지 않았고, 관련 지침을 정교하게 작성하여 제시하지도 않았을 뿐만 아니라 관련 도구 및 자료에 대한 준비도 미흡하였다. 따라서 현장에서는 수행평가 실시에 대한 제반 여건을 갖추지 못한 채 수행평가를 시행하게 됨으로써 ‘고행평가’를 할 수밖에 없었고 이 때문에 각 단위 학교의 현실에 맞추어 경우에 따라서는 부적절하게 시행되기도 했다. 이 외에도 ‘위로부터의 개혁’이라는 점도 문제점으로 제시할 수 있다.

3. 수행평가 정책의 공식적 고찰

현재의 수행평가 정책은 학교 생활 기록 작성 및 관리 지침을 규정한

<훈령 719호>에 잘 드러난다. <훈령 719호>는 ‘교과학습발달상황의 평가’에 관한 주요 방침으로 ‘교과학습발달상황평가는 지필평가와 수행평가로 구분하여 실시한다.’를 제시하고 있다. 1999년 수행평가가 도입되면서 개정된 <훈령 587호> 이래 학교 생활 기록 작성 및 관리에 관한 훈령은 여러 번 제·개정되었지만 이와 같은 평가의 방침은 계속 유지됨으로써, 현재 개별 단위 학교에서 수행평가는 평가의 또 다른 한 축으로서 안정적인 기반을 확보했다고 판단된다.¹¹⁾ <훈령 719호>는 크게 다섯 가지 항목으로 구성되어 있는데, 개념, 계획의 수립 및 실시 방법(영역, 방법, 횟수, 배점, 반영 비율 등), ‘계획 및 결과’의 공개와 이의 신청 처리, 수행평가 자료의 보관, 수행평가 점수의 반영 방법 등으로 나눌 수 있다. 이 지침이 갖고 있는 특징은 크게 두 가지로 정리할 수 있다. 하나는 수행평가의 개념을 구체적으로 명시하고 있다. 둘째는 수행평가에 관한 모든 결정은 사실상 국가 수준이 아닌 개별 학교 단위 수준에서 이루어진다.

<훈령 719호>는 본래 생활 기록부에 교과 발달 상황을 어떤 방식으로 기재할 것인가를 규정한 것이다. 따라서 여기에서 수행평가의 개념을 구체적으로 기술할 논리적 근거는 적다. <훈령 719호>의 지필평가에 관한 지침은 ‘평가 문제 출제 시 주의 사항, 배점 표시, 보완, 평가 결과의 처리, 자료의 보관’ 등으로 구성되어 있다. 지필평가의 경우, 그 개념을 규정하고 이를 홍보해야 할 이유가 없기 때문에 관련 지침은 주로 지필평가 구현상의 ‘주의 사항’ 중심으로 되어 있다. 훈령에서 수행평가의 개념을 규정한 이유 중의 하나는 현장 교사들이 개념을 제대로 인식하지 못하고 있음에 대한 배려에도 있을 것이다.

최근 서울시 교육청에서는 서술형·논술형 평가 확대 정책을 추진하여, 서술형·논술형 평가의 비중을 전체의 50% 이상으로 반영하도록 하고 있다. 그런데 현재 각 학교마다 ‘서술형·논술형 평가’를 수행평가의 일종으로 보기도 하고 그렇지 않기도 하여 평가 영역 및 배점 안배에 혼

11) 그런데 <훈령 719호>에서 평가를 ‘지필평가’와 ‘수행평가’로 구분한 것의 타당성에 대한 문제 제기는 여전히 가능할 것이다. 예컨대 수행평가의 대표적 예로 볼 수 있는 ‘논술형 시험’은 지필로 시행된다.

란이 있다. 예컨대 서울의 A 중학교의 경우 평가를 크게 지필평가와 수행평가, 서술형·논술형 평가의 3가지로 구분한 뒤 각각을 30%, 20%, 50%로 배정하고 있는가 하면 또 다른 B 중학교에서는 지필 평가 50%, 수행평가 50%(서술형·논술형 평가 포함)로 규정하고 있다. 그런가 하면 서울의 한 Y중학교에서는 지필고사에 서술·논술형 평가를 포함시켜 지필 평가 80%, 수행평가 20%로 배점을 정하고 있다(백남진, 2007 : 36). 현장에서 수행평가 개념을 얼마나 혼동하고 있는지를 잘 보여 주는 사례이기도 하다. 이와 같은 개념의 혼란은 사실상, 앞서 언급한 것처럼, 10여 년간 지속되어 온 수행평가 정책의 흐름에서 이미 배태되었으며 현 정책의 안이한 조항들 속에서 확대되고 있다. 그러나 기술(既述)한 것처럼 수행평가가 단순히 기법적, 방법적 차원으로만 논의될 수 있는 것은 아니라는 점에서, 수행평가의 개념은 여전히 애매하거나 모호한 상태로 정책적으로 구현되고 있다고 볼 수 있다.

<훈령 719호>에 따르면 수행평가 실시 과정에서 부딪히게 되는 실질적인 문제들은, 요컨대 교사와 학생들이 피부로 느끼게 되는 구체적인 내역들은, 모두 ‘학업성적관리위원회’에서 정하게 되어 있다. ‘학업성적관리위원회’는 개별 단위 학교장을 위원장으로, 교감을 부위원장으로 하며 각 위원은 교직원(교사)으로 하는 단체이다. 위 지침에 따르면 수행평가는 각 교과 담당자(교사)들로 구성된 교과 협의회에서 수행평가 실시 영역·방법·횟수·세부기준(배점)·반영 비율 등을 정한 뒤 학교장의 결재를 받아 추진하게 되어 있다. 수행평가에 관한 실질적인 결정을 단위 학교 수준에서 하도록 하는 이러한 지침은 정책 도입의 초기인 1999년 <훈령 587호>부터 유지되어 온 것이다. 그런데 수행평가에 관한 실질적인 결정을 단위 학교에서 모두 결정하는 것은, 단위 학교의 자율성을 최대한 살린다는 측면에서 긍정적으로 평가할 수 있으나, 수많은 단위 학교들이 자율적으로 수행평가를 실시할 경우 학교별로, 시·도교육청별로 여러 문제점이 파생될 수 있다.¹²⁾

12) 예컨대 어떤 도에서는 ‘기본 점수 부여’를 규정하고 있는가 하면 반대로 다른 도에서는 ‘기본 점수 부여’를 금지하고 있다. 상세한 것은 뒤의 <표 2> 참조 그런데 <훈

<훈령 587호>와 비교할 때 <훈령 719호>에서는 ‘마’항이 새롭게 추가되었다. 이 조항은 수행평가의 점수 부여 방식과 수행평가 결과물의 처리 방식을 규정하고 있다. 이 조항은 일단 수행평가를 실시할 때 점수화가 가능한 영역만 선택해서 실시하라는 지침을 주고 있다. 다음 기본점수를 줄지 말지와 이의 신청 등에 대비하여 수행평가 결과물을 어떻게 보관할 것인지를 학업성적관리위원회에서 결정하도록 하고 있다. 이는 수행평가의 공정성, 객관성 문제와 관련이 있다고 판단된다. 즉 그간 수행평가 점수가 공정하게 객관적으로 부여되지 못하고 있다는 불만이 꾸준히 제기되어 왔기 때문에 이를 해결하기 위한 하나의 방책으로 제시된 조항이라고 볼 수 있다.

수행평가 실시와 더불어 제기되어 온 각종 문제점의 실상은 2003년 9월 4일 교육인적자원부 학교정책실 평가관리과 회의 자료를 통해 확인할 수 있다.

교육인적자원부 학교정책실 평가관리과 회의 자료(2003년 9월 4일)

1. 근거 : 중·고교 학업성적관리 문제점 및 대책
2. 학생, 학부모의 평가 과정·방법 등의 공정성에 대한 이의 제기 예방 -중략-
3. 수행평가 시행 횟수 과다 및 시기 집중 예방 -중략-
4. 가정 여건에 따른 과제 수행의 차이, 수행평가 과제 대행 아르바이트 성행 등 수행평가 가 빈부 격차 및 사교육 수혜 여부에 따른 학력차를 조장한다는 민원 예방 -하략-

위의 회의 자료에는 사실상 수행평가가 실시되면서 부각된 문제들의 실상이 종합적으로 제시되어 있는데, ‘수행평가의 객관성·공정성 문제, 학습 부담의 문제, 사교육 성행 문제’ 등으로 정리할 수 있다. 위의 <훈령 719호>의 ‘마’ 항은 이 중 특히 ‘수행평가의 객관성·공정성 문제’를 강조한 독립 조항인 것이다.

령>에 의하면 기본 점수 부여 방식의 결정은 단위 학교 소관 사항이다. 따라서 정책 구현의 일관성이라는 측면에서 문제가 제기될 수 있다.

각 시도 교육청에서는 <훈령 719호>에 근거해 학업성적관리규정을 정한다. 대부분의 경우 <훈령 719호>와 유사하지만, 꼼꼼히 살펴보면 각 시도 교육청별로 위와 같은 문제들을 해결하기 위한 관련 규정을 따로 명시하여, 각 시도 교육청별로 수행평가와 관련된 집중적인 문제가 무엇인지를 보여 주고 있다. 그러한 문제들은 대부분 위 의 회의 자료에 명시된 것들로 수렴된다. 각 시도 교육청의 학업성적관리규정 중 <훈령 719호>에서 언급되지 않은, 각 시도 교육청별 특성으로 볼 수 있는 내용만 간단하게 정리하면 다음과 같다. 위 의 회의 자료에 명시된 ‘학습 부담의 문제’와 ‘사교육 성행 문제’는 크게 보면 결국 ‘객관성·공정성 문제’로 정리되지만 여기서는 위 의 회의 자료의 틀도 타당성이 있으므로 이에 준하여 정리하도록 한다. 위 의 회의 자료에서는 언급되지 않았지만 각 시도교육청별 수행평가 관련 지침의 특징이 있는 경우 기타 특징으로 제시하였다.

<표 2> 각 시도교육청별 수행평가 관련 지침의 특징

교육청	특 징	
경기도 교육청	평가의 객관성· 공정성 문제	<p>—평가 문항 출제 시 동일 학년 동일 과목을 2명 이상의 교사가 담당하고 있을 때는 반드시 공동 출제 및 공동 편집을 하고, 채점도 공동으로 하도록 한다.</p> <p>—채점기준표에는 모범 답안, 유사 답안의 처리, 풀이 과정에 따른 부분 점수, 단계별 점수 등을 제시하여야 하고 채점기준표 작성의 번거로움이나 채점의 객관성을 지나치게 의식한 단답형 출제 등은 지양한다.</p> <p>—그룹평가(또는 조별평가)는 학생 개개인의 노력 정도 및 성과에 대한 평가가 이루어지도록 한다.’</p>
	학습 부담의 문제	—정기 교사 준비 기간과 수행평가 시기가 겹치지 않게 적절히 조정하여 시행한다.
	사교육 성행 문제	<p>—반드시 수업 시간 내에 수행평가가 이루어지도록 한다.</p> <p>—예체능 교과외의 경우, 영역별로 골고루 평가하고 반드시 수업 시간 내에 이루어지게 한다.</p>
	기타 특징	—태도 점수는 가급적 교과 내의 정의적 영역과 관련성이 있는 것으로 한정한다.
부산 광역시	평가의 객관성· 공정성 문제	—점수화가 가능한 영역만을 평가하되, 계획과 결과를 학교 홈페이지 등을 통해 학생·학부모에게 공개한다.

교육청	특 징	
부산광역시	기타 특징	—정의에서 ‘선택형 지필 평가 이외 다른 모든 평가 방법’이라고 규정하면서 수행평가의 개념 규정뿐만 아니라 수행 방법도 예시(훈령 587호의 예를 가저음)
인천	평가의 객관성·공정성 문제	—과도한 기본 점수 부여로 인한 성적 부풀리기가 발생하지 않도록 유의한다.
전라북도	평가의 객관성·공정성 문제	—그룹평가(또는 조별평가)는 학생 개개인의 노력 정도 및 성과에 대한 평가가 이루어지도록 한다. —학교 수행평가 신뢰도 제고를 위하여 학생 및 학부모를 대상으로 수행평가 인식에 대한 모니터링을 실시하여 그 결과를 차기 수행평가 계획 수립에 반영한다.
	학습 부담 문제	—중간고사 및 기말고사 때 수행평가가 겹치지 않도록 시기를 적정화한다.
	사교육 성행 문제	—과제물 위주의 수행평가를 지양하고 정규 수업 시간을 활용하여 평가 실시한다.
	기타 특징	—도서관 활용 수업 강조(독서 과제·결과물의 수행평가 반영)(가급적 전 교과에 적용)
전라남도	평가의 객관성·공정성 문제	—수행평가의 기본 점수는 인정하지 않는다.
	사교육 성행 문제	—교수·학습과 평가의 연계로 과제물 위주의 수행 평가를 지양하고 결과와 과정을 함께 중시하는 수행 평가를 권장한다.
대구	학습 부담 문제	—과제물은 적정량을 제시한다.
경상북도	사교육 성행 문제	—과제물 위주의 수행평가를 지양하고 정규수업 시간을 활용하여 평가한다. * 부득이 과제형 수행평가를 실시할 경우의 지침을 박스 처리하여 제시 —학생 수준에 적합한 수행평가 과제를 부여한다. —학년협의회 및 교과협의회를 통해 과제의 적절성을 사전에 협의한다. —학생 스스로 해결할 수 없거나 부모 등 외부의 도움이 필요한 과제는 지양한다. —예· 체능 교과의 수행평가는 영역별로 균형 있게 시행한다.
		—수행평가에 대한 개인별 과제물이 적당하게 부과될 수 있도록 학교 차원의 대책이 수립되어야 한다. —과제 중심 수행평가는 지양한다. —예체능 교과의 수행평가는 영역별로 균형 있게 시행한다.
제주도	사교육 성행 문제	—수행평가의 개념 규정뿐만 아니라 수행 방법도 예시(훈령 587호의 예를 가저음)
	기타 특징	수행평가의 개념 규정뿐만 아니라 수행 방법도 예시(훈령 587호의 예를 가저음)

위의 표에서 드러난 것처럼 각 시도 교육청별 학업 성적 관리 규정은 사실상 위의 회의 자료에서 언급된 수행평가 실시 관련 문제들을 해결하기 위한 주요 대안의 성격을 지니고 있다. 그 외의 특징은 다음 두 가지로 정리된다.¹³⁾ 첫째는 경기도 교육청에서 특별히 ‘태도 점수는 가급적 교과 내의 정의적 영역과 관련성이 있는 것으로 한정한다.’고 규정한 것이고, 또 하나는 전라북도 교육청에서 특별히 독서 활동을 강조하여 국어 교과 외의 기타 다른 교과에서도 독서 활동을 수행평가에 반영하도록 권장하고 있는 점이다. 후자는 각 시도 교육청의 중점 수행평가 관련 정책을 보여 준다는 점에서 의미가 있다. 전자는 역시 수행평가와 관련하여 제기되는 문제 중의 하나를 해결하기 위한 방안에서 비롯된 것으로 보인다. 즉 단순히 수업 태도 점수를 수행평가 점수로 부과하는 폐단을 의식한 조항이라고 볼 수 있다.

각 시도 교육청별로 다소의 차이는 있지만, 각 시도 교육청별 수행평가 관련 정책이 갖는 가장 중요한 특징은 10여 년 동안 수행평가가 실시되면서 생겨난 각종 문제점들을 해결하기 위한 각종 방안들의 상세화라는 데에 있다. 그 방안은 인천광역시처럼 과도한 기본 점수 부여를 지양하도록 하거나 전라남도처럼 아예 기본 점수 부여를 인정하지 않는 등 다소 차이를 보이기도 하지만, 대체로 그 초점은 같다.

한편 단위 학교에서는 소속된 시도 교육청의 지침에 근거하여 자체적으로 학업성적관리 규정을 정한다. 대체로 시도 교육청의 지침을 그대로 준용하는 경우가 많은데 각 학교별로 필요한 내용들만 발췌하거나 학교별 특수성을 반영한 새로운 내용을 중심으로 각 지침을 다소 상세화하기도 한다. 예시로 전라남도에 소속된 두 학교의 수행평가 관련 지침을 제시하면 다음과 같다. 전자는 간략화의 예이고 후자는 상세화의 예이다.

13) 예컨대 부산광역시처럼 수행평가의 방법을 1999년에 제시된 <훈령 587호>에 근거해 좀 더 자세하게 제시한 것이 있으나, 이 역시 각 시도 교육청의 독립된 규정이라기보다 과거의 훈령에 근거하고 있다는 점에서 논의에서 제외하였다.

<전남 ○○ ○○ 중학교의 수행평가 관련 지침>

제15조 (수행평가) -중략

- ③ 교수·학습과 평가의 연계로 과제를 위주의 수행 평가를 지양하고 결과와 과정을 함께 중시하는 수행 평가를 권장한다. -중략
- ⑥ 모둠별 평가의 경우는 개인별 능력보다는 모둠별로 채점하여 공동점수를 부여한다.
- ⑦ 수행평가의 영역은 3개 이상 실시하고 평가는 3단계 이상 차별화하여 점수를 부여한다.
(밀줄 필자)

<전남 △△ 여고 수행평가 관련 지침>

제13조(수행평가)

- ① 수행평가는 모든 교과에서 시행한다. -중략
- ⑧ 수행평가의 점수는 점수화가 가능한 영역의 점수만 반영하되, 태도평가 등과 같이 점수화가 어려운 영역이라도 별도의 누가 기록에 의해 등급화하여 점수를 반영할 수 있다. -중략
- ⑩ 단일별 또는 월별 평가 계획을 세밀히 수립하여 시행하되 학생의 부담이 과중되지 않도록 하고, 정기고사 준비기간과 수행평가 시기가 겹치지 않게 적절히 조정하여 시행한다.
(밀줄 필자)
- ⑪ 과제 제시형 수행평가를 지양하고 수업 시간 학습활동 위주의 과정평가를 권장한다. -중략
- ⑫ 수행평가의 기본점수는 인정하지 않는다. -하략 (밀줄 필자)

위의 전남 ○○ ○○ 중학교는 ③번처럼 전라남도 교육청의 특별 지침을 제시하면서 ⑥번, ⑦번처럼 교육청 지침에서는 구체적으로 언급하지 않은 내용을 학교 상황에 맞게 새로 설정하고 있다. 전남 △△ 여고도 ⑪번, ⑫번처럼 전라남도 교육청의 특별 지침을 제시하면서 ⑧번, ⑩번처럼 교육청 지침에서는 구체적으로 언급하지 않은 내용을 새로 설정하고 있다. 그런데 각 단위 학교에서 제시하고 있는 지침 내용도 큰 틀에서 보면 수행평가와 관련하여 제기되는 각종 민원들을 해결하기 위한 방안 안에 들어 있음을 알 수 있다.

한편 수행평가 정책은 초등학교와 중·고등학교의 경우 다소 다르다. <훈령 719호>는 ‘② 초등학교의 교과학습발달상황은 각 교과의 학습활동

발달 정도, 수행평가 결과, 특징 등을 종합하여 ‘세부 능력 및 특기 사항’란에 과목별로 간략하게 문장으로 입력한다.’고 규정하고 있다. 중·고등학교의 경우 ‘점수화가 가능한’ 경우에 한해 수행평가를 실시하는 것과는 매우 다르다. 그런데 현재 초등학교에서 수행평가는 사실상 공개 요구에 대비하여 보존되고, 학기말의 통지표 작성 근거로 쓰이는 것 이외에는 별 활용성이 없다. 수행평가 결과를 기록하는 칸이 매우 작아 간단한 의견을 제시하는 것 외에 다른 기술은 어려운 상황이다. 이 때문인지 통지표의 기술은 학습자의 긍정적인 특성에만 초점을 맞추어 기술되고 학습자들의 국어적 사고력의 수준에 대한 상세한 기술 및 그와 관련된 교육적 조언은 이루어지지 못하고 있다. 수행평가 결과의 송환은 딱 통지표에 드러나는 만큼만 이루어지고 일방향적인 것이다(이수진, 2008 : 46).

중·고등학교의 경우, ‘점수화가 가능한’ 경우에 한해 수행평가를 실시하라는 조항은 앞서 언급한 것처럼 1999년의 <훈령 587호>에는 없던 것이며, 10여 년 동안 수행평가를 실시하면서 평가의 객관성·공정성에 대한 문제 제기가 끊임없이 이루어지자 이에 대한 문제점을 해결하기 위해 새롭게 부과된 것이다. 그런데 양적 평가가 질적 평가보다 더 객관적이라는 교육학적 증거는 사실상 없다. 4지 선다형 객관식 평가나 수량화된 점수가 그렇지 않은 평가보다 더 객관적이고 공정한 평가라는 인식은 평가의 목적과 대상, 방법적 특징을 무시한 안이한 판단이라고 아니할 수 없다. 더욱이 ‘양적 점수화’는 수행평가의 본질과는 거리가 멀다(백순근, 2002).

지금까지의 논의를 통해 알 수 있는 것은 현재의 수행평가 정책은 수행평가와 관련하여 제기되는 여러 가지 문제점들을 해결하기 위한 대안적 성격이 강하다는 점이다. 수행평가의 본질을 정책적으로 구현하고자 하는 의도보다는 관련 민원을 해결하고자 하는 수세적, 방어적, 소극적 제안들로 채워져 있다는 점이다.

이경아(2005)는 우리나라 수행평가 정책에 대한 평가 연구에서, 현장 교사들의 설문 조사 결과를 바탕으로 ‘수행평가 정책은 정책 목표의 타당도가 높고 교육적으로 매우 의미 있는 정책으로 인식되고 있으나 정책 환경의 변화에 수행평가 정책이 적절히 대응했다고 보기는 어렵다’는 결론

을 제시한 바 있다. 현재의 수행평가 정책은 수행평가 관련 문제점을 해결하면서도 수행평가 도입의 본래 취지를 살릴 수 있는 방안을 탐색하여 평가의 본질을 구현하는 데에는 상당히 미흡하다. 정책을 추진할 의도가 있다면, 정책 환경의 변화를 면밀히 분석하여 수세적, 방어적, 소극적 정책 보다는 좀더 전문적, 분석적, 적극적인 정책을 제시할 필요가 있을 것이다.

4. 수행평가 정책의 개발 논리에 대한 재고

우리나라의 수행평가 정책 시행은 수행평가에 관한 서구의 교육 이론 및 실체가 국내에 소개되면서 이루어졌다. 특히 우리나라의 경우 교육 제도가 유럽보다는 미국과 유사하기 때문에 수행평가 정책 시행 과정에서 미국의 교육 정책이 일정 부분 영향을 미친 것으로 판단된다. 그러나 물론 미국과 우리나라의 교육 현황과 수행평가 실시 과정은 서로 다르다.

남명호(2007)에서 논의된 것처럼 미국에서 수행평가가 1980년대 초반 다시 부각되게 된 데에는 ‘책무성’에 대한 교육계의 반발이 중요한 요인이 되었다. 수행평가는 오래 전부터 엘리트 교육을 표방하는 미국의 사립 학교에서 활발하게 이루어져 왔다. 그런데 중등교육의 성격이 엘리트 교육에서 대중을 위한 보편 교육으로 바뀌면서 비교적 비용이 적게 들고 시행이 간편한 표준화 검사에 관심을 갖게 된다. 이런 상황에서 1965년, 연방 정부로부터 재정 지원을 받는 학교에 대해 평가를 의무화하도록 하는 내용의 초·중등교육법이 통과되면서, 학교에서는 학생들의 점수를 올리기 위한 ‘시험 준비 수업’(teaching to tests), ‘시험에 끌려 다니는 수업(measurement-driven instruction)’을 할 수밖에 없었다(남명호, 1995). 따라서 선택형 표준화 시험에 근거한 책무성의 압력에 대한 교원들의 반발이 거세졌고 1980년대 초반 교육 개혁 운동의 전략적 도구로 부각된 것이 바로 수행평가이다.¹⁴⁾

최근 미국에서는 1983년 ‘위기에 처한 국가’(A Nation at Risk, NCEE 작성)

라는 보고서 이후 다시 책무성의 논리가 부각되고 있다. 책무성의 논리에 대한 강조는 한편으로 국가 수준의 성취 기준 제정을 강조하는 방향으로 흐르고 있다. 미국에서는 1989년 ‘Goals for the year 2000’을 채택하면서, 국어, 수학, 과학, 역사, 지리 등 5개 교과에서 국가 기준을 마련키로 결정했고 이는 1994년 ‘Goals 2000-Educate America Act’라는 법으로 구체화되기에 이른다. 현재 이 법에 따라 성취 기준이 설정 및 상세화되어 있는 상태이다.

허운나·김영옥(1998 : 134)은 미국 교육 개혁의 흐름을 논하면서 1960년대 교육 개혁의 주제였던 ‘평등’은 1970년대 말 이후 특히 1983년 이후 ‘수월성’이라는 주제로 대치되었다고 평한 바 있다. 여기서 ‘수월성’은 ‘책무성’의 또 다른 얼굴이라고 볼 수 있다. 그간의 논의를 중심으로 책무성 논리와 그 대(對)가 되는 지원 및 평등의 논리가 갖는 특징을 정리하면 다음과 같다.

<표 3> ‘책무성 논리’와 ‘지원 및 평등의 논리’의 특징

	책무성 논리	지원 및 평등의 논리
기본 특징	위기에 처해 있는 학생의 학업 수준을 높이기 위해 학교는 책임을 다해야 한다.	정부는 학교 지원에 초점을 두어야 한다. 정부는 교육적 불평등을 심화시켜서는 안 된다.
주장 내용	1. 학생의 학업 성취는 국가 경쟁력과 직결된다. 2. 학교는 국가와 사회가 투자한 만큼 학생 성취도 제고라는 결과물을 생산할 책무가 있다. 3. 학생의 학업 성취 정도와 학교의 책무성을 평가하기 위해서는 교육 내용과 평가 기준을 국가 수준에서 표준화해야 한다.	1. 교육에 대한 중앙집권적 통제 강화로 교사의 수업 자유를 침해할 우려가 있다 2. 표준적으로 평가할 수 있는 내용만 평가함으로써 교육 내용이 왜곡될 수 있고 교육 수준이 오히려 저하될 수 있다 3. 학생의 가정 환경이나 개인적인 차이를 무시하고 일률적인 기준의 성취를 강요함으로써 오히려 교육적 불평등을 심화시켜 소외 계층의 학생들을 학교에서 배제할 가능성이 높다

미국의 성취 기준 제정 작업은 수행평가의 실시와 전혀 다른 두 가지

- 14) 교육자들 또한 다시 선다형, 규준 지향적 평가 시험이 책무성의 증거로 사용되는데 반감을 표출하게 되고 책무성 전략에 대한 재고를 요구하게 되었다(Khattari & Sweet, 1996, 김정자, 1999 : 160에서 재인).

방향으로 영향을 미쳤다고 판단된다. 첫째는 성취 기준을 수행 기준으로 제정비하는 계기가 되어 수행평가를 공고히 하는 데 기여하는 방향이다. 김영천(2002 : 150)에 의하면 Maryland의 MSAP, Kentucky의 CAT 평가 시스템은 학생들의 학업 성취도 검사를 수행 과제 중심으로 실행하는 대표적인 평가 제도로 정착했다고 한다. 평가 문항이 선다형에서 개방형 지필평가나 수행 과제로 변화되면서 학교의 교육과정과 수업 그리고 학습의 방법이 변화되고 있다고 한다.

이는 수행평가가 제대로 실시되기 위해서는 성취 기준이 수행 기준화될 수 있어야 함을 시사한다. 특정 평가 방법을 통해 소기의 성과를 내고자 하는 교육 개혁은 교육과정 특히 교육 내용과의 일관성 속에서 함께 논의되어질 때 그 타당성을 인정받을 수 있다(김경자, 1999 : 157~160). 미국에 수행평가가 안정적으로 정착한 학교가 존재하게 된 데에는 수행평가의 실시가 교과 내용의 개혁(성취 기준의 제정비)과 함께 추진된 데에도 그 이유가 있다고 판단된다.¹⁵⁾

우리나라의 경우에도 성취 기준의 제정 작업이 이미 오래 전부터 추진되었다(이에 대한 상세한 논의는 김진숙, 1999 참조). 하지만 그러한 작업의 방향이 미국의 몇몇 주처럼 수행 기준화의 방향으로 이루어지지는 않았다. 우리나라의 교육 현장에서 수행평가가 ‘수행’ 그 자체만을 강조하면서(물론 진정한 의미에서의 ‘수행’이라고 보기도 어렵다) ‘교육 내용’은 전혀 고려하지 못하는 것도(예를 들어 교과서나 노트 검사, 수업 태도 검사를 수행평가라고 하는 것) 교육 내용은 변하지 않은 상태에서 평가의 변화만을 유도하려고 한 교육 정책에 기인하는 것인 지도 모른다. 이러한 측면에서 김경자(1999 : 157)에서 수행평가의 안정적 정착을 위해서는 교육 내용에 대한 개혁도 함께 수반되어야 함을 논한 것은 의미가 있다. 강승호(2001)에서도 수행평가 실시상의 문제점을 해결하기 위한 대안을 제시하면서 ‘교육과정의 개편이나 재구성’이 필요함을 논한 바 있다. 현행 교육과정과 교과서의 경우 서

15) 우리의 경우 수행평가 실시가 1999년부터 이루어진 데 비해 제7차 교육과정이 고등학교의 경우 2002년부터 적용되었고 각종 교과서의 개편도 2000년 이후에 이루어졌다. 수행평가만 유독 먼저 시행되면서 초기 학교에서는 어려움을 겪을 수밖에 없었다.

로 중복되는 학습 내용들을 다루고 있어 깊이 있는 학습이 어려울 뿐만 아니라 그 목표와 내용이 수적으로 많기 때문에 수행평가는 교사와 학생 모두에게 부담이 될 수 있다는 것이다. 학교 현장에서 굳이 수행평가로 할 필요가 없는 것을 수행평가로 실시하기도 하고 경우에 따라서 무엇을 수행평가로 평가해야 하는지를 판단하지 못하는 것은 ‘수행평가’의 개념에 대한 불철저한 인식 외에 ‘무엇’이 수행평가의 대상이 되는지를 주어진 교육 내용(성취 기준)의 틀 안에서 골라내기 어렵기 때문일 수도 있다.

한편 미국의 성취 기준 제정 작업은 수행평가의 실시에 부정적인 방향으로 영향을 미치고 있기도 하다. 성취 기준의 제정은 평가의 방향을 전반적으로 양적 평가를 강조하도록 만들 수 있다. 현재 미국의 많은 주에서 수행평가가 약화되는 경향이 있는 것은 성취 기준의 강조가 양적 평가에 대한 강조로 이어지고 있기 때문이다.

미국 내에서도 위에서 논한 책무성 논리 강화를 비판적으로 보는 시각도 적지 않다. <표 3>에 제시된 ‘지원 및 평등의 논리’가 이를 대표한다. McIntush는 ‘A Nation at Risk’라는 보고서를 기점으로 교육에 대한 인식이 사회 정치적 평등을 위한 수단에서 경제적 번영을 위한 수단으로 전환되었다고 평했다(Ericsson, 2005 : 226 제인). Eisner(1993)는 교육은 학생이 가지고 있는 다양성과 이질적인 속성을 장려하고 육성해야지 일사불란한 기준과 목표, 시험, 성적표에 의해 학생을 한쪽으로 몰아가서는 안 된다고 하면서, 기준은 양적 측정 도구로 기능하기 때문에 수 계산이나 맞춤법, 어법의 정확성을 판단하기에는 적합하지만, 보고서, 작품, 한편의 완성된 글 등 보다 고차적인 수준의 학생 수행에 대하여 질적인 판단을 하기에는 적합하지 않다고 지적하였다. 작품에 대한 평가는 기준이 아닌 준거에 의해 이루어지는데 준거는 경험 많은 교사의 전문성이나 식견에 기반한 판단에 의지하는 것이기 때문이다(밑줄 필자). 배려 교육을 주장한 Noddings(1997) 역시 기준 설정보다 더 중요한 것은 학생들이 학습을 하기 위해서 학생과 교사를 어떻게 지원할 것인가와 같은 문제를 해결하는 일이라고 하였다(이상 정혜승, 1997 참조). 그러나 미국에서는 현재 책무성의 논리가 여전히 우세하다고 판단된다.

이러한 미국 교육 정책의 흐름에서는 ‘책무성 논리—국가 수준의 단일화된 성취기준 강조—양적 평가 활성화—수행평가의 약화’와 ‘지원 및 평등의 논리—교사의 평가권과 자율권 강조—질적 평가 활성화—수행평가의 강화’의 흐름을 추출, 비교해 볼 수 있다. 이를 통하여 수행평가가 활성화되기 위해서는 국가 수준의 일반화된 성취 기준보다는 교사의 전문성에 대한 신뢰에 기반하여 교사의 평가권과 자율권을 강조하고, 양적 평가보다는 질적 평가를 강조해야 함을 알 수 있다.

다시 우리나라 수행평가 도입을 공시한 ‘교육비전 2002 : 새 학교문화 창조’로 돌아가 보자. 이 문서에서는 새 학교문화 창조를 위해 6가지의 추진 과제¹⁶⁾를 제시하고 있는데 이중 ‘4) 평가의 다양화 및 투명성 보장’ 부분에 수행평가 관련 내용이 들어 있다. 수행평가 도입의 취지와 관련하여 주목되는 것은 ‘목표지향평가를 위한 학업 성취 기준 상세화 자료 개발, 보급’이다. 그리고 수행평가는 다섯 번째 추진 과제인 ‘5) 교원의 전문성, 책무성 제고’와 연계하여 추진되는 것이었다. 그리고 수행평가 실시를 공식적으로 고시한 1999년의 <훈령 587호>를 보면 성적 채점 기준의 예시가 있다.

<표 4> ‘훈령 587호(1999)’에 제시된 성적 채점 기준의 예시

1999학년도 제1학기 국어과 성적 일람표													
제1학년 ()반					교과(과목)담당교사()인								
평가방법 (반영비율)	선택형 지필평가 (60%)				수 행 평 가 (40%)						합 계	성 취 도	석차/ 재적수
명칭, 영역 (반영비율) 번호, 성명	1회 (30%)	2회 (30%)			말하기 (10%)	듣기 (10%)	읽기 (10%)	논술 (10%)			100%		
1 김길동	28.5	29.4			8.8	9.6	8.8	10			95.1	수	4(15)/532
2 나갈길	25.5	19.2			6	8	7	5			70.7	미	273/532
3													

16) 1) 학교 토론 문화 형성, 2) 교수—학습 과정의 개별화 실현, 3) 다양한 체험 학습 수행, 4) 평가의 다양화 및 투명성 보장, 5) 교원의 전문성, 책무성 제고, 6) 학교 경영의 자율성 증진

이를 보면 알 수 있듯이 수행평가 결과는 수량화되어 반드시 양적 점수로 환산되어야 한다. 따라서 교사들에게 수행평가란 성적 산출을 위한 ‘선택형 지필평가가 아닌’ 다른 모든 종류의 평가 방법이나 기법을 가리키는 것으로 이해될 수 있다. 물론 수행평가 결과가 10점, 20점처럼 수량화될 수도 있지만 수행평가라면 ‘어떠어떠한 점은 훌륭하나 어떠한 점은 부족하다’는 식의 질적 기술도 할 수 있도록 허용되고, 권장되어야 한다. 그런데 위의 정책을 보면 사실상 그런 평가는 쓸모가 없어 보인다. 또한 위의 정책에서 국가 수준의 성취 기준을 상세화하여 교사로서의 책무를 다해야 함을 강조하고 있기 때문에, 교사들에게 수행평가 정책이란 국가 수준의 성취 기준을 상세화하여 이 성취 기준 도달 여부를 측정하되 단지 선택형 지필평가로만 하지 말고 여러 가지 다양한 방식으로 하라는 지침 정도로 이해되기 쉬운 것이다. 현장의 교사들이 수행평가를 의당 성취도 평가의 일종으로 인식하는 것은 위의 정책 내용을 분석해 볼 때에 자연스러운 일이다.

수행평가는 지식이 상황이나 맥락에 따라 변하며 개개인에 의해 창조되고 구성되고 재조직되는 것이라고 가정하며(지식관, 절대 평가, 질적 평가를 평가 체제로 하고 여러 줄 세우기를 목적으로 하는 평가이다. 평가 방법에 있어서도 객관성과 공정성보다는 전문성과 타당성, 적합성을 강조한다(백순근, 2002 ; 천경록, 2001 ; 최미숙, 1999 등). 그런데 위에서 논한 바와 같이 우리나라에서 수행평가란 개인과 독립적으로 존재하는 지식(즉 교육과정 및 교과서에 제시된 지식, 요컨대 국가 수준의 성취 기준)을 평가하되 결국은 수량화해야 하며 그 결과 학생들의 성취도 정도를 판단하는 ‘객관적, 투명한 평가’가 되어야 한다. 학생들의 성적을 제시해야 하므로 한 줄 세우기는 어찌면 수행평가를 실시하기 전에 이미 전제되어 있는 것이다. 따라서 우리나라의 수행평가 정책은 그 특성상 수행평가의 본질적인 개념에 맞게 실행되기 어렵다는 심각한 문제점을 안고 있다.

진정한 평가의 목적이 구현되기 위해서는 평가 보고 양식이 갖는 주관성을 인정해야 한다. 그리고 등급을 매기고 성적을 보고하는 것이 교수에 필수적인 과정은 아니며 평가에 대한 최상의 증거는 시험이 아닌 다른

방법으로 이루어져야 한다(Guskey eds., 1996 : 14~19, 김영천, 2002에서 재인). 따라서 다음과 같은 지적은 의미심장하다.

성적 : 중요한 것은 ‘어떻게’가 아니라 ‘왜’이다.

오직 전통적인 방식의 성적 양식을 버리는 것만이 교육적인 목적에 부합하는 수행평가를 성취시킬 수 있다(Kohn, 1994 : 38).

미국의 경우, 1990년대 중반부터 수행평가 이론가들이 수행평가에 맞는 성적표 양식의 개발 작업을 시작하였고 새로운 양식의 성적표들이 제작되었는데, 그 형식들은 수행평가가 추구하는 학습 결과를 적절하게 기록하고자 심도 있게 만들어졌다.¹⁷⁾ 우리의 경우 기술(既述)한 것처럼 초등의 경우에도 수행평가 결과는 오로지 긍정적 기술 몇 마디로만 이루어진다. 실상 학습자들의 국어 수준에 대한 제시와 능력 향상을 위한 교사의 조언은 없다.

수행평가 결과를 성적 보고 시 의미 있게 기술하기 위해서는 수행평가의 평가 기준표가 좀 더 체계적이고 명료할 필요가 있다. 그런데 현장의 교사들에게 평가 기준표를 만드는 일은 매우 지난한 작업이다. 미국의 경우 하나의 교과 단원의 주제에 대한 평가 기준표를 제작할 때에, 아이디어를 제공할 수 있는 전문가 집단이 그 작업에 공동으로 참여하며 개발 기간도 몇 달에서 몇 년에 이른다(김영천, 2002 : 146). 이에 비해 우리의 경우 평가 기준표의 개발 작업은 개별 교사에 의해 이루어지기 때문에 집단 또는 학회 수준에서 만든 평가 기준표에 비하여 그 수준이 낮을 수밖에 없다. 정책 당국에서 좀 더 적극적으로 이에 대한 지원을 펼칠 필요가 있음이 명확하다. 이러한 점에서 국어과 평가의 개선 방향을 논한 최미숙(2006)에서 성적 기록 방식의 문제점을 지적하고 이를 개선해야 함을 강조한 것은 의미가 있다.

17) 김영천(2002)에 의하면 Guskey, T. eds.(1996)는 전통적인 성적표 양식의 문제점은 무엇이고 그 대안은 무엇인가에 대한 이론적/실제적 이슈들을 논의함으로써 ASCD의 1996년 최고 연구 작품으로 뽑혔다고 하며 그 외에도 ‘Azwell & Schmar eds.(1995), Tschudi ed.(1997) 등의 연구가 있다고 한다.

정부 지원의 불충분에 관한 논의는 오래 전부터 제기되어 왔다. 특히 실험 실습에 필요한 기자재의 부족이나 실기 장소의 부족과 미정비 등에 대한 지적은 사실상 어제오늘의 교육 현장 요구가 아니다. 수행평가의 시행을 위해서는 시설 등의 여러 지원 방안이 특히 더 요구된다고 볼 때, 우리의 경우 ‘지원’에 대한 논의 없이 ‘책무성’만 강조하는 상태로 수행평가 정책이 추진되었다고도 말할 수 있을 것이다. 앞서 살핀 바, 이는 정책을 추진해 놓은 뒤에야 현장 교사들에게 수행평가를 이해시키기 위해 부랴부랴 관련 자료를 배포한 사실에서도 드러나고 있다. 정책 초기에 수행평가가 교사들에게 ‘압력’으로 작용하기도 한 것은 어쩌면 자연스러운 귀결일지도 모른다.¹⁸⁾

Cuban(1993)은 교육 개혁을 점진적 변화와 근본적 변화를 요구하는 두 가지 타입으로 나누었는데, Cuban의 범주를 따르면 수행평가 실행은 근본적인 변화를 요구하는 개혁이다(최진영, 2007 : 351). 단순히 기존의 전통적인 학교 구조 안에서의 변화가 아닌 전통적인 학교 구조의 전면적인 변화를 요구하는 개혁이라는 뜻이다. 학교 구조의 전면적인 변화는 교수·학습에 대한 새로운 관점에 기반한 새로운 패러다임을 요구하며 수행평가 역시 이러한 새로운 패러다임 안에서 실행 가능한 개혁이라고 볼 수 있다. Wilson(1990)이 정책 실행 과정에서 교사들이 보이는 양상을 연구한 바에 따르면, 새로운 교육과정 실행에 있어 교사가 ‘했던 것’은 교사가 ‘할 수 있는 것’이었다(최진영, 2007 : 356). 따라서 우리나라에서 수행평가가 제대로 실시되기 위해서는 근본적 변화를 담보하는 지원과 교육 구조의 개선이 필요하다.

18) Firestone et al(2001)은 교사 변화를 위해서는 압력과 지원이 함께 필요하다고 하였다(최진영, 2007 : 351에서 재인). 수행평가의 실시는 교사들에게 수행평가 실행을 위한 압력으로 작용할 수 있다. 그런데 위에서 아래로의 평가 변화 지시는 교사들로 하여금 근본적 변화가 아닌 형식적 실행을 하게 할 뿐이다.

5. 결론

우리나라의 경우 수행평가 정책을 추진하는 과정에서 교과 교육 내용의 특성이나 성격을 고려한 교육 내용의 정비 작업이 함께 이루어지지 못했다. 이 점은 수행평가가 그 본질을 구현하기 어렵게 만드는 요인 중의 하나였다. 또 양적 평가를 중시하는 교육 제도의 틀과 전통을 고려하지 않은 채 평가 방법만을 혁신하려고 한 점 역시 실패의 요인이 되었다고 판단된다. 무엇보다 교사의 수준이나 인식, 학교의 행정적 여건, 수능 시험 준비로서의 학교 운영의 실체 등 학교 내적인 변수들을 교육 정책에 충분히 반영하지 못한 점이 현재와 같은 수행평가 실태를 낳는 요인이 되었을 것이다.

정책의 목표가 아무리 타당하다고 해도 이의 실행을 위한 정책 수단이 제대로 준비되지 않는다면 정책은 집행 단계에서 혼란을 겪을 수밖에 없다. 정책의 형성·결정 단계에서 정책 수단의 실현 가능성을 충분히 점검해 보아야 한다. 그리고 정책 집행 단계에서 정책의 환경 변화에 보다 주도면밀하게 대처하는 능력이 필요하다. 그리고 정책 환경의 변화를 정확히 읽는 것도 중요하지만 이에 대한 대안을 제시할 때에는 신중을 기해야 한다. 정책 추진의 취지는 뒤로 한 채, 단순히 임기응변식으로 ‘문제만 생기지 않도록’ 하는 데 급급한다면 그 정책은 실패하기 쉽다. 정책 집행 후 단계에서도 주기적인 정책의 효과성 진단이 필요할 것이다.

서민원(2001)에서 수행평가의 학교 현장 적용상의 문제를 수행평가 자체에 대한 이해 수준으로 파악하여 이를 분석의 내적 준거로 설정하고 분석한 결과에 의하면, 수행평가의 문제 중에서 교사들의 이해 수준이 가장 낮은 요인은 수행평가의 목적이었다. 서두에서 논의한 것처럼 국어 교사들의 상당수는 현재 수행평가를 수업 분위기를 좋게 하는 일종의 도구로 사용하고 있기도 하다. 미시적 측면에서, 수행평가 도입의 취지를 구현할 수 있는 국어교육적 개선 방안에 대한 탐색 역시 여전히 유효한 연구 주제일 것이다.

본고는 현 국어과 수행평가 실행 양상의 원인을 정책을 중심으로 검토하다보니, 국어교육적 특수성에 대한 심도 있는 논의는 하지 못했다. 앞으로 국어교육적 수행평가가 갖는 특수성과 문제점에 대한 논의도 좀 더 깊이 있게 이루어질 필요가 있을 것이다.*

* 본 논문은 2008. 10. 30. 투고되었으며, 2008. 11. 4. 심사가 시작되어 2008. 11. 26. 심사가 종료되었음.

■ 참고문헌

- 강승호(2001), “학교현장에서의 수행평가 실시상의 문제점과 정착 방안”, 교육평가연구 14(1), pp.1~21.
- 교육부(1999), 수행평가 이렇게 한다.
- 교육부(1999), 수행평가의 문제점과 개선 방향.
- 구민경(2006), 중학교 국어과 수행평가의 실태와 개선 방안 연구, 경남대 석사 논문.
- 김경자(1999), “초등학교 수행평가의 의미와 그 개혁 전략 : 미국 사례를 중심으로”, 초등교육연구 13-1, pp.157~184.
- 김대조(2004), “읽기 영역 수행평가 목표의 위계화 방안”, 한국초등국어교육 26, pp.39~67.
- 김명숙(2001), “수행평가의 가치와 질 관리”, 교육평가연구 14-2, pp.17~46.
- 김슬아(2007), 국어과 수행평가 실태 및 개선 방안 연구, 한양대 석사 논문.
- 김영천(2002), “미국 수행평가 연구 방향에 대한 분석적 고찰: 우리에게 주는 시사점”, 열린교육 연구 10-1, pp.131~162.
- 김진곤·김혜영(2005), “고등학교 국어과 수행평가 실태 조사 연구”, 교과교육연구 26-1, 조선대학교.
- 김진숙(1999), “미국의 기준 운동에 비추어 본 한국의 절대평가기준 개발”, 교육과정 연구 17-2.
- 김혜정(2008), “고등학교 국어과 평가의 문제점과 개선 방안—평가 사례를 중심으로”, 국어교육학연구 32.
- 남명호(1995), 수행평가의 타당성 연구, 고려대 박사 논문.
- 남명호(2007), “다시 생각해 보는 수행평가”, 교육과정 운영에서의 수행평가의 적용 실태와 개선 방안, 한국교육과정평가원 연구자료 ORM 2007-8.
- 박인기(2000), “국어교육 평가의 패러다임 변화와 실천”, 국어교육 102.
- 박인기(2008), “국어과 평가의 반성과 전망”, 국어교육학연구 32.
- 백남진(2007), “수행평가 관련 지침과 중학교에서의 적용 현황”, 교육과정 운영에서의 수행평가의 적용 실태와 개선 방안, 한국교육과정평가원 연구자료 ORM 2007-8.
- 백순근(2002), 수행평가: 이론적 측면, 교육과학사.
- 서민원(2001), “수행평가의 학교현장 적용상의 문제와 장애요인 분석 연구”, 교육평가 연구 14-1, pp.47~80.
- 서지영·신명선·김소영(2008), 학교 교육 내실화를 위한 수행평가 개선 연구, 한국교육과정평가원 연구 자료.
- 유영희(2008), “평가 주체의 호환 작용에 의한 능동적 평가의 필요성—중학교 국어과

- 평가를 중심으로”, 국어교육학연구 32.
- 이경아(2005), 수행평가 정책의 평가 연구, 서울교대 석사 논문.
- 이수진(2008), “초등 국어과 수행평가의 현황과 과제”, 국어교육학연구 32.
- 이희숙(2004), 중학교 쓰기 영역의 수행평가 실태와 개선 방안, 홍익대 석사 논문.
- 정해승(2007), “성취 기준 중심 국어과 교육과정 구성에 대한 비판적 고찰”, 국어교육 123.
- 천경록(2001), 국어과 수행평가와 포트폴리오, 교육과학사.
- 최미숙(1999), “국어과 수행평가”, 열린교육연구 7-1.
- 최미숙(2006), “국어과 평가의 반성과 탐색”, 국어교육 121.
- 최진영(2007), “학교맥락, 교사 능력 및 의지, 교사학습기회 요인들이 초등교사의 사회
과 수행평가 실행에 미치는 영향력 분석”, 초등교육연구 20-3.
- 허경철 · 백순근 · 박경미 · 최미숙 · 양길석(1999), 수행정책 정책 시행 실태 분석과 개
선 대책 연구. 한국교육과정평가원. 연구보고 CRE 99-2.
- 허운나 · 김영옥(1998), 정보 시대와 미국의 교육혁명, 교육과학사.
- Allen, D. ed.(1998), *Assessing student learning : From Grading to understanding*, N.Y. :
Teacher College Press.
- Azwel, T. & Schmar, E. eds.(1995), *Report Card on Report Card : Alternating to Consider*,
Virginia:ASCD.
- Cuban, L.(1993), *How Teachers Taught : Constancy and Change in American Classrooms*
1890-1990, New York : Teachers College Press.
- Eisner, E. W.(1993), Why Standards may not Improve Schools, *Educational Leadership*,
50(5), 22-23.
- Ericsson, P. F.(2005), Raising the Standards for Standards : A Call for Definitions, *English*
Education, 37(3), 223-243.
- Firestone, W. A. et al(2001), *Pressure, Support and Instructional Change in the Context of a*
State Testing Program, Paper Presented at the American Educational
Research Association, Seattle, WA.
- Guskey, T. eds.(1996), *Communicating Student Learning*, Virginia : ASCD.
- Kohn, A.(1994), *Reporting What Student are Learning*, Educational Readership.
- Noddings, N.(1997), Thinking about Standards, *Phi Delta Kappan*, 79(3). pp.184-189.
- Tschudi, S. ed.(1997), *Alternatives to Grading Student Writing*, Illinois : National Council of
Teachers of English.
- Wilson, S.(1990), A Conflict of Interest : The Case of Mark Black, *Educational Evaluation*
and Policy Analysis, 12(3). pp.293~301.

<초록>

우리나라 수행평가 정책의 특징과 발전 방향 연구

—국어과 수행평가의 실행 양상의 원인에 대한 정책적 반추를 중심으로—

신명선

본고는 국어과 수행평가의 실행 양상에서 드러난 여러 문제점에 대한 원인을 정책을 중심으로 검토한 것이다. 수행평가 정책의 공식적·통시적 고찰을 바탕으로, 정책 개발의 논리를 분석하여 문제점을 지적하고, 바람직한 발전 방향 탐색의 통로를 찾고자 했다.

우리나라 수행평가 정책이 갖고 있는 문제점은 다음과 같이 요약할 수 있다. 첫째, 정책 도입의 초기에 정책 수단의 실현 가능성이 충분히 점검되지 못했으며, 정책 집행 단계에서 정책의 환경 변화에 보다 주도면밀하게 대처하지 못했다. 둘째, 수행평가 실행은 전통적인 학교 구조의 전면적인 변화를 요구하는 개혁임에도 불구하고, 우리의 경우 교육 제도나 교육과정 등에 대한 개혁 없이 교사의 책무성에 기반하여 단순히 평가 체제만을 바꾸려 했다. 양적 평가 및 객관적 평가를 중시하는 교육 전통, 성취 기준의 수행 기준화의 어려움, 교사의 평가 전문성에 대한 학생과 학부모의 불신, 학교의 행정적 여건, 교사의 수행평가에 대한 불철저한 이해, 입시 준비로서의 학교 운영의 실체 등과 같은 학교 내외적인 변수들을 교육 정책에 충분히 반영하지 못했다. 우리나라에서 수행평가가 제대로 실시되기 위해서는 근본적 변화를 담보하는 지원과 교육 구조의 개선이 필요하다.

【핵심어】 국어 교육, 국어 교육 평가, 평가, 양적 평가, 질적 평가, 수행평가, 수행평가 정책, 성취 기준, 평가 체제

<Abstract>

A Study on the Features and Reform Direction
of Performance Assessment Policy

—centering around the policy rumination about the cause of performance
assessment's practice aspect—

Sin, Myeong-seon

This paper intends to analyze the cause of performance assessment policy failure. The cause of policy failure is as follows. First, policy's realization possibility was not examined completely at the early stage of policy. And that policy have been practiced not voluntarily but defensively at the execution stage of policy. Secondly, that policy has been practiced depended on a teacher's sense of responsibility without the amendment of the education system and curriculum. It was not considered that the education tradition in Korean society such as the emphasis of quantitative and objective assessment, the disbelief of teacher's assessment ability, the abstract features of curriculum as a performance standards and the school as a educational organization for CSAT arrangement and so on will be influence on the policy. We have prepared the support system of teacher for a righteous practice of performance assessment policy. And It's also necessary to improve educational system.

【Key words】 Korean education, assessment, Performance assessment,
Performance assessment policy