교육에서 평가정책의 역사와 미래

– 표준화시험을 중심으로

2015. 1. 31.

연구책임자: 이윤미(홍익대학교)

교육을 바꾸는 사람들

목 차

Ⅰ. 서론 ····· 1

Ⅱ. 연구 요약: 표준화시험에 대한 역사적 재조명 – 한국교육에의 시사점 ····· 5

1. 서론 ····· 5
2. 미국에서 표준화시험의 역사적 전개 ····· 8
3. 주요 쟁점: 찬반 논쟁 ····· 19
4. 결론: 한국교육현실에 대한 시사점 ····· 42

Ⅲ. 미국에서 표준화시험의 형성과 전개 ····· 03

1. 공교육의 확대와 표준화시험의 형성 ····· 03
2. 계량심리학, 과학적 경영과 표준화시험 ····· 43
3. 표준화시험에 대한 반대와 재강화 ····· 63
4. 표준화시험과 책무성의 결합 ····· 83
5. 교육과정과 학생평가 패러다임 ····· 24
6. 평가일반과 평가학의 전개 ····· 61

Ⅳ. 미국에서 표준화시험을 둘러싼 논란: 옹호와 저항 ····· 84

1. The tyranny of testing(1962)– 표준화시험 비판의 고전 ····· 84
2. 찬반 양론의 대립 ····· 49
3. 비판 입장 ····· 53
4. (재)옹호 입장 ····· 58
5. 대안의 모색 ····· 59

Ⅴ. 표준화시험 적용의 국가별 사례: 프랑스, 독일, 일본 ····· 26

1. 표준화시험 적용의 맥락 ····· 26
2. 각 국별 사례 ····· 62
   가. 프랑스 ····· 62
   나. 독일 ····· 68
   다. 일본 ····· 73

Ⅵ. 주요 쟁점 및 정책적 시사점 ····· 77

1. 요약 ····· 78
2. 주요 쟁점 ····· 80
3. 한국 교육현실에 대한 시사점과 변화의 방향 ····· 38

부록1: 미국에서의 표준화시험 – 대학입시와 학력평가 ····· 98

참고문헌 ····· 98

표 목차

<표 IV-1> '표준화시험은 학습을 돕는가'에 대한 찬반 양론 ·····9 4
<표 IV-2> '표준화시험은 교사들에게 유익한가'에 대한 찬반 양론 ·····1···· 5
<표 IV-3> '표준화시험은 대학입학을 위해 가치가 있는가'에 대한 찬반 양론 ·····2····· 5
<표 V-1> 프랑스 초등 및 전기중등 최종 시험(모니터링 주기 평가) ·····6····· 6
<표 V-2> 프랑스 초등 및 전기중등 최종시험 (기초 역량) ·····7····· 6
<표 V-3> 프랑스 초등 및 중등 진단평가 ·····7·· 6
<표 V-4> 독일 전기중등 졸업자격 시험 ·····2·· 7
<표 V-5> 일본 전국 학력·학습상황조사 ·····6·· 7

그림 목차

<그림 II-1> 20세기 교육과정, 학습이론, 평가이론의 지배적 패러다임 변화 ·····3····· 1
<그림 III-1> 20세기초 미국 객관식 시험의 사례 ·····3···· 4
<그림 III-2> 20세기 교육과정, 학습이론, 평가이론의 지배적 패러다임 변화 ·····4····· 4
<그림 III-3> 20세기의 지배적 패러다임 ·····5··· 4
<그림 III-4> 새로운 구성주의적 패러다임 ·····5··· 4

Ⅰ. 서론

개혁이 어려운 이유 중의 하나는 많은 사람들이 오래된 이론들에 의존하기 때문이다. 평가는 객관적이고 공정해야 한다는 생각과 명시적이고 반복적인 시험이 학습에 유익하다는 행동주의적 가정은 많은 사람들의 상식 속에 강하게 자리잡고 있다. 이러한 상식은 역사적, 경험적으로 뿌리 깊게 내면화되어 있다.

이 연구에서는 교육을 넘어 거의 모든 공무와 사업 영역에서 인간 활동을 제어하는 장치(방안)로 채용되고 있는 평가의 역사에 대해 국내외 문헌을 통해 분석하고, 향후 방향을 제안하는 것에 목적을 두고 있다. 특히 관리주의(managerialism)가 팽배하게 된 역사적 과정과 이른바 신자유주의와 '평가 체제'가 결탁하게 되는 역사적 과정이 어떠한가에 대한 국내외 사례를 분석하고자 한다. 이 연구를 이끄는 주요 질문들은 아래와 같다.

- 평가의 역할과 비중: 교육 또는 학교 교육의 역사에서 평가가 차지해 온 역할이나 비중은 어떤 것인가? 어떤 배경 또는 논리로 평가는 오늘날과 같은 영향력을 지니도록 변화해 왔는가? 평가를 위한 방법이나 절차 등에 대한 생각은 어떤 역사적 변전을 거쳐 왔는가?
- 평가에 대한 수용과 비판의 담론: 평가에 대한 수용과 비판의 담론은 역사적으로 어떤 양상을 보여 왔는가? 평가의 가능성과 한계를 지적하게 되는 실지 근거들은 어떤 것이었는가? 평가의 오용과 남용 그리고 선용 등에 대한 역사적인 사례들은 어떤 것인가?
- 역사적 검토를 통한 시사점: 학교 교육과 관련된 평가의 발전을 도모할 때, 역사적인 정리를 통하여 얻을 수 있는 시사는 어떤 것인가?
- 주요 외국 국가의 경향: 오늘날 주요 외국들이 평가와 관련하여 취하고 있는 정책은 어떤 것들인가? 특히, 학교에서의 성취도 평가, 입학 전형, 교육의 질 관리, 교사나 학교 등에 대한 책무성 확인 등에 관련된 평가 정책은 어떤 것인가?
- 평가와 관련한 새로운 경향: 최근 세계적으로 주목 받는 평가의 새로운 경향에는 어떤 것들이 있는가? 주요 국가에서 추진하고 있는 평가 개혁 정책은 어떤 것들인가? 이 정책들은 어떻게 평가될 수 있으며, 우리나라 평가 발전에 어떤 시사를 주는가?

이상의 문제의식에 기초하여 이 연구에서는 오늘날 교육논의에서 핵심적 지위를 지니는 평가와 평가문화의 역사를 살펴봄으로써 각종 학문적, 정책적 논의의 기초를 제공하고 향후 방향을 설정하는데 도움이 되고자 한다. 교육에서 평가는 불가결한 하나의 과정이면서 그것이 교육의 전과정에 미치는 영향 때문에 주목된다. 한국에서 평가가 유난히 주목되는 이유는 선발 때문이라고 할 수 있으며, 상급학교 진학에서 요구되는 전형자료가 하급단계 학교들에서의 교육에 직접적으로 영향을 주기도 한다.

이 연구가 주로 다루고자 하는 초점은 표준화시험의 역사에 관한 것이다. 표준화는 20세기초 이래 과학의 사회적 적용과 함께 중시되어온 원리로서 산업경영, 행정분야에서 발달해왔고 교육에도 지대한 영향을 미쳤다.

시험의 역사를 볼 때 표준화시험의 대표적 사례는 중국에서 7세기 이후 시행된 과거시험인

것으로 알려져 있다. 과거시험과 유사한 사례는 다른 문명권에서는 시도되지 않은 것으로 나타나며 18세기말과 19세기에 들어 유럽에서 도입되기 시작한다. 과거시험이 관리선발시험이었던 점을 감안할 때, 이를 직접적으로 수용한 것은 19세기 영국의 관리선발시험이었던 것으로 알려진다. 중국의 공개경쟁형 과거시험에 대한 관심은 18세기 영국의 아담 스미스(Adam Smith)나 프랑스 사상가들(François Quesnay 등)에 의해 표명되기 시작했고 이는 직간접적으로 유럽의 제도에 영향을 준 것으로 분석된다(Teng, 1943). 직접적인 관료선발시험은 아니지만 관료선발의 기초로 도입된 시험들에는 독일의 아비투어(1788년)와 프랑스의 바깔로레아(1808년)와 같은 시험들이 있다. 이러한 시험들은 엘리트교육의 질을 관리하고 능력주의(메리토크라시)의 원리를 실현하는 목적과 관련되어 있다(Heafford, 1995; Atia, 2008).

중국의 과거시험은 평민이상이면 누구든 응시할 수 있는 개방적이고 공개적 시험으로 국가에 복무할 수 있는 우수한 재능을 가진 자를 가려내고자 하는 것이다. 이는 세습적 혈통중심으로 특권과 권력을 장악하는 방식과는 대비되며, 서구에서 시민혁명 이후에 도입하고자 한 능력주의의 이상에 부합하는 측면이 있다. 즉 객관적이고 공정하게 인재를 가려내는 방식인 것이다(Martin, 1870). 이러한 객관성, 공정성에 기초한 사회공학적 필요는 표준화된 시험에 대한 정당화와 밀접하게 결합되어 있다고 볼 수 있다.

20세기초 미국을 중심으로 대두된 과학적 경영주의(scientific management)는 교육과정, 교육평가 패러다임에 큰 영향을 주었다. 소위 교육과정에서의 IQ 검사, 과학적 측정방법론 발달, 교육행정에서의 체제이론 등은 이러한 경향을 대표한다. 표준화 검사와 시험은 능력주의(meritocracy) 원리와 결합하여 주관성을 배제하고 객관성과 공정성을 보장하는 장치로 부각되어왔다.

이러한 표준화의 한계는 학생의 다양성, 교육내용의 다원성, 심도있는 이해와 고급한 사고능력 등을 이끌어내지 못한다는 점과 교사와 학생의 상호작용으로 구성되는 교실내 활동을 획일화시킬 수 있다는 점에서 비판되어왔다. 다중지능이론으로 잘 알려진 하워드 가드너(Gardner, 1992)는 표준화시험에 대해 비판적 입장을 견지해왔는데, 가드너의 비판에서 두 가지 점을 주목할 수 있다. 첫째는 폭넓은 평가(assessment)가 아니라 제한적 시험의 형식(testing)을 취한다는 점이다. 가드너는 형식적 시험 모델(formal testing model)과 도제적 모델(apprenticeship model)을 구분하고, 객관적이고 탈맥락화한 형식을 취하는 시험(testing)이 도제적 모델보다 지배적 영향력을 갖게 된 점을 비판한다. 시험은 도제모델에 비해 학습자의 특성에 대해 맥락내적이고 전면적(인격적, 주관적)으로 파악하지 못한다는 것이다(Gardner, 1992: 78). 둘째는 이러한 형식적 시험을 강조하게 된 것이 학교교육에 대한 획일적 관점(uniform view of schooling)때문이라고 본다. 즉 모든 학생에 대해 동일한 교육내용을 부과하고 모든 학생에게 그것이 동일한 의미를 가질 것이라고 가정하는 것은 문제라는 것이다. 이 과정에서 형식적 시험을 통해 평가되기 어려운 교과영역(예술 등)은 학교교육에서 평가절하되는 경향이 있다(Gardner, 1992: 80-81).

가드너는 인간 지능이 다양함에도 불구하고 시험은 이 중 일부 능력(논리수학적)만을 중시한다고 비판한다. 따라서 개개인의 지능이 공정하게 다루어져야 하며(intelligence-fair), 학습자가 처한 맥락이 중시되는 개인중심 접근이 대안으로 모색되어야 한다고 주장한다(Gardner, 1992:

85-86). 이처럼 표준화시험은 교수-학습의 맥락(context)과 거리가 있다는 점에서 비판되어 왔고, 변별을 위한 활용의 목적이 강할 뿐, 학습자 개인의 발달에는 긍정적인 영향을 주지 못한다는 점에서 교육적이지 못하다는 지적을 받아왔다.

이 연구에서는 표준화시험의 형성과 전개를 중심으로 하여 교육에 대한 거버넌스, 교육과정 및 학생평가, 대학선발기준, 책무성평가 등이 역사적으로 어떻게 서로 관련되어왔는지를 검토한다. 이를 통해 역사적 전개과정이 평가정책의 방향 정립에 주는 시사점을 탐색하고자 한다. 이 연구의 주된 대상은 표준화체제의 형성과 발달이 가장 두드러지게 나타난 미국을 중심 사례로 주목하고 있으며, 프랑스, 독일, 일본 등 다른 사례들도 참고적으로 검토하고 있다. 이 연구는 주요 쟁점과 관련한 자료와 문헌들을 수집·분석함과 동시에, 평가의 역사와 관련하여 쟁점을 제기해온 주요 연구서들을 재검토하고 이들 논의를 비판적으로 확장·심화하는 방식으로 다루고 있다.

몇 가지 선행연구들은 이 연구의 문제의식에 중요한 통찰을 제공한다. 특히 능력주의의 등장이 지니는 문명사적 의미를 재구성한 영(Michael Young)의 <메리토크라시의 발흥(The rise of the meritocracy)>, 미국 SAT시험의 역사가 지니는 사회적 의미를 다룬 리만(Nicholas Lemann)의 <더 빅테스트(The big test: the secret history of the American meritocracy)>, 미국의 전통 명문대학들에서 20세기초에 나타난 선발 전형 변화의 이면을 다룬 캐러벨(Jerome Karabel)의 <선택된 자들(The chosen: the hidden history of admission and exclusion at Harvard, Yale and Princeton)> 등은 능력주의가 과학주의와 어떻게 결합되며, 사회적, 정치적 의도를 실현하는데 어떻게 기여하는지에 대해 비판적으로 조명하고 있어 평가가 지배적 사회문화와 연계된 방식을 드러내준다.

한편, 한국의 사례로 적용하여 논의하기 위해서는 학력주의 및 능력주의라는 근대 '보편적' 원리들이 동아시아적 '특수' 맥락과 접맥되어 동아시아적 교육문화를 이루어왔음을 주목해야 할 것이다. 타케우치요(竹内洋)의 『일본의 메리토크라시(日本のメリトクラシー)』는 일본의 사회시스템 속에 메리토크라시라는 보편적인 개념장치가 어떻게 기능하는지 보여주는 연구로 동아시아적 특수성과 한국 교육문화분석에 의미있는 시사점을 주며, 젱(Kangmin Zeng)의 <드레곤 게이트(Dragon gate: competitive examinations and their consequences)>도 대학입학시험을 중심으로 일본, 한국, 대만을 비교 역사적으로 조명함으로써 동아시아적 유사성에 접근하고 있다. 동아시아국가들에서는 과거제도와 연계된 유교문화의 영향으로 전통적으로 시험을 중시해왔고 이는 근대이후 능력주의와 연관되어 연속성을 보여왔다. 동아시아적 유사성 및 특수성은 한국적 조건을 논의할 때 매우 중요한 차원이기 때문에 비교 역사적으로 분석할 필요가 있다.

이 연구에서는 선행연구들에서 제기된 문제들을 중심으로 표준화시험이 학력평가, 책무성평가, 대학선발 등과 관련되어 온 방식을 검토하고 현대 평가문화가 지니는 특징들을 논의한다. 이를 통해 교육정책 논의에 주는 시사점을 제시하고자 한다. 이 연구는 다음과 같은 의의가 있다. 첫째, 평가정책을 역사적 맥락에서 검토함으로써 주요 쟁점에 대한 이해를 제고할 수 있다. 둘째, 국내외 사례에 대한 비교 분석을 통해 한국적 특수성과 보편성을 확인할 수 있다. 셋째, 역사적 검토를 통해 평가정책에 대한 향후 방향 정립의 기초 자료로 활용할 수 있다.

이 보고서의 체제는 다음과 같이 구성되어 있다. 제2장은 연구 요약에 해당하는 내용으로, 표준화시험의 등장, 발달, 주요 논쟁, 한국적 시사점 등을 다루고 있다. 이 장은 연구 전체에 대한 요약이라기보다는 보고서 전체 내용을 일정하게 포괄하면서도 표준화시험에 대한 역사적 재조명과 시사점을 다룬 독자적인 논문의 형식으로 구성되어 있다. 이 보고서의 내용을 요약적으로 담고 있는 별도의 글로 재구성한 것으로, 표준화된 평가산업이 가장 발달한 미국의 사례를 중심으로 다루면서 한국적 시사점을 논의하는 방식으로 기술되어 있다. 표준화시험과 관련하여 이 장에서 충분히 상술되지 않은 논의들이나 사례는 제3-5장에서 보다 구체적으로 다루고 있다.

제3장과 제4장은 미국에서의 표준화시험 발달과정과 주요 논쟁점에 대해 다루고 있다. 이 장들에서는 표준화시험의 역사와 현재 논쟁의 이해를 위해 기본이 되는 주요 문헌들에 기초하여 역사적 흐름과 현재적 쟁점들을 정리하였다.

제5장은 표준화시험의 다양한 적용을 살펴보기 위해 프랑스, 독일, 일본의 사례를 검토했다. 이들 나라에서 대학입학시의 자격·선발시험으로 표준화시험이 기능하는 방식과 최근의 학력관리 정책의 일환으로 시행되고 있는 국가수준학업성취도평가 운영 실태를 포함하고 있다.

제6장은 이 연구에서 종합적으로 도출할 수 있는 쟁점들을 확인하고 한국 평가정책의 방향 설정에서 참고할 사항들을 시사점으로 논의하고 있다.

부록에는 미국의 주요 표준화시험과 관련한 정보들을 정리하여 수록했다.

Ⅱ. 연구 요약:

표준화시험에 대한 역사적 재조명
– 한국교육에의 시사점

1. 서론

한국 교육문제에 대한 논의에서 평가는 매우 중요한 의미를 지니고 있다. 한국교육은 흔히 '암기중심 입시위주교육'으로 특징지워지고 비판되어온 경향이 있다. 상급학교 입시가 학교교육에 영향을 미쳐왔고, 학교교육도 표준화된 시험에 의해 지배되어왔다. 표준화된 시험은 제한적인 능력을 측정하고 교육과정을 협소화하며 깊이있는 학습과 성찰을 유도하지 못한다는 점에서 비판되어 왔다. 그럼에도 불구하고 객관성과 공정성을 중시하는 평가와 선발문화로 인해 표준화시험의 중요성은 꾸준히 인정되고 있는 것이 현실이다.

표준화시험(standardized testing)은 표준화된 방식으로 실시되고 채점되는 평가형태를 말한다. 표준화되었다는 것은 이미 정해져 있고, 누구에게든 동일한 문제가 주어지며 동일한 시간동안 해결해야 하고 동일한 기준으로 채점된다는 것을 포함한다. 즉 평가대상이 일관되고, 동일(획일)하며 객관적으로 채점된다는 특징이 있는 것이다. 이러한 특징을 가진 표준화시험은 교수-학습의 관점에서 볼 때는 외부적 평가라는 특징이 있다. 즉 교수-학습이 실제로 이루어진 상황이나 맥락의 특수성과 무관하게 출제와 채점이 이루어지게 된다는 것이다.

표준화시험은 몇 가지 점에서 문제점이 지적되고 대안에 대한 논의들이 이루어져 왔다. 다중지능이론으로 잘 알려진 하워드 가드너(Gardner, 1992)는 표준화시험에 대해 비판적 입장을 견지해왔다. 가드너의 비판에서 두 가지 점이 특히 주목된다. 첫째는 폭넓은 평가(assessment)가 아니라 제한적 시험의 형식(testing)을 취한다는 점이다. 가드너는 형식적 시험 모델(formal testing model)과 도제적 모델(apprenticeship model)을 구분하고, 객관적이고 탈맥락화한 형식을 취하는 시험(testing)이 도제적 모델보다 지배적 영향력을 갖게 된 점을 비판한다. 시험은 도제모델에 비해 학습자의 특성에 대해 맥락내적이고 전면적(인격적, 주관적)으로 파악하지 못한다는 것이다(Gardner, 1992: 78). 둘째는 이러한 형식적 시험을 강조하게 된 것이 학교교육에 대한 획일적 관점(uniform view of schooling)때문이라고 본다. 즉 모든 학생에 대해 동일한 교육내용을 부과하고 모든 학생에게 그것이 동일한 의미를 가질 것이라고 가정하는 것에 문제가 있다는 것이다. 이러한 관점에서는 질보다는 양적인 평가가 이루어지기 때문에 대단위 표준화를 하게 되고 비교를 중시하게 된다. 이 과정에서 형식적 시험을 통해 평가되기 어려운 교과영역(예술 등)은 학교교육에서 평가절하되는 경향이 있다(Gardner, 1992: 80-81).

가드너는 인간 지능이 다양함에도 불구하고 시험은 이 중 일부 능력(논리수학적)만을 중시한다고 비판한다. 따라서 개개인의 지능이 공정하게 다루어져야 하며(intelligence-fair), 학습자가 처한 맥락이 중시되는 개인중심 접근이 대안으로 모색되어야 한다고 주장한다(Gardner, 1992: 85-86). 이처럼 표준화시험은 교수-학습의 맥락(context)과 거리가 있다는 점에서 비판되어 왔

고, 변별을 위한 활용의 목적이 강할 뿐, 학습자 개인의 발달에는 긍정적인 영향을 주지 못한다는 점에서 교육적이지 못하다는 지적을 받아왔다.

국제적으로 볼 때, 1980-90년대에 들어서는 표준화시험이 학교교육의 책무성 점검과 연결됨으로써 일부 시험들이 교수-학습지원보다는 학교에 대한 통제와 모니터링의 목적과 강하게 연계되는 현상이 심화되었다. 이는 평가관리문화의 국제적 확산으로 인해 전세계적으로 나타난 양상이다.

교수-학습활동 및 학교운영에서 평가의 비중이 강화되면서, 표준화시험에 기반한 평가문화가 교육문화에 주는 영향은 높아지고 있다. 표준화시험은 그 효용성으로 인해 널리 활용되고 있지만, 표준화시험의 내용(표준화 정도)과 형식(문항출제방식, 대단위시험 여부, 채점의 객관성 등), 활용방식(고부담 여부)등이 어떠한가에 따라 교육문화에 해악적인 영향을 미치기도 하는 것이다.

우리나라의 경우에도 대학수학능력시험을 비롯한 학교 외부의 표준화된 각종 학업성취도평가뿐 아니라 학교 내부의 표준화된 시험(획일적으로 실시되는 중간고사, 기말시험 등) 등이 실시되고 있다. 이러한 시험들이 교수-학습과정의 맥락성을 고려하지 못하고 교사의 자율성, 전문성을 저해한다는 지적을 받아온 만큼 표준화시험에 대한 비판적 재검토가 필요하다고 하겠다.

이 연구에서는 표준화시험의 형성과 전개과정을 역사적으로 살펴보면서 관련 쟁점들과 대안들이 논의된 방식에 대해 검토한다. 특히 20세기 들어 표준화시험의 실시와 활용이 가장 활발하게 이루어진 미국의 사례를 중심으로 살펴보면서 미국에서의 전개과정과 현황이 우리 교육에 주는 시사점이 무엇인지에 대해 접근하는 형식을 취하고자 한다.

시험의 역사를 볼 때 표준화시험의 대표적 사례는 중국에서 7세기 이후 시행된 과거시험인 것으로 알려져 있다. 과거시험과 유사한 사례는 다른 문명권에서는 시도되지 않은 것으로 나타나며 18세기말과 19세기에 들어 유럽에서 도입되기 시작한다. 과거시험이 관리선발시험이었던 점을 감안할 때, 이를 직접적으로 수용한 것은 19세기 영국의 관리선발시험이었던 것으로 알려진다. 중국의 공개경쟁형 과거시험에 대한 관심은 18세기 영국의 아담 스미스(Adam Smith)나 프랑스 사상가들(Quesnay 등)에 의해 표명되기 시작했고 이는 직간접적으로 유럽의 제도에 영향을 준 것으로 분석된다(Teng, Ssu-yu, 1943). 직접적인 관료선발시험은 아니지만 관료선발의 기초로 도입된 시험들에는 독일의 아비투어(1788년)와 프랑스의 바깔로레아(1808년)와 같은 시험들이 있다. 이러한 시험들은 엘리트교육의 질을 관리하고 능력주의(메리토크라시)의 원리를 실현하는 목적과 관련되어 있다.[1)]

중국의 과거시험은 평민이상이면 누구든 응시할 수 있는 개방적이고 공개적 시험으로 국가에 복무할 수 있는 우수한 재능을 가진 자를 가려내고자 하는 것이다. 이는 세습적 혈통중심으로 특권과 권력을 장악하는 방식과는 대비되며, 서구에서 시민혁명 이후에 도입하고자 한 능력주의의 이상에 부합하는 측면이 있다. 즉 객관적이고 공정하게 인재를 가려내는 방식인 것이다.[2)]

1) 프랑스학자인 브뤼네티에르는 프랑스의 교육제도와 시험에 의한 경쟁적 선발제도가 중국체제에서 영향을 받은 것이라고 본다. Brunetière, F., "L'Orient dans la littérature française" in Études critiques sur l'histoire de la littérature française, 8e serie, Paris, 1910: 197-201; Teng, 1943: 302 재인용.

2) 이러한 분석들은 19세기 근대화과정에서 중국정체의 원인을 과거시험에서 찾고자 하는 관점들에 비추어 볼 때 아이러니한 논의이지만, 비교교육사적 관점에서 매우 흥미로운 주제이다. 1870년대에 북경 Tungwen College의 총장으로 있던 마틴(Martin)은 중국 과거시험의 장점을 다음과 같이 정리하여 소개한 바 있다. 첫째 공개경쟁이기

이러한 객관성, 공정성에 기초한 사회공학적 필요는 표준화된 시험에 대한 정당화와 밀접하게 결합되어 있다고 볼 수 있다. 한편, 객관성과 공정성의 이면에는 지식의 표준화에 따른 비용의 문제가 있다. 표준화시험이 전면화되면 표준적 지식기준으로 인해 확산적이고 깊이 있는 사고를 저해할 수 있으며 획일적 기준에 의한 결과 판정 문제(석차, 수직 서열화 등)가 논란이 될 수 있다. 물론 시험의 내용과 형식, 결과활용 방식 등에 따라 사회마다, 혹은 문화마다 표준화시험의 효과는 다르게 나타날 수 있다.

미국에서는 19세기 중반 표준화시험이 처음 도입된 이래 20세기 초에는 선발과 변별의 목적으로 적극 활용해왔다. 특히 프랑스의 비네(Binet)가 개발한 지능검사(IQ)와 그 원리를 군대와 학교에 적용하기 시작하면서 표준화시험은 빠르게 확산되었으며, 표준화시험을 개발하는 민간 평가산업도 크게 발전하는 등 그 영향력이 커졌다. 표준화시험의 발달은 다른 한편으로는 1960년대 이후 표준화반대운동이 거세게 전개되는 계기가 되기도 했다. 그러나 1980년대 이후에는 교육책무성을 강화하는 신자유주의적 흐름으로 인해 표준화시험의 중요성은 더욱 강조되고 있는 추세이다. 특히 표준화시험은 소위 '낙오자방지법(NCLB)'과 같은 고부담책무성평가와 연결되면서, 그 찬반에 대한 양론은 팽팽하게 맞서고 있는 상황이다. 더욱이 영미를 중심으로 발달된 국가수준의 학업성취도평가는 1990년대 이후 유럽연합에서도 도입되고 확대되는 추세이기도 하다.

이 연구에서는 이러한 흐름을 역사적으로 살펴보고, 이러한 흐름이 나타나게 된 배경이 무엇이며, 교육적으로 볼 때 논란이 되는 지점이 무엇이고, 이를 해결하기 위한 대안적 논의가 무엇인지 검토하고자 한다. 그리고, 이를 기초로 하여 한국교육현실에서의 시사점을 논의하고자 한다. 미국은 계량심리학과 측정기술 및 문화의 발달로 표준화시험 산업이 가장 발달한 나라라고 볼 수 있기 때문에 그 역사가 드러내 주는 쟁점이 많다. 흥미로운 것은 표준화시험 자체는 미국보다 유럽에서 먼저 시작되었다는 점이고, 형식적 시험의 역사는 중국의 과거시험이 가장 오래된 것으로 간주된다는 점이다. 한국 교육은 해방 이후 계량심리학의 발달로 미국식 표준화시험의 영향을 직간접적으로 받아왔지만, 보다 심층적 문화의 차원에서는 10세기 가량 지속된 과거시험의 역사와 일제시기 상급학교 입학시험, 고등문관시험 등과 같은 시험제도들에 영향을 받은 바 크다. 특히 평가가 변별과 선발의 함의를 강하게 가지고 있고, 객관성과 공정성을 우선시하는 평가문화를 유지하고 있는데 그 역사적 뿌리는 깊다고 하겠다.

이 연구는 미국 표준화시험의 역사 및 논쟁과 관련한 주요 문헌들을 중심으로 논의를 검토하고 시사점을 도출하고자 했다. 이 연구를 이끄는 주요 연구문제는 다음과 같다. 첫째, 미국에서 표준화시험이 발달한 배경은 무엇인가? (측정문화발달, 국가수준의 표준필요, 평가산업의 발달) 둘째, 1960년대 이후 표준화시험에 대한 반대와 옹호의 양상은 어떻게 나타났는가? (비판의 역사) 셋째, 찬반 논란에도 불구하고 표준화시험이 유지, 강화되는 맥락은 무엇인가? (옹호자들의 재옹호 논의) 넷째, 표준화시험의 지배에 따른 교육적 문제들을 해결하기 위한 방안은 무엇인가.

때문에 국가에게는 안전장치가 될 수 있다. 자칫하면 불만세력이 될 수도 있는 사람들이 국가시험을 준비하기 위해 순치된다. 둘째 혈통에 의한 귀족세력을 견제하고 국가(국왕)의 권력을 강화할 수 있다. 낮은 사회적 배경을 가진 자들도 재능이 있으면 선발될 수 있고, 그렇기 때문에 인민대중의 필요에 대해 더 잘 알 수 있다. 셋째 국가는 교육받은 층이 제도유지에 복무하도록 할 수 있다. 혁명적 상황으로 인해 제도가 붕괴되는 것을 원하지 않게 된다(Martin, 1870: 75-76). 과거제도가 지닌 객관성 및 공정성, 지식인층에 대한 인센티브, 국가권력과 사회공학적 효용성 등이 잘 나타나 있다.

2. 미국에서 표준화시험의 역사적 전개

가. 미국에서 표준화시험의 시작

미국은 다른 나라에 비해 표준화시험이 체계적으로 발달해온 국가로 알려져 있다. 미국에서 많이 통용되는 표준화시험들에는 SAT, ACT, GRE 시험과 같이 대학 및 대학원 입학에 활용되는 시험도 있고, 학력진단 배치 등을 목적으로 실시되는 시험들도 있다. 이러한 시험들은 민간 평가기구들에서 제작되어 보급되고 있는 것이 특징이다.[3] 대표적인 표준화시험인 SAT, GRE 등은 지능검사인 IQ 시험의 원리에 기초하여 시작되었는데 20세기초에 이미 실시되기 시작했다. 미국에서는 1950년대까지 표준화시험의 활용이 교육에서 매우 일반화되었던 것으로 알려져 있고, 1960년대 이후에는 표준화시험에 대한 찬반 논란이 활발히 이루어져 왔다.

표준화시험은 흔히 표준화검사로 번역되기도 하는데, 이를 대표하는 것은 프랑스의 비네에 의해 개발된 지능검사이다. 미국에서는 이를 제1차 대전 전후에 군대와 학교에서 활발히 적용하게 된다. 이러한 이유로 표준화시험은 측정기술이 고도로 발전하는 20세기 이후의 산물로 인식되기도 하지만, 표준화시험의 실시 자체는 19세기초중반으로 거슬러 올라갈 수 있다.

미국에서 실시된 최초의 표준화시험은 1845년 매사추세츠 보스톤에서 중등학교를 대상으로 한 시험이었던 것으로 알려져 있다(Reese, 2013). 이 시험은 당시 매사추세츠 교육위원회의 비서였던 호레이스 만(Horace Mann)의 주도로 실시되었고, 전통적인 문법교육을 실시하는 중등학교를 대상으로 하는 것이었다. 호레이스 만은 교육계밖에 있던 인사들을 평가위원회에 임명하여 보스톤 시내 학교들에 대한 학력평가를 실시하도록 했고, 이 결과를 출간하여 전국적으로 보고한 바 있다. 이 시험의 결과로 당시 중등학교에서 이루어지던 교육의 실상이 대외적으로 알려지게 되는 계기가 되었고, 보스톤 지역의 학력이 낮게 나타나 교육의 질 개선을 위한 다양한 조치들이 뒤따르게 했다.

이 시험은 미국 교육의 역사에서 몇 가지 중요한 의미를 지닌다고 본다. 첫째는 최초의 표준화된 학력평가라는 의미가 있다. 이 시험이 도입되기 이전 미국에서 일반적으로 이루어지던 평가의 형태는 구술평가이거나 일년 한 두 번씩 지역주민을 대상으로 한 발표회의 형식이었던 것으로 알려져 있다. 특히 중등교육은 소위 마스터(master)로 불리는 엘리트 중등교사들에 의해 이루어졌으며 개별 학교단위에서 이루어졌을 뿐 지역 수준에서 학력을 진단하는 것은 없었다. 이 시험은 보스톤 지역 학교들을 대상으로 하는 최초의 대단위 평가로 표준화된 문항의 출제와 채점이 평가위원회에서 이루어졌다. 이 시험 이후 표준화시험은 미국에서 점진적으로 확대되기 시작한 것으로 알려진다. 특히 1870년대에 매사추세츠 퀸시(Quincy) 지역에서 진보적 교육운동을 이끌었던 파커(Parker)에 의하면 미국 학교들은 시험에 열광하고 있다고 비판될 정도로 지필고사형태의 시험은 많은 학교들에서 보편화되었다.

둘째는 당시 실시된 시험이 교수학습의 직접적 개선을 위한 목적 외에 교육의 질을 행정적으

3) Riverside publishing(Iowa Tests of Basic Skills), Educational Testing Service(SAT; Graduate Record Examinations, GRE), Scholastic Testing Service(Highschool Placement Test, HSPT), American College Testing(American College Test, ACT) 등이 대표적이다. Dolezaler, 2009: 43.

로 모니터링하는 기능을 겸하고 있다는 점이다. 특히 시험의 출제, 채점, 분석을 담당한 평가위원회 소속의 인사들은 교육계밖 타 분야 전문가들이었다. 외부에 의한 평가를 통해 학교의 실상을 드러내고자 했다는 것이다. 이는 교육계 내부의 맥락이나 당사자들과는 다른 관점에서 이루어졌다. 이 시험의 결과로 보스톤 지역 학교들에서 교육이 형편없이 이루어지고 있다는 점이 밝혀짐으로써 교육의 질에 대한 논란이 촉발되었다. 특히 이를 계기로 보스톤지역의 교육을 감독 지원할 교육장직이 만들어진 점도 주목할 만하다. 당시의 시험이 외부 표준화시험에 의해 교육의 질 관리를 하는 책무성평가의 기능을 하고 있었음을 알 수 있는 것이다.

셋째는 교육 운영상의 주도권이 기존의 엘리트주의적 정교사(master)들로부터 행정기관으로 옮겨지는 데 기여했다는 점이다. 특히 이 평가로 교육계 내부의 반발이 매우 심했던 점을 주목할 수 있다. 당시 중등학교 정교사협의회에서는 호레이스 만의 개혁시도에 대해 강력하게 반발했다. 특히 현장교육의 사정을 모르는 자들에 의한 것이라는 점이 주된 비판의 논지였다. 그러나 결국 이 평가 이후 정교사협의회를 주도하던 핵심 교사들의 입지는 약화되었고, 이들이 대표하고 있던 전통적 문법교육의 내용과 형식이 지닌 권위도 쇠퇴하게 되었다. 정교사협의회가 전통적 교육의 옹호자들이었다면, 표준화시험의 실시를 주도한 세력은 보다 체계적인 개혁을 추진하고자 하는 의도를 지니고 있었다. 호레이스 만 등은 유럽교육에 대한 시찰 경험에 기반하여 당시 프러시아 교육의 개혁에 감화를 받고 있었다. 이들은 미국의 교육현장에 개혁이 필요하다고 강하게 믿고 있었다. 표준화시험의 실시 이전부터 호레이스 만 등은 보고서를 통해 개혁의 필요를 주장하고 있어 정교사협의회로부터 강한 반발에 부딪치고 있었다. 호레이스 만의 입장은 당시 개혁을 추진하고자 했던 휘그당의 입장을 대변하는 것이기도 했다. 결국 시험의 결과가 공개된 이후 전통적 문법교육의 권위는 객관적 데이터를 통해 도전받게 되었고 정교사협의회의 영향력은 크게 약화되었다.

넷째는 측정문화의 발달을 반영하고 있다는 것이다. 특기할 것은 호레이스 만이 임명한 평가위원회 인사 중 통계 전문가들이 포함되어 있다는 점이다. 미국은 실증적 데이터를 통해 효과성을 확인하는 프래그머티즘적 문화가 발달되어 있었다. 이러한 특징은 1840년대의 평가결과를 분석하고 해석하는데에도 반영되었다. 20세기에 비할 때 통계학의 발달 수준은 높지 않지만 당시 위원회에 소속된 전문가들 중 통계전문가가 포함되었던 점은 주목된다. 인간의 능력에 대한 측정 욕구는 20세기초 지능검사의 적용과정에서 매우 강렬하게 표출되며 미국에서 표준화시험의 발달과 밀접하게 연계된다.

이상의 특징들은 1840년대에 실시된 표준화시험이 단순히 일회적 사안이 아니었음을 알게 하며 이후 미국교육 및 평가활동의 전개과정에서 중요한 전환점이 되었음을 알게 한다. 흥미로운 것은 이러한 표준화시험이 미국에서 유독 빨리 시작된 것이 아니라는 점이다. 유럽에서도 표준화시험의 역사는 18세기말, 19세기초로 거슬러 올라간다. 독일의 중등졸업시험이자 대학입학시험인 아비투어시험[4]이나 프랑스의 바깔로레아[5]가 그것이다. 아비투어나 바깔로레아의 경

4) 아비투어는 독일에서 1788년에 처음으로 실시된 것으로 알려져 있다. 아비투어 출현과정에서는 학교체제와 대학의 개혁 필요가 핵심적이었다. 교육내부의 연계성이 부족하고 독일대학의 강의 질 등이 낮아지고 있다는 비판이 있었고, 대학의 특권화로 학생에 대한 군입대가 면제되자 질 낮은 학생들이 입학한다는 문제가 있었다. 이에 대응하면서 유능한 관료충원을 목적으로 교육개혁을 단행하게 되는데 이를 위해 중등문법학교의 교육 개선을 위한 국가기구가 설치되었다(Oberschulkollegium). 아비투어는 중등문법학교와 대학입학의 질 관리를 위한 기준을 국가가 정하고 관리하기 위한 시험이었다. 아비투어의 역사에 대해서는 Heafford, 1995 참조.

우 현재까지도 존속하는 시험으로 중등교육과 대학입시를 연계하는 중요한 시험이다. 이 시험이 고등학교 교육과정과 밀접하게 연계되어 있고 출제나 채점과정에서 고등학교 교사들의 참여가 높다는 점으로 인해 대규모 선다형시험 형식이 기본인 미국식 표준화시험과는 다르게 간주되기도 한다. 그러나 이들 시험도 표준에 기반하여 모든 학생에게 동일하게 적용되는 시험이라는 점에서 표준화시험이라고 간주된다. 18세기말과 19세기에 출현한 이러한 시험들의 공통점은 교수학습이 이루어지는 맥락에서 이루어지는 시험이 아닌 일정한 표준에 근거한 외부 평가라는 점이다. 그리고 외부평가의 목적이 교수학습의 질관리에 있다는 것이다. 그러나 문제의 출제 및 채점 방식, 결과활용 방식 등에 따라 시험의 효과는 매우 달라질 수 있다. 미국에서 실시된 표준화시험들은 학력진단과 선발 배치 등과 관련이 높았으며, 공공기관보다는 민간기구들에 의해 개발되고 시행되었다는 특징이 있다.

나. 표준화시험의 확대

□ 사회공학적 관심

보스톤에서 실시된 최초의 표준화시험은 시험형식 상의 큰 변화였을 뿐 아니라 사회통합을 위한 사회적 공학의 발달을 상징하는 것이기도 하다(Garrison, 2009). 호레이스 만은 표준화시험을 통해 공립학교의 실패를 드러내어 새로운 시민형성을 위한 공통의 기준을 도입하고자 했다. 캔델(Kandel)의 분석에 의하면 호레이스 만은 표준화시험의 의의를 다음과 같이 보고 있다. 우선, 모든 학교와 학생에게 균등하게 실시되며, (구술시험에 비해) 1시간동안 충분한 시간을 사용할 수 있어 공정하다. 넓은 범위와 충분한 문항수를 확보할 수 있어 철저한 평가를 할 수 있으며, 교사의 개입이 제한되고 편견이나 선호를 방지할 수 있다. 학생들이 신뢰할만하고 유능하게 가르쳐졌는지 확인이 가능하고, 성적표가 영구적 기록으로 남아 학교간 비교나 개선 상황의 확인이 가능하다(Kandel, 1936; Garrison, 2009 재인용).

이러한 점들은 당시 도입된 표준화시험과 기존의 구술시험 방식을 비교해서 도출된 특징이다. 표준화시험이 객관적이라는 점과 함께, 균등하고 공정한 것이라는 관점은 주목할 만하다. 획일적인 것이 곧 공정한 것이라는 관점은 그 이후에도 지속적으로 영향을 주고 있다. 이러한 객관성, 공정성에 대한 강조는 '표준'이 지니는 외관적 중립성에 의문을 품는 입장에서는 비판의 표적이 된다.

당시로서는 표준을 설정하는 자들이 휘그당 즉 공화주의자들이었다. 이러한 표준화시험은 '(획일적) 균등한 적용 = 공정성'이라는 관점에서 정당화되어 왔고, 이는 미국적 메리토크라

5) 바깔로레아의 경우 나폴레옹 집권시기인 1808년 3월 17일 제국칙령에 의해 실시되기 시작했다. 프랑스대혁명 시기부터 국가의 교육적 책임에 대한 논의는 활발히 이루어지기 시작했는데 나폴레옹시기에 이르러 정부의 제도로 확고히 하고자 했고, 교수 학습 평가를 동일한 수준으로 국가전체에 적용하고자 했다. 바깔로레아는 국가주도의 표준화시험으로 출발했고 봉건적 귀속성에 영향을 받지 않는 능력주의를 표방했다. 1830년대 이후 필기시험이 도입되고, 1902년에는 고전 대신 과학교과와 현대어가 주요 과목이 되었다. 대학진학을 위한 일반 바깔로레아 외에 1968년에 기술학교 진학을 위한 기술바깔로레아, 1985년에는 직업분야 진출을 위한 전문 바깔로레아가 도입되어 학생의 특강점에 따라 서로 다른 진로 프로그램에 배치하고 있다. 바깔로레아의 역사에 대해서는 Atia, 2008 참조.

시 이념과 맞물려 사회공학적 차원에서 명분을 유지해온 것이라 볼 수 있다.

1840-50년대 이후 남북전쟁시기까지 미국 동북부에서는 표준화시험이 상당히 보편적으로 실시된 것으로 나타나며, 19세기말까지 미국에는 다양한 종류의 시험들이 학교에서 사용되고 있었다.[6] 표준화시험의 보급은 교육계 외부로부터의 개혁요구와 관련이 있다. 예컨대 19세기말 언론을 통해 미국 교육을 비판하여 큰 반향을 일으킨 조세프 메이어 라이스(Joseph Mayer Rice)도 당시 미국의 시험운동 역사에서 기억되는 인물이었다(Cremin, 1964). 미국 전역 36개 도시 학교에서 1,200명의 교사를 면담한 후 라이스는 교육에 대한 대중적 무관심, 정치적 개입, 부패와 무능력을 폭로했다. 교육계는 이에 대해 교육현실을 모른다고 비판을 가했다. 표준화시험의 실시와 확대의 이면에는 교육계 외부에서 교육을 개혁하려는 움직임과 교육계 내부 세력 간의 갈등이 자리잡고 있었음을 알 수 있는 것이다.

20세기 이후 표준화시험의 확산은 19세기말 이래의 다양한 사회문화적 변화를 반영한다. 첫째는 계량심리학의 발달로 인간의 자연적 능력을 재려고 하는 관심이 높았던 것과 관련된다. 이러한 관심은 미국적 메리토크라시 즉 공정한 기회배분원리에 대한 요구와 결합되는 측면이 있다. 둘째는 과학적 경영과 사회적 효율성 원리 등이 교육행정 및 교수원리에 영향을 미치게 되는 것과 관련이 있다. 셋째는 20세기 이후 교육기회의 확대 특히 중등교육의 확대로 학생들을 능력과 잠재성에 따라 분류, 배치하고자 하는 사회적 관심과 결합된 측면이 있다.

□ 계량심리학과 주요 표준화시험의 개발

19세기에서 20세기초로 넘어오는 과정에서 교육에 가장 크게 영향을 준 것이 계량심리학(pyschometrics)의 발달이라고 해도 과언이 아니다. 이러한 계량심리학의 발달은 지능검사의 발달, 지능검사와 유사한 원리의 학교 적용 등을 통해 '과학적 방법'으로 미국 교육계에 뿌리를 내리게 된다.

지능검사는 인간의 정신 능력을 재고 수치화함으로써 객관화하기 위한 시도이다. 18세기 계몽주의사상이 인간의 계몽가능성을 신뢰하며 프랑스대혁명 시기 공교육체제의 기획에 힘을 실어주었다면 대중교육이 확대되는 19세기후반에 와서는 인간의 교육불가능성(ineducability)에 대한 생각들이 나타나게 된다. 프랑스 제3공화국 하에서 교육이 불가능한 자들을 가려내려는 정책적 관심에서 탄생한 소위 지능(IQ)검사는 인간에 대한 분류를 위한 기제로 발달해왔다(Garrison, 2009). 프랑스에서 비네가 지능검사의 창시자로 알려져 있다면, 영국에서도 갈톤(Galton), 스피어맨(Spearman) 등이 이와 관련되며, 이는 우생학주의(Eugenics)에 영향을 주게 된다.

지능은 인간의 능력을 분류하는 기준이 되었을 뿐 아니라, 유사과학적인 객관성을 통해 엘리트들이 지배적 특권을 누릴 수 있도록 하는 근거로 작용하기도 했다. 지능은 성취가 아니라 잠재성을 의미하는 것으로 인간의 능력에 대해 자연적으로 이해하도록 한다. 이는 두 가지 의미를 갖는다. 한편으로는, 사회경제적 환경으로 인한 불평등을 자연적 능력을 통해 교정할 수 있

6) 1920년에 50만 종류의 시험들이 존재했으며, 1930년대 1,300종의 성취시험과 400여종의 심리검사가 있었다(Sacks, 2000: 70).

다. 즉 경제적 부의 여부에 의해 부풀려지거나 가려진 실제 능력을 드러낼 수 있다는 것이다. 다른 한편, 능력을 자연적인 것으로 보게 함으로써 인간에 대한 분류 자체와 능력의 차이에서 발생하는 불평등을 정당화한다. 무엇보다 IQ검사의 특징은 능력이 숫자로 표시된다는 점인데, 이러한 능력의 측정 지표, 측정 방법, 표기 방법 등은 지속적으로 논란이 되어 왔다.

인간의 능력을 단일한 숫자로 표시하여 분류하는 것이 미국에서 본격적으로 실용화한 것은 제1차대전 전후라고 할 수 있다. IQ검사지를 1905년에 처음 고안한 비네는 1911년에 사망했지만, 미국 스탠포드대학 교수 루이스 터먼(Lewis Terman)에 의해 비네-스탠포드 척도가 만들어지고 IQ(Intelligence Quotient, 정신연령 대 신체연령)라고 명명되었다. 미국에서 IQ검사는 군대에서의 직책구분을 위해 실시되었고, 이후 학교에도 적용되어 잠재적 능력을 재는 검사로서 미국적인 시험문화 형성에 기여했다.

미국에서 IQ검사의 원리는 학업성취의 예언을 위해 SAT, GRE 등의 표준화시험에 적용되었다. 프린스턴대학의 심리학자인 브릭험(Brigham)은 1차 대전동안 군대 IQ검사에 참여한 본인의 경험을 토대로 1926년 군대검사를 변환하여 SAT시험을 개발하였다. 그는 유제닉스에 관심을 갖고 미국인의 지능을 종족별로 구분하는 연구를 수행했다(A study of American intelligence). 흥미로운 것은 1932년 <오류에 대한 연구(A study of error)>에서 그의 전작을 비판하고 SAT나 IQ로부터 거리를 유지하기 시작했다는 것이다. 브릭험은 IQ검사에 대한 신뢰가 아무런 생물학적 근거도 갖지 않는 위험한 것이라고 보게 되었고, 이의 실용화에 대해 여러 차례 경고했다. 아이러니한 것은 SAT 개발자가 회의를 갖는 동안 SAT 산업은 지속적으로 확대되고 있었다는 것이다(Lemann, 2000).

한편 대학원입학 전형자료로 사용되는 GRE를 개발한 우드(Wood)의 경우는 회의적인 브릭험과 달리 표준화시험의 강한 지지자였다. 심리학자 손다이크(Thorndike)의 제자였던 우드는 학교교육이 비논리적이고 비조직적으로 이루어지는 점을 비판했다. 그는 국가수준의 시험체제를 통해 두 가지 목적을 달성하고자 했다. 첫째는 고등교육을 위한 가장 뛰어난 학생을 선발하는 것이었으며, 둘째는 교사들의 임의적 권력을 박탈하고자 하는 것이었다. 그는 이것이 학생들의 교과숙달도를 객관적으로 드러냄으로써 이루어질 수 있다고 보았다.

이렇게 계량적 방식으로 인간의 자연적 능력을 재고자 하는 관심은 이 방법에 대한 강한 믿음을 가진 사람들에 의해 수행되었다. 1930년대 하버드대학의 총장이었던 제임스 브라이언트 코넌트(James Bryant Conant)는 자연적이고 잠재적인 능력에 대한 신봉자로 알려져 있고, 지능검사의 원리에 기반한 SAT를 신뢰하여 이의 발달에 기여했다. 코넌트는 토마스 제퍼슨의 자연적 귀족주의(natural aristocracy)를 하버드대학에서 실현하고자 했던 것으로 알려져 있다.

당시 하버드대학은 전통적인 종교적 세력의 영향력이 강했고 미국 동부지역의 부유한 학생들이 많이 다니고 있었다. 코넌트는 가난하지만 학구적인 학생을 위한 전액장학금을 지원함으로써 자연적 능력에 기반한 (능력주의에 기반한) 새로운 엘리트를 키워내고자 했고 SAT를 선발과정에서 중요한 지표로 간주했다. 그는 선천적 능력인가, 성취(업적)인가라는 문제에서 전자를 선호했고 개인의 능력을 성취가 아닌 IQ에서 찾고자 했다. 성취에는 다양한 사회문화적 요인들이 작용하기 때문이다. 그는 하버드대학에서 기존의 교회세력이 아닌 민주적으로 선출된, '지능검사로 검증된' 탁월한 엘리트가 주도권을 갖기를 바랬다. 그는 자연적 귀족주의에 기반한

새로운 SAT 엘리트("new SAT elite")를 형성하고자 한 것이다(Lemann, 2000). 이러한 믿음은 자연적 능력, 메리토크라시 등이 객관성, 공정성과 결합되는 방식을 보여준다.

▫ 과학적 경영, 사회적 효율성원리에 따른 교육

셰퍼드(Shepard, 2000)에 의하면 20세기의 지배적 패러다임은 사회적 효율성(social efficiency) 교육과정, IQ 유전이론, 과학적 측정주의(scientific measurement) 등을 중심으로 구축되어 왔다. 이러한 체제는 1980-90년대를 지나면서 새로운 교육과정, 학습이론, 평가이론 등에 의해 도전을 받아 왔다. 이를 도식화하면 다음 그림과 같다.

<그림 II-1> 20세기 교육과정, 학습이론, 평가이론의 지배적 패러다임 변화

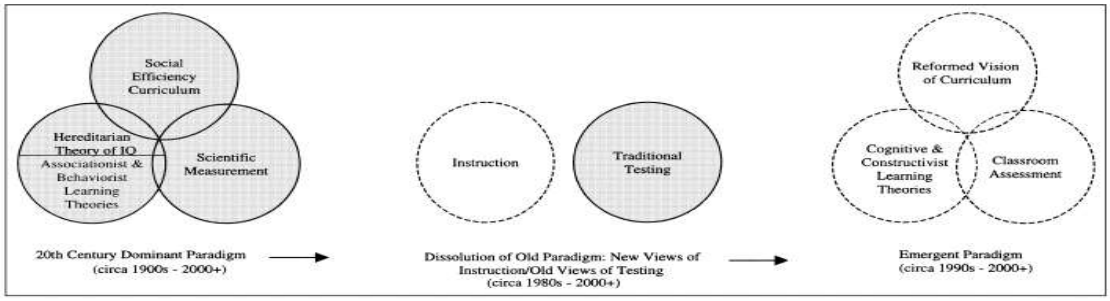


FIGURE 1. *An historical overview illustrating how changing conceptions of curriculum, learning theory, and measurement explain the current incompatibility between new views of instruction and traditional views of testing.*

(출처: Sphepard, 2000: 5)

20세기 동안 지배적 영향을 발휘한 교육과정 패러다임은 사회적 효율성주의에 입각한 과학적 경영론이었으며 이는 개인차에 대한 유전이론(IQ)과 연합주의 및 행동주의 학습이론과 긴밀하게 연관되어 있었다. 이러한 심리학적 이론들은 능력과 성취에 대해 과학적으로 설명하고자 했다.

20세기초 미국은 과학의 효능에 대한 신념이 매우 높았던 시기이다. 사회적 효율성운동은 과학이 산업화와 도시화로 인한 문제를 해결해줄 것이라는 신념에 기초했고, 산업현장의 효율성을 높이기 위해 사용된 근대과학의 원리들이 학교에도 적용되어야 한다는 관점들에 나타났다. 이는 교육심리학, 교육과정, 교육행정 등의 원리에 반영되었다.

학생의 진로에 직접 대응하는 교육(Taylor), 기능 숙달을 위한 정확한 측정 표준의 설정(연합주의 심리학) 등이 교육에서 중시되었다. 사회적 효율성운동의 대표자인 보빗(John R. Bobbitt)은 낭비의 제거(1912)를 강조했는데, 그는 평생 동안 전혀 쓰지 않을 것을 가르치는 교육과정은 개인과 사회에 낭비이며, 개인들이 각자 자신의 능력에 따라 교육을 받아야 한다고 보았다. 이러한 관점은 고도로 분화된 교육과정으로 나타났고, 학문적 교과는 대학을 가는 학생에게만 가르쳐져야 한다는 실용적 접근을 주장했다.

한편, 심리학에서 연합주의와 행동주의는 행동주의 교수-평가 모델을 개념화했다. 손다이크(Thorndike) 등의 연합주의(1922), 헐(Hull)(1943), 스키너(Skinner)(1938, 1954), 가네(Gagne)

(1965) 등의 행동주의는 학습을 자극-반응의 연합이라고 보았고, 이에 따라 교수-평가를 개념화했다.

이러한 관점에 따르면, 1) 학습은 작은 단위로 분화된 지식의 누적에 의해 발생한다. 2) 학습은 긴밀하게 순차적이며 위계적으로 이루어진다. 3) 전이는 제한적이기 때문에 각 목표는 명시적으로 가르쳐져야 한다. 4) 다음 단계 목표로 나아가기 전에 숙달도 확인을 위해 시험의 활용빈도를 높여야 한다. 5) 시험은 학습과 동형적이다 (시험=학습). 6) 동기는 외재적인 것이며, 다수의 작은 단계들로 구성된 긍정적 강화에 기초한다(Shepard, 2000: 5).

연합주의학습이론의 주창자이면서 과학적 측정주의의 아버지라고 불리는[7] 손다이크는 미국 성취평가의 특징인 소위 객관식 시험("objective" test)의 발달에 기여했다. 초기 객관식 문항들은 단순 기억에 의존하고 있었다. 셰퍼드에 의하면 이 당시의 관점은 무엇이 학습되어야 하는가에 대한 전통적 생각들을 반영하고 있는 것이라고 봐야 한다. 즉 기억하게 하고, 완성시키고, 매칭하는 형태의 시험이나 선다형 문항 등은 100여년 전 당시 무엇이 중요한 학습인가에 대한 그들의 관점을 반영하는 것이었다는 것이다.

그러나 이러한 요소들이 교육과정을 통해 고정되고 실천되면서 교과지식을 진부화시킨 경향이 있고, 평가의 객관성과 공정성에 대한 신념을 형성해왔다. 교사들이 교수활동과 분리된 별도의 공식적 시험에 대해 당연하게 생각하고, 공정성을 위해 획일적인 집행이 이루어져야 한다는 것에 찬동하는 경향이 있다는 것이다. 즉 객관식 시험 문화는 교사들의 생각을 지배하게 되어 홀리스틱하게 학생의 활동을 평가하는데 있어서의 주관성에 대해 부정적 인식을 주는데 기여한 측면이 있다(Shepard, 2000: 5-6).

□ 중등교육기회확대와 분류·배치에 대한 강한 관심

표준화시험이 20세기 전반에 빠르게 확산된 데에는 미국에서 중등교육의 급속한 확대가 중요한 배경이 된다. 미국 중등교육은 19세기말까지만 해도 유럽국가들과 유사하게 엘리트적 교육기관이었으나 1890년대에 들어 취학율이 9배 상승했고 제1차 대전후 25년간 중등교육이 급속히 확대되고 대중교육화하기 시작한다. 이에 따라 고전어를 중심으로 한 기존의 교육과정이 도전을 받게 되어 재학생중 90%의 학생이 대학진학을 하지 않는 상태에서 이들을 위한 교육과정의 문제가 논란이 된다. 1918년 중등교육의 7가지 기본원리(Seven cardinal principles in secondary education)의 발표는 변화하는 사회의 요구를 반영하는 교육과정의 시작을 알리는 것이었다. 동등한 교육내용과 수준으로 모든 학생을 끌어올려야 하는가, 아니면 각자의 능력과 진로에 맞는 차별화된 교육과정이 필요한가 하는 문제는 미국 교육과정 논쟁사의 핵심 질문이다. 결국 20세기 이후의 방향은 동일한 학교 내에서 차별화된 계열구분이 이루어지는 방식으로 전개된다. 자신의 직업진로에 따라 학생들을 분류하고 배치할 필요가 높아지게 된 것이다. 보다 많은 학생이 교육을 받게 되면서 교육체제는 보다 효율적으로 관리될 필요가 있었고, 이러한 필요는 과학적 효율성을 사회에 적용하려는 운동에 기반하여 이루어졌다(Cremin, 1964).

표준화시험은 지능이나 성취에 따른 학생 분류 및 진로 지도의 기초가 되었다. 즉 심리검사

7) 이는 1918년 Ayers에 의해 붙여진 별칭이다. Sphepard, 2000: 5.

및 표준화시험은 교육가능성을 진단하고 진로를 확인하고 분류, 배치하는데 이용되었으며, 이러한 구분은 과학주의로 정당화되었다. 학생의 분류와 배치에 있어 각종 시험이 교육기회를 제한하는 방식으로 적용될 것인가, 기회를 확대하는 방식으로 할 것인가의 문제는 논란이 되어 왔다.

중등교육의 대중화를 우려하는 엘리트주의자들은 기존에 비해 너무 많은 학생들이 학교에 가고 있다는 주장을 했고, 교육이 불가능한 자(uneducable)들이 분류되어야 한다고 보았다. 이에 대해 컬럼비아대학 티처스칼리지의 베이글리(Bagley) 등은 엘리트주의적 교육제한주의(educational restrictionism)를 비판했다. 이들은 IQ를 제대로 해석하게 된다면 교육을 제한하는 것이 아니라 확대하는 방향으로 나가게 된다고 주장했다. 즉 민주주의사회에서 교육기회에 제한은 없으며 시험의 기능은 다만 교육자들로 하여금 진단을 통해 교육을 어디서 시작하는지 알게 해주는 것이라고 보았다. 존 듀이(John Dewey)는 과학주의와 시험의 용도에 대해 언급하면서, 개인들을 분류하기 위한 목적으로 시험을 사용하는 것은 민주적 사회정책에 반한다고 지적한 바 있다(Cremin, 1964: 191). 진단과 형성의 목적이든, 분류 배치 선발의 목적이든 시험은 인간들의 능력에 접근하는 과학적 도구로서 뿌리내려지게 되었다.

여기서도 특기할 것은 이러한 경향이 미국에서만 나타난 것은 아니라는 것이다. 계량심리학과 표준화시험의 영향은 유럽에서도 상당한 정도의 영향을 미쳤지만 그 방식에는 편차가 있다. 이는 룬달(Lundahl & Waldow, 2009: 366)이 '빠른 언어(quick language)'라고 부른 바 있는 특징과 관련이 된다. 계량심리학의 결과는 복합한 현실을 단순화하여 제시해줌으로써 이해를 돕는다는 '장점'이 있다. 이러한 '장점'의 수용은 해당 사회의 지배적 문화와 수용의 조건에 따라 적용되는 방식이 사회별로, 시기별로 다르게 나타났다.[8)]

다. 표준화시험에 대한 비판: 1960-80년대 표준화시험에 대한 반대운동

미국의 경우 표준화시험이 빠르고 강하게 확산되었지만, 표준화시험에 대한 반대의 역사 또한 길다. 미국에서 표준화시험에 대한 반대는 1920년대부터 제기되어온 것으로 알려져 있다. 1920년대초 언론인 리프만(Lippmann)이 <새로운 공화국(the new republic)>에 연속적으로 기고하면서 IQ의 남용에 대해 비판한 바 있다. 그는 IQ검사의 남용이 지능카스트로 갈 것이라고 경고한 바 있다(Sacks, 2000).

이미 1920년에 50만 종류의 시험들이 존재했으며, 1930년대 1300종의 성취시험과 400여종의

8) Lundahl & Waldow(2009)에 의하면, '빠른 언어'로서의 계량심리학적 언어는 20세기 초 유럽에서도 논쟁적으로 회자되었다. 스웨덴과 독일의 대비는 계량심리학 적용과정이 각 국의 조건에 따라 달리 나타나는 것을 보여준다. 계량심리학은 1940년대 스웨덴의 교육개혁과정에 영향을 많이 주었으나 독일에서는 그 정도의 영향을 발휘하지 못했다. 1940년대 스웨덴은 전통적 교사의 권위나 계열화를 개혁하는데 계량심리학의 결과를 활용했다. 반면 독일은 계량심리학 자체의 발전이 이루어진 나라 중 하나이지만, 이를 교육에 활용하는 데는 소극적이었는데 이는 도야(bildung) 중심의 철학-해석학적, 역사적 전통이 워낙 강했기 때문이라고 할 수 있다. 스웨덴에서는 1970-80년대 동안 계량심리학 및 표준화의 영향은 신좌파적 흐름(평가거부운동 등)에 의해 힘을 잃었다가 90년대 우파정부의 신자유주의 개혁(목표-결과중심개혁)과정에서 다시 강조되었고, 독일에서는 PISA쇼크 이후("쓰나미와 유사한 영향(Tsunami-like impact)"이라고 불리는) 책무성담론이 부각되면서 재도입되게 되었다. 독일에서 1920-30년대 과학적 측정주의에 대한 논쟁이 뜨거웠던 것과 달리 현재는 교육적 표준에 대한 보편적 지지가 이루어지고 있다고 해도 과언이 아니라고 본다.

심리검사가 있었음은 앞에서 언급한 바 있다(Sacks, 2000: 70). 듀이와 같은 학자도 통계적 수치(평균, 퍼센트 등)에 대한 과도한 관심을 비판하며 이러한 범주들(평균, 열등성, 우월성 등) 속에 교육과 사회체제가 종속될 수 있음을 경고한 바 있다(Sacks, 2000: 73 재인용).

1950년대 중반에 이르러서는 시험의 과도함에 대한 비판적 논의가 증가했다. 특히 1962년 물리학자 호프만(Banesh Hoffmann)의 <시험의 압제(The Tyranny of testing)>가 출간된 이후 표준화시험에 대한 반대운동은 매우 활발하게 전개됐다. 호프만은 객관식시험이 주장하는 객관성과 선다형 문항유형을 비판했으며, 특히 SAT 등에 나타난 문항의 오류를 지적했다. 그는 표준화시험이 깊이 생각하는 천재를 가려내기보다는 오히려 차별한다고 보았다(Hoffmann, 1962).

미국에서 1970년대와 80년대에는 표준화시험에 대한 반대 분위기가 강했고 시험반대정서(Anti-test sentiment)도 강하게 나타나 지능검사에 대한 불신과 대학입시에서 표준화시험을 사용하는 것을 기피하는 경향도 나타났다. 반시험운동을 주도한 논의들은 다양한 저서나 보고서로(The mismeasurement of man, The case against SAT, The reign of ETS, Ralph Nader Report) 나타났다. 또한 시험을 반대하는 활동가들이 페어테스트(FairTest)라는 표준화시험반대를 위한 단체를 구성하게 된 점도 특기할 만하다.

표준화시험을 반대하는 논의들은 다양하다. 삭스(Sacks, 2000)에 의하면 주요 비판 관점으로는 다음과 같은 것이 있다. 첫째, 표준화시험이 학문적 성공에 대한 예측력을 지녔다고 보기 의심스럽다. 둘째, 표준화시험에서의 수행은 사회경제적 지위와 상관이 높다. 셋째, 표준화시험은 수동적이고 피상적인 학습에 대해 보상한다. 넷째, 교육을 바람직하지 않은 방향으로 유도한다. 다섯째, 의미있는 교육개혁을 방해한다.

표준화시험에 대한 비판은 1980년대 이후 참평가, 수행평가 등에 대한 강조로 나타났다. 선다형시험에 대한 대안이 모색되기 시작했으며 일부 주에서는 (켄터키, 버몬트, 캘리포니아) 선다형시험을 일시적으로 폐지하기도 했다. 일부 대학들도 1990년대 중반에 표준화시험을 입학사정에서 제외하는 등 표준화시험에 대해 반대하거나 경계하며 신중하게 접근하는 태도들이 강화되었다.

이러한 미국 표준화시험 반대운동의 역사에서 일종의 고전처럼 언급되는 책이 호프만(Hoffmann)의 <시험의 압제(The tyranny of testing, 1962)>이다. 1962년에 쓰여진 이 책은 직접적으로는 강력한 영향을 발휘하고 있던 평가기구인 ETS에 의해 실시되던 시험들(SAT, GRE 등)을 비판하는 것이었다. 저자인 호프만은 물리학자로 언론 매체인 Harper, The American Scholar 등에 기고하면서 객관식 시험(objective tests)에 대해 강하게 반대했다.

미국에서 제2차 대전 이후 시험은 하나의 산업으로 발전했다. 진위를 체크하는 시험이 전통적인 에세이시험을 대체하여 사회의 전 분야에 적용되었다. 그러다 보니 2차 대전 이전 몇몇 논자들(Jacques Barzun 등)이 시험에 대해 반대하는 글을 쓸 때는 낡은 주장이라고 아무도 관심을 갖지 않다가 1960년대에 이르러서는 시험제공자들이 수세에 몰릴 정도로 시험에 대한 비판적 관점이 확산되었다. 당시 외국(영국, 독일, 프랑스 등)은 모두 아동중심적인 교육을 추구하는 반면 미국에서는 시험과 학력이 강조되고 있다는 비판이 가해졌다(Hoffmann, 1962).

호프만은 표준화시험이 개성을 억압하고 학생의 수행에 대해 오독하게 한다고 비판한다. 그리고 더 큰 문제는 정말 우수한 학생을 가려내지도 못할 뿐 아니라 독창성을 가진 학생을 잘못

된 시험의 형식으로 인해 실패시키는 것이라고 본다.

그는 1959년에 한 유력 장학기금(National Merit Scholarship Corporation)이 장학생 선발(지원자의 2%)을 하면서 객관식 시험점수만으로 적임자를 가려낸 방법에 대해 비판하며 논쟁을 제기했다. 객관식 시험을 통해서는 시험에 유리한 능력만을 가려내지 심오함이나 독창성에 대한 평가를 하기 어렵고, 그렇기 때문에 그러한 능력을 가진 학생들에게 불리한 점이 지적되었다. 호프만에 의하면 이러한 평가에서는 아인스타인(Einstein)처럼 너무 깊이 생각하는 사람은 최종 선발의 대상이 되지 못한다. 뿐만 아니라 더 큰 문제는 아인스타인같은 천재가 100명 중 10위에 그칠 수 있다는 사실이 아니라 잘못된 선다형 문항으로 인해 정신박약아(moron)가 80위를 하게 되는 경우라고 본다. 후자는 더 잘 못된 배치라는 것이다.

호프만은 통계에 대한 맹목적 신뢰를 경계해야 한다고 본다. 통계를 통해 밝혀내지 못하는 질적인 부분들이 있다는 것이다. 시험은 자신의 머리로 독창적으로 사고하고 행동하는 창의적 사람보다는 약삭빠르고 평이하며 순응적인 사람에게 유리하다고 말한다. 보다 치밀하게 사고하는 사람들에게는 오히려 불리하다는 것이다. 특히 선다형 문항은 고르기만 하고 왜 그 답을 골랐는지 설명을 할 수 없게 되어 있어 문제라고 지적한다.

그의 책에서 호프만은 SAT등 각종 표준화시험의 문항에 나타난 문제들을 지적하는데 많은 지면을 할애한다. 문항들이 인위적으로 난해하게 만들어져 있고 의미를 모호하게 하여 풀기 어렵게 한다고 비판한다. 또한 잘못되거나 좋지 않은 문항의 사례들을 예시하고 이에 대한 시험 출제자들의 변명을 정면으로 비판한다. 그는 문항의 문제점을 지적할 뿐 아니라 테스트회사인 ETS와 싸워서 해명을 받아내기도 하고 방송출연 등을 통해 문제를 공공연하게 확산하는데 기여했다. 시험 문항의 문제점을 지적하기 위해 동일한 문항(외교관시험)을 여러 전문가들에게 테스트해보고 서로 다른 답이 나온 사례를 제시하기도 한다. 시험들이 객관성을 주장하고 있지만 모호한 문제들로 인해 전문가들조차도 서로 다른 응답을 할 수밖에 없는 상황들을 폭로한 것이다.

호프만은 그의 책 말미에서 테스트의 옹호자가 되어서는 안 되고 저항해야 한다는 메시지를 '테스트(test)'라는 용어를 재치있게 활용하여 제시한다. “시험을 옹호하지 마라(Don't be pro-test), 저항하라(protest).” 그는 당시 ETS의 권위가 막강하다는 점을 지적하며 이 기구를 집중적으로 공격했다. 특히 고등사고를 재지 못하는 객관식시험과 이에 연관된 통계기법에 대해 강하게 문제를 삼았다. 그는 객관식문제의 출제자들을 비판하면서, 그들이 거만하며 과학적인 양하고, 통계기법을 연막으로 사용하고 있다고 지적한다. 특히 그들이 객관성이나 과학성을 가정하면서 이를 배타적으로 사용하는 것에 대해 경계해야 한다고 본다.

호프만에 의하면 선다형문제는 아무리 잘 만들어도 본질적으로 결함이 있기 때문에 대안적 평가체제가 만들어져야 한다고 주장한다. 고등능력을 가려내는 방법은 객관식, 선다형 방식이 아닌 보다 개별적인(개인맞춤적인) 방식으로 접근돼야 한다고 본다. 평가에서 사고의 깊이(depth)와 정교함(subtlety)이 고려돼야 한다고 강조한다. 다른 나라들(덴마크, 영국, 프랑스, 러시아)이 에세이시험을 광범하게 활용하는 사례들을 제시하며 미국도 대안을 찾을 것을 제안한다.

라. 표준화시험의 재강화

▫ 1990년대 이후 표준화시험의 재강화

미국에서 표준화시험에 대한 반대운동은 1960-80년대에 걸쳐 비교적 활발히 전개되었다. 최근의 비판론자들에는 삭스(Peter Sacks), 콘(Alfie Kohn), 폽햄(W. James Popham), 코레츠(Daniel Koretz), 브레이시(Gerald Bracey) 등이 있다. 시민단체인 <FairTest and its assessment reform network> 등은 시험을 합리적으로 사용해야 한다고 주장하고, 표준화시험의 한계를 드러내기 위한 노력을 경주해왔다.

이렇게 한편으로는 표준화시험에 대한 반대가 꾸준히 전개된 가운데, 1990년대 이후에는 표준화시험이 학교책무성과 결합되면서 재강화된 경향이 있다. 사회문화적으로도 신자유주의의 영향으로 미국사회에서 자유경쟁이 강조됨에 따라 불평등의 자연적 기원을 찾으려는 시도가 다시 발흥하는 양상이 나타난 것도 주목된다. 이러한 경향을 반영하는 저술로 대표적인 것이 1990년대 이후 미국사회에서 큰 논란이 된 <종형분포(The Bell Curve)> (Murray & Herrnstein)이다. 이 책에 의하면 미국에서의 불평등은 서로 다른 지능을 타고난 자연적 결과에 기인한다는 것이다. 누가 능력이 있고 누가 그것을 가졌는가 하는 능력주의 관점에서 볼 때, SAT 등의 표준화시험은 여러 비판에도 불구하고 핵심적 위치를 차지해온 것이다.

또한 반동적 상황은 차별철폐조치 혹은 적극적 평등조치(Affirmative Action)에 대한 반대운동으로 전개되었다. 여러 주(California, Texas, Mississippi, Washington 등)에서 이러한 경향이 두드러졌다. 주요 주장은 인간이 능력에 의해 평가되어야지 젠더, 인종 등의 범주에 의해 평가되어서는 안 되며, 공정한 평가기준은 객관적 학점과 시험점수라는 것이다. 객관성과 공정성이라는 기준이 다시 강조된 것이다 이러한 전반적 보수화를 배경으로 1990년대에는 여러 주에서 표준기반 시험이 도입되었고, 1982년에서 1994년까지 표준화시험지는 대학교재보다 훨씬 많이 팔린 것으로 보고되기도 한다(Sacks, 2000).[9]

▫ 대규모 표준화시험과 책무성의 연결

2000년대 이후 미국은 NCLB 등의 전격 도입으로 표준화시험과 학교에 대한 책무성(accountability)을 연결시키는 정책이 강한 힘을 발휘하고 있다. 대규모 표준화시험과 학교의 책무성을 연결시키는 시도는 이미 1960년대부터 시작되었다고도 본다. 1960-70년대를 거치면서

9) 미국사회에서 제기된 '여러 비판에도 불구하고 왜 표준화시험이 계속 되는가'라는 질문에 대해 효율성(비용)을 흔히 언급하지만, 표준화시험 반대론자인 삭스에 의하면 다음과 같은 이유들이 있다(Sacks, 2000). 첫째, 미국이 유난히 심리측정을 선호하는 경향이 있다는 것이다. 유럽에서는 수업에서 배운 것을 기초로 한 에세이시험이 일반적인 데 비해 미국인들은 IQ에 집착(논리분석력, 단일한 숫자)하는 경향이 있다. 둘째, 성취보다는 잠재성에 훨씬 비중을 두고 있다. 셋째, 다른 사회들에서는 사람들의 성취를 본다면, 미국은 지능검사에 비중을 두는 이례적 사회이다. 넷째, 능력과 성취는 높은 점수를 의미한다는 인식이 대중문화 속에 있고, 이것이 표준화시험을 정당화하여 학생, 교사, 학교, 대학 등을 평가하게 한다. 다섯째, 대학들의 경제적 이익을 유지시켜준다. 대학랭킹과 SAT 점수 등과 연관해, 높은 점수는 학생의 수준을 재는 척도이고 대학의 질을 드러내준다고 믿게 하는 힘이 있다. 여섯째, 엘리트들이 유사과학적 객관성을 통해 자신의 계급적 특권을 유지하고 있다.

시험이 정보제공으로부터 교사, 학생의 책무성을 묻는 것으로 변화하게 되는데, 평가학자 코레츠(Koretz, 2008)에 의하면 이는 평가 역사에서 반세기동안의 가장 중요한 변화이다.

1960년대에 있었던 변화는 연방정부의 교육정책과 관련된다. 미국은 연방정부가 교육에 대한 직접적인 행정적 개입을 하지 않는다. 그러나 1965년의 흑인인권운동의 결과로 시민권법(Civil Rights Act)이 개정되고 소득불평등시정에 대한 요구가 높아지면서 저소득층학생의 학력에 대한 연방정부의 교육적 책임을 강화하는 1965년 초중등교육법(ESEA)이 의회에서 통과되었다. 이로부터 타이틀원(Title I)으로 불리는, 저소득층 학생의 학력 향상을 위한 보상교육프로그램이 시작되었다. 저소득층 학생에 대한 지원을 위해 연방정부는 표준화시험을 이용해 전국기준에 의한 상대평가를 도입하도록 주에 권고했다. 타이틀원 정책 자체는 시험정책도 아니고, 책무성정책도 아니었지만 이 정책의 시행으로 시험과 책무성정책의 연결이 이루어지기 시작한 것이다.

1974년에는 타이틀원에 대한 평가보고체제(TITLE 1 evaluation and reporting system, TIERS)를 만들어 표준화시험과 상대평가에 의한 성취평가로 학생점수를 평가하도록 했다. 그 이전에도 1960년대에 연방정부는 표집평가인 '학업발달에 대한 국가수준 평가제도(National Assessment of Educational Progress, NAEP)'를 실시하였다. 1990년대에는 학업성취도 결과가 상대평가로 언론에 발표됨에 따라 주별 비교의 빈도가 높아졌다. 그러나 초기에는 주나 학군단위의 정보는 제공되지 않았고 대부분의 의사결정이 평가 단위 내부에서 이루어졌기 때문에 책무성에 큰 영향을 주었다고 볼 수는 없다. 다만 정책결정자들에게 학생성취와 시기별 변화에 대한 정보를 제공하는 것에 목적을 둘 뿐이었다.

국가의 재정지원정책과 연동된 표준화시험의 실시는 책무성강화정책이기는 했으나 NAEP나 TIERS 모두 교사, 학생에게 평가결과를 가지고 직접 영향을 줄 수 있는 '고부담 평가'는 아니었다. 그럼에도 불구하고 이 두 가지는 변화의 기폭제가 되었다. 즉, 시험의 목적을 진단평가·지역단위평가로부터 대규모 모니터링·시험기반 책무성을 위한 것으로 전환시키는데 기여했다는 점이다(Koretz, 2008).

3. 주요 쟁점: 찬반 논쟁

가. 찬반 논쟁

표준화시험에 대한 찬반 양론은 팽팽하다. 흥미로운 것은 동일한 사안이나 이슈에 대해 상반된 관점이 대립한다는 점이다. 빌리(Bily, 2011)는 표준화시험과 관련한 각종 논점들에 대한 상반적 시각들을 대비하여 다루고 있다. 표준화시험이 학습을 돕는가, 교사들에게 유익한가, 대학입학시험으로 가치가 있는가 등에 대해 상반된 논의들이 전개되어왔음을 알 수 있다. 제시된 입장들을 주요 논점들을 중심으로 살펴 보면 다음과 같다.

첫째, 학습을 돕는가라는 문제에 대해 표준화시험을 옹호하는 입장에서는 학습에 도움이 된다는 입장을 취한다. 표준화시험은 사람들의 상식과는 달리 복합적 사고를 요구한다는 것이며 좋은 교사는 사실의 암기가 아닌 복합사고기술을 가르칠 수 있어야 하고 교사에게 잘 배운 학

생이라면 시험에서 성공적일 수 있다는 것이다. 또한 학생의 낮은 성취는 교사들에게 학생의 부족한 부분을 진단 가능하게 하기 때문에 학습지원에 도움을 준다는 것이다(K. Tankersley, 아리조나주립대교수).

한편, 반대입장에서는 표준화시험이 학습을 제한한다고 본다. 시험이 대부분 한 개의 정답을 요구하기 때문에 복합사고보다는 지식암기에 중점을 두도록 하는 경향이 있고 이로 인해 교육과정이 협소화되고 표준화되는 점을 지적한다. 표준화시험은 학교가 가르칠 수 있는 범위를 제한한다는 것이다. 따라서 학생들이 학업수행에 대해서 보다 정교하게 평가할 수 있는 대안적 전략이 개발되어야 한다고 주장한다(H. Cole 외, 테네시링컨메모리얼대학 교수진).

학습과 관련하여 표준화시험의 적용이 저소득층의 교육을 개선했는가, 장애학생에게 도움이 되는가의 문제에 있어서도 찬성입장은 책무성강화를 통해 공정성과 평등성에 기여한다는 논리를 견지하고(Education Equality Project, 시민단체) 반대측은 창의적이고 비판적 사고를 교육하지 않고 학생 개개인의 필요를 반영하지 못한다고 비판한다(Monty Neill, 시민단체대표). 특히 저소득층 지역의 경우 점수를 높이기 위해 다양한 활동을 제한하는 문제가 나타나지만 부유층 지역에서는 시험자체를 위한 교육과정 협소화가 덜 나타나고 있어 교육의 질적 불평등이 오히려 심화되는 점을 지적한다.

둘째로, 교사들에게 표준화시험이 유익한가의 문제에 대해 찬성 입장은 표준화시험에서 요구하는 지식과 기능을 가르치는 것은 교사들에게 중요하다는 입장을 취한다. 시험과 관련되지 않은 내용으로 수업의 상당 부분을 할애한다면 교사로서는 무책임하며, 교사의 시간과 주의를 어디에 기울여야 하는지를 결정하는데 있어 가장 적절한 방법이 시험을 준비하는 것이라고 본다(W. Gardner, 교사출신 대학강사 칼럼니스트).

반면 반대입장에서는 시험을 위한 교육은 교육을 약화시킨다고 본다. 특히 고부담책무성평가를 대비한 교육은 맥락과 개연성을 벗어난 지식이나 개념에 대한 단순 암기를 증가시킨 경향이 있다고 지적된다. 이러한 교수방법은 뇌가 원하는 새로움이나 즐거움의 단서를 제공해주지 못하고 주의를 끌지 못한다고 본다. 이러한 해독성은 표준화시험으로 인해 심화되므로 시험대비 교육은 학생의 삶, 흥미, 경험과 연관된 교육을 통해 대체되어야 한다고 본다(J. Willis, 중학교 교사, 뇌과학기반교수전략 전문가).

교사의 역할과 관련하여, 학생의 성취와 교사의 효과성을 연결시킬 수 있는가(시험에 기반한 교사평가)에 대해서도 관점이 대립된다. 일부 주에서 학생데이터와 교사데이터 간에 방화벽을 두고 있는데, 이를 연결하여 어떤 교사가 잘 가르치는가를 확인해야 한다는 관점이 있는가 하면(A. Duncan, 미국 교육장관), 학생의 시험성적은 어느 교사가 일을 잘 하는지를 판단하기 위한 공정한 자료가 아니라는 관점이 있다. 후자의 경우 시험결과는 학생이 한 순간 알고 있는 것에 대해 보여줄 뿐이며 그들이 잘 교육 받은 것인지를 알려주지는 못한다는 관점을 취한다(W. J. Popham, 미국교육학회 전 회장).

셋째는 대학입학과 표준화시험에 관한 문제인데, 빌리(Bily, 2000)등이 선정한 자료들이 표준화시험에 대한 입장은 다르지만, 표준화시험보다는 고등학교성적을 더 중시해야 한다는 점에서는 공통적 관점을 지니고 있어 주목된다. 찬성입장에서는 성취시험이 대학입시의 좋은 도구라고 보는데 고등학교에서 학생들이 배우는 내용과 긴밀히 연관된 시험은 대학입시에서 유용하다

고 본다. 이 논의는 학교에서 배운 것을 중시하며 절대평가를 강조하고 있으며 시험결과가 예측적 타당도보다는 액면타당도에 기초해야 한다고 주장하고 있어 고교에서 직접적으로 배운 결과를 더 중시하고 있음을 알 수 있다(R. C. Atkinson, 캘리포니아주립대 전 총장). 반대입장은 표준화시험이 대학입시에서 지나치게 강조된다고 지적하는데, 특히 저소득층이나 소수자들이 대학교육을 받을 수 있는 수준이지만 시험성적이 낮게 나오는 경향이 있다고 비판한다(E. Landau, CNN 리포터).

SAT가 대학에서의 성공을 예측해주는가에 대해 찬성입장에서는 표준화시험점수가 대학졸업여부를 예측하는데 훨씬 가치가 있다는 주장을 제시한다(P. Salins, 뉴욕주립대 교수). 대규모주립대학에서 SAT성적이 높은 학생을 더 많이 선발한 결과 대학졸업율이 높아졌다는 것이다. 그러나 이는 캘리포니아대학과 대비되는 주장으로 표본에 따른 차이가 있음을 알 수 있다. 반대측에서는 SAT가 대학에서의 성공에 대한 예측도구로 취약하다고 주장한다. 고등학교 학점, 석차, 학생태도 등이 학생의 수행을 예측하기에 보다 명확한 자료이며 시험은 예측력이 낮은 지표라고 본다. 특히 여성, 유색인종, 성인학생을 평가절하하기 때문에 표준화시험점수의 제출은 선택화해야 한다고 주장한다(Fair Test, 시민단체).

이처럼 관점에 따라 동일한 사안에 대해서도 표준화시험의 가치나 효과에 대해서는 대립되는 입장이 맞서고 있다. 특히 실증적 자료의 경우 표본에 따른 차이들이 있어 일반화하기 어려움을 알 수 있다.

나. 표준화시험에 대한 비판의 주요 논리

표준화시험은 '빠른 언어'로 복잡하고 복합적인 현실을 단순하게 파악하도록 해준다는 장점이 있음은 앞서 언급한 바 있다. 따라서 전체적 경향을 파악하고 개선 방안을 도출하기 위한 자료로서 의미가 있다고 하겠다. 또한 교육과정이나 교수맥락과 연계성이 높은 평가일 경우 학습 지원에도 도움이 될 수 있다는 것이 옹호론자들의 입장이라고 할 수 있다. 물론, 지능검사나 인적성검사처럼 성취보다는 잠재적 역량을 확인하고, 현재보다는 미래를 예측하기 위한 목적으로도 표준화시험(검사)은 교육에서 활용 가치가 있다. 이러한 논리가 근간이 되어 표준화시험은 여러 비판에도 불구하고 정당화되고 존속하고 있다고 볼 수 있겠다.

그러나 표준화시험이 논란이 되는 것은 그것이 그 기본적 장점에 비해 지나칠 정도로 강조되고 있기 때문이라고 할 수 있을 것이다. 표준화시험에 대한 의존이 과도하게 이루어질 경우 어떤 문제들이 있는지에 대해서는 다양한 비판들이 제기되어 왔다. 표준화시험은 근본적 가정에서부터 그것이 활용되는 방식에 이르기까지 다양하게 문제시되어왔다. 이러한 문제들을 정리해보면 다음과 같이 논의해볼 수 있을 것이다.

첫째는 재고자 하는 인간의 능력이 무엇인가 하는 점이다. 표준화시험을 통해 잴 수 있는 능력은 특정한 것(수리언어적)에 제한될 수밖에 없는데 이를 능력전체로 해석하는 경향이 있다는 점이다.

둘째는 표준화시험은 가르치고 배우는 맥락과는 다른 외부평가이기 때문에 교수나 학습에 대해 설명할 수 있는 것이 제한될 수밖에 없다는 것이다. 또한 성취에 대한 직접적 평가가 아니

라 간접적 측정으로서의 한계를 감안해야 한다.

셋째는 표준화시험의 장점이 객관성과 공정성이라고 알려져 있는데 시험점수만으로 성취에 대한 정보의 객관적 확인이 가능한지 표준을 누가 정하며 숙달도는 어떤 의미인지 등이 객관적으로 정당화 가능한가의 문제가 있다.

넷째는 평가결과에 따른 위험 부담이 큰 고부담평가의 경우 표준화시험 결과만으로 중요한 의사결정을 해도 좋은가에 대한 문제가 있다: 학생의 지식과 기능이 제한된 문제 안에서 확인 가능한가; 왜 특정 기능을 평가하며 누가 그것을 어떻게 결정하는가; 문제에 답한다는 것은 어떤 문제를 맥락적으로 발견하고 해결하는 것과 동등한가; 표준화시험은 학교에서의 학업과 생애전체를 통한 성공을 얼마나 잘 예견하는가 등의 문제가 있다.

이상의 문제들은 표준화시험이 교육과정이나 교수학습과정에 영향을 주고, 고부담평가가 되었을 때 고려할 수 있는 것들이다. 표준화시험은 그 적용에 있어 신중하게 접근되어야 함에도 불구하고 교육외적 요인들에 의해 비교육적 오용이 계속되는 경향이 있다(Harris & Smith, 2011).

보다 급진적인 논자들은 표준화시험은 교육에서 절대악이라고 주장하기도 한다. 미국 표준화시험 반대운동의 대표주자라고 할 수 있는 알피 콘은 표준화시험에는 어떠한 장점도 없다고 말하며 일관되게 입장을 견지한다(Kohn, 2000). 콘의 주요 주장은 다음과 같다.

첫째, 표준화시험은 불필요할 뿐 아니라 매우 위험하다는 것이다. 그 이유는 상관이 없는 것을 측정하기 때문이고 미국의 특수성상 어느 나라보다 일찍 모든 연령의 아이들에게 선다형시험을 실시하고 있어 위험하다고 본다. 빠르고 쉽게 낮은 비용으로 실시할 수 있다는 점에서 효율적 시험을 실시하지만, 이를 통해 중요한 능력들은 평가되지 않거나 가르쳐지지 않는다고 주장한다. 결국 시험의 배후에서 이익을 보는 집단은 영리를 추구하는 평가회사들과 정치가들이라고 본다. 숫자가 과학이라는 문화적 경향은 시험성적을 마치 판매실적과 유사하게 간주하여 성공과 실패를 판가름하는데 사용한다. 내적 동기, 지적 탐구 등 잴 수 없는 것들은 평가되지 못하며 결국 측정가능한 것들은 학습에서 가장 중요성이 낮은 것들이 된다는 것이다.

둘째는 시험의 객관성에 대한 가정이 오류라는 것이다. 객관적이기 위해서는 규범이나 표준에 대한 동의가 있어야 하지만 현실에서는 이것이 일치되고 있지 않다. 중요한 시험일수록 학생의 시험불안이 높기 때문에 성적의 타당성이 낮을 가능성이 지적될 수 있고, 정해진 시간안에 기억을 통해 즉각적으로 정답을 찾는 능력을 요구하기 때문에 사회에서 필요로 하는 능력과는 다르다. SAT 등도 사고력이나 적성을 재지 못하고 언어섹션의 경우 어휘력확인에 그치고 있어 누가 대학을 가야 하는지를 결정하는 데에 필요한 시험이라고 보기 어렵다.

셋째는 시험결과가 말해주는 것은 학생의 집이 얼마나 큰가에 대한 것 뿐이라는 것이다. 성적은 학교교육의 질을 알려주지 않으며 학생의 사회경제적 지위와 동원가능한 자원에 대해 알려줄 뿐이다. 단기기억에 의존한 시험준비결과로 고차원적 능력의 확인은 어렵고, 과정보다는 결과만을 중시하기 때문에 정답을 맞췄다고 이해를 했다고 보기 어렵다. 시험은 깊이 있는 이해보다는 학습에 대한 얄팍한 접근을 위주로 하며, 잘 교육된 사람을 가려내는 과정이 아니라는 것이다.

넷째는 표준화시험 자체가 문제이지만 그 중에서도 최악의 시험이 있다고 본다. 콘에 의하면

최악의 시험은 '상대평가로, 선다형에 의해서, 제한된 시간에, 반복적으로, 1학년부터 실시되는' 시험이다. 모든 표준화시험은 좋은 학습자의 특성을 무시하며, 잴 수 없는 것들이 많다는 것을 간과한다. 동기, 창의성, 상상력, 개념적 사고, 호기심, 노력, 아이러니, 판단, 의지, 뉘앙스, 선의, 윤리적 성찰 등이 그것이다. 반면 시험을 통해 가르치기도 한다. 사실에 대한 암기가 중요하구나, 많은 것을 아는 것이 똑똑하구나, 빠른 시간안에 하는 것이 중요하구나, 모든 것에는 정답이 있고 누군가가 답을 가지고 있구나라는 것 등이다. 따라서 독창적인 생각을 하지 않고 해석적인 활동을 하기보다 정답을 찾거나 추측하려고 한다는 것이다. 또 다른 두 가지 교훈은 개인에게 부과되는 시험이기 때문에 협력하는 것을 경계하게 하고, 사전에 문제를 알지 못하기 때문에 스트레스 레벨이 높고 암기를 하게 된다는 것이다.

다섯째는 고부담시험이 지니는 문제로, 인센티브를 통해 동기화하는 방법은 중요한 목표 도달에 도움이 되지 못한다고 경고한다. 고부담평가가 학업성취도 제고에 기여하지 못하고, 좋은 교사와 행정가들을 학교에서 떠나게 하거나 방어적이고 경쟁적으로 만든다. 교육에 대한 대화를 협소화시키고 시험 잘 보는 것에 집중시키며, 가장 심각하게는 교사들이 시험을 위한 교육을 하도록 만든다. 위로부터의 통제적 압박은 학생들의 학습에 대한 의욕을 높이고 성취동기를 부여하기보다는 오히려 질낮은 교육을 하게 한다. 높은 점수가 높은 질의 학습을 의미하는 것은 아니고, 오히려 높은 점수는 깊은 사고와는 부적 상관이 있다고 주장한다. 결국 시험은 학교에서 배우는 것이나 교육과정과 무관하게 시험보는 기술을 가르치게 된다는 것이다.

다. 표준화시험에 대한 재옹호 입장

교육에서의 책무성정책이나 고부담평가의 폐해에 대한 비판 여론이 강해지고, 교육현장이 받는 압박도 커지면서 표준화시험을 옹호하는 입장과 반대하는 입장이 팽팽하게 맞서 왔음은 앞의 사례에서도 확인할 수 있었다. 앞서 살펴본 비판론과 정반대의 입장에서 지지자들은 표준화시험의 가치를 전적으로 옹호하는 태도를 취하고 있다.

이러한 옹호입장 중 펠프스(Phelps, 2006)등의 논의는 주로 전문 심리학자 및 통계학자들의 관점을 담고 있어 주목된다. 이들은 표준화시험이 교육적으로 지닌 가치와 그 활용에 대해 전적으로 긍정적인 관점을 취한다.

이들 논의의 핵심은 대중이 표준화시험에 대해 잘 모르고 있다는 것이고, 잘 개발된 시험이 지닌 교육적 의의는 매우 높기 때문에 교육적 가치가 옹호되어야 한다는 것이다. 이들은 표준화시험이 유익하고, 어떤 다른 대안보다 우수한 테크놀로지를 갖추고 있다고 주장한다.

펠프스(Phelps, 2006)는 1960년대부터 40여년 동안의 여론조사에서 대중이나 교육자들로부터 표준화시험 자체에 대한 반대가 없었다고 주장한다. 최근 들어 교사들의 비판은 증가했으나 교육소비자들의 표준화시험에 대한 지지도는 반대에 비해 3배 이상 높다고 주장한다. 또한, 표준화시험의 장점은 개선된 진단, 예측과 선발(대학, 장학금, 고용), 성취향상(가장 논쟁적) 등이며, 표준화시험이 있는 학교체제를 다닌 학생이 그렇지 않은 학생보다 더 많이 배우는 것으로 보고되고 있다고 강조한다.

이들은 표준화시험에 대해 이루어져온 각종 비판들을 재반박한다. 예컨대, 시험을 준비하는

교육(teaching to the test)이 교육과정의 필요를 충족한다는 점, 통과점수 설정의 임의성 문제를 반박하면서 표준설정을 위한 잘 고안된 절차가 있다는 점, (언어) 소수자, 장애인들에게 유익한 사례들이 있는 점 등을 강조하고, 책무성평가에 대한 비판자들의 논거를 반박한다. 이들은 논란이 되어온 NCLB를 지지하며, 그 역사와 종합적 의의를 재확인한다(Phelps, 2006).

4. 결론: 한국교육현실에 대한 시사점

가. 객관성과 공정성의 평가문화 재고

미국에서 발달한 표준화시험은 객관성과 공정성의 확보를 가장 중요한 정당화 근거로 갖고 있다. 이는 주관적이고 맥락내적인 평가와 대조된다. 평가는 그것이 교수학습의 직접적 지원을 목적으로 하는가, 변별과 선발의 도구로 사용하는가, 교육전반의 질에 대한 모니터링을 목적으로 하는가 등에 따라 보편적 기준(표준)에 의한 객관성이 강조될 수도 있고 교사의 전문적 관찰과 개입에 의한 주관성이 강조될 수도 있다.

미국의 경우 주별 자치에 의해 교육제도가 다양하게 운영되고 있기 때문에 일반적 수준에서 상대적으로 객관성과 공정성이 강조되어 왔다고 볼 수도 있다. 복잡한 현실을 단순화하고 빠른 정보를 얻어내고자 하며 즉각적이고 효과적인 처방을 구하고자 하는 실용주의적 문화와 이를 뒷받침하는 정교한 측정방법이 발달해온 것도 이러한 조건과 관련이 될 것이다. 또한 20세기초에 중등교육이 빠르게 확산됨에 따라 학생들의 진로에 따른 교육과정차별화와 분류배치의 필요가 높았고, 대중교육체제를 관리하기 위한 효율적 행정의 필요 등이 표준화시험의 발달을 부추겼다고 볼 수 있을 것이다. 1980년대 이후 경제위기로 교육의 질 제고가 문제시되었을 때도 이를 국가 수준에서 효율적으로 관리하기 위한 방편으로 표준화시험에 의한 학업성취도와 학교의 책무성을 직접 연결시키는 조치들이 손쉽게 취해진 것도 이미 발달해 있는 표준화시험 문화가 뒷받침이 되었기 때문이라고 할 수도 있다.

다른 한편, 미국은 표준화를 반대하는 저항적 흐름도 거셌다고 할 수 있다. 19세기말 이후의 진보적 교육은 엘리트주의적 전통교육에 대립하는 공통 요소와 함께 표준화, 과학화, 효율성을 중시하는 하나의 흐름과 개성, 다양성, 시민성, 인권, 참여, 민주주의 등의 가치를 중시하는 다른 흐름 간에 차이를 드러내고 있었다. 표준화에 대한 요구가 한편에서 높아졌다면 다른 한편에서는 이에 대한 반대 흐름도 강하게 형성되어 왔다. 1960년대에서 1980년대까지의 기간은 표준화에 반대하는 경향이 상당히 지배적이었던 시기이다. 표준, 중립성, 객관성에 대한 근본적 문제제기들이 이루어지고, 공정성에 대한 기준들이 다양하게 논의되었다. 학력과 국가경쟁력의 문제를 심각하게 거론하면서 강한 책무성 정책을 사용하게 된 1980-90년대 이후, 일련의 담론과 정책들은 표준화시험의 활용을 극대화시켜왔고 이에 대한 비판도 매우 강렬하게 제기되고 있다고 하겠다.

1980년대 이후 미국과 영국을 중심으로 시작된 표준화시험과 책무성의 연계 강화 경향은 국제적으로 수렴되는 경향을 보이고 있다. 1990년대 이후 유럽연합에서는 대부분의 국가가 표준

화시험을 학력진단과 교수학습지원의 목적 이외에 교육의 질관리 기제에 활용하고 있다. 이 과정에서 미국을 중심으로 이루어져온 선다형 문항유형을 도입하는 사례들도 늘어나고 있다.

한국의 교육문화는 미국 교육사와는 매우 다른 흐름을 가지고 있지만 외관상의 유사성도 있다. 한국의 경우 객관성, 공정성을 중시하는 평가문화를 가지고 있다는 점이 특징적인데 이는 선발이 매우 중요한 비중을 차지해온 것과 관련이 된다. 한국에서는 상급학교 입학 선발이 학교교육에 결정적인 영향을 미쳐왔을 뿐 아니라 취업 선발과정도 시험의 형식으로 이루어져 있어 일관적인 평가문화가 형성되는데 영향을 주었다고 할 수 있다. 객관성, 공정성을 표방하는 시험은 대체로 점수 등으로 표시되는 결과에 의존하게 되고 사고나 학습의 과정을 파악하기는 어렵다. 또한 특정 영역에 필요한 지식, 기능, 태도와는 무관하게 일반적인 능력 혹은 일부 능력을 중심으로 선발을 하게 되는 경우가 많다. 이는 표준화시험의 단점으로 지적된 문제들과 일치한다.

가드너는 형식적 시험 모형과 도제적 모형을 구분하면서 특정한 분야의 직능을 중심으로 선발을 해야 할 경우 도제적 모형이 타당하고, 형식적 시험모형은 맥락을 배제한 평가이기 때문에 타당도가 낮다고 보았다. 역사적으로 이러한 형식적 시험모형의 대표적 사례는 중국의 관리선발제도(과거시험)이다(Gardner, 1992). 앞서 언급했듯이 중국에서 천년 이상 유지되어온 이 제도를 근대 이후 서구 국가들이 이러한 제도를 수용하게 되었다는 논의들이 있다(Teng, 1943).

한국은 고려 광종 이후 시험에 의한 선발제도를 유지해왔으며 최종 선발자의 성적은 엄격하게 서열화되어 관직부여의 기준이 되었다. 갑오경장으로 과거시험은 제도적으로 폐지되었지만, 식민지시기에는 일본의 고등문관시험(사시, 외시, 행시)이 실시되어 고급 관리들에 의한 선발이 시험에 의해 이루어졌다. 이는 일본 관료 선발을 목적으로 한 것이지만 학벌과 무관하게 시험점수에 의한 당락으로 선발이 이루어지기 때문에 나름대로 객관성과 공정성을 표방하는 제도였던 것이다. 또한 교육기회가 제한되어 있는 조건에서 각급 학교의 입학시험이 존재했고, 고등교육단계에서도 대학별 입학시험이 있었다. 이러한 관례는 해방후까지 존속해왔다. 학교에서의 교수학습의 맥락과 선발에서 요구하는 요건의 차이는 진학이나 취업의 욕구가 강한 상태에서는 교육보다는 선발에 치중하도록 요청한다. 특히 해방후 교육기회가 급격하게 개방·확대되고 학력이 지위결정의 핵심적 요인이 되면서 상급학교 입시경쟁은 과열양상을 띠게 되었다. 중등교육 평준화, 대학정원자율화 등으로 경쟁요인을 완화하는 조치들을 취했음에도 불구하고 제한된 기회배분을 둘러싼 경쟁은 실질적으로 줄어들지 않고 경쟁으로 인한 각종 사회적 비용을 지속적으로 발생시키고 있다. 특히 과열경쟁 속에서 시험점수가 능력을 재는 데 있어 객관적이고 공정한 것이라는 생각은 강화되어 왔다. 각종 경쟁 완화 조치에도 불구하고 경쟁이 줄어들지 않는 것은 학교간 위계가 존재하고 학벌의 효과가 크기 때문이라고 할 수 있다. 한 줄 세우기가 아닌 여러 줄 세우기라는 표어가 등장하여 제도적으로 시행된지 10여년이 지났지만 여전히 점수가 실력이고 객관적인 것이며 공정성의 기준이라는 인식은 근본적으로 바뀌지 못하고 있다. 대학입학선발고사가 학교 교육과정을 지배하는 양상도 거의 개선되지 못하고 있다. 입학사정관제도, 학생부전형강화, 성취평가제 도입 등 학교교육을 정상화하려는 시도들은 꾸준히 제시되어 왔지만 이를 다시 '표준화' 하려는 문화가 함께 작동하고 있다. 사회의 경쟁, 변별, 선발 압박에

학교가 자유롭지 못하고 끊임없이 맞추게 되는 조건들이 형성되고 있다.

제도적 외피는 바뀌었지만 시험의 형식만을 놓고 볼 때 전통사회로부터 현대까지 한국 사회에서는 유사한 평가문화가 이어지고 있다고 해도 과언이 아니다. 적어도 형식적 시험인가, 도제적 평가인가의 관점에서 볼 때 한국에서는 후자의 관점이 역사적으로 득세하지 못했다. 조선시대에도 유형원같은 학자는 단회적인 필기시험에 의해 현덕자를 가려내기 어렵다고 보았고 중국에서 과거제 이전에 활용되던 인재 천거방식의 도입을 주장했다. 학문의 초기단계부터의 양성과정을 강화하고 단계적으로 상급기관으로 천거해 올리는 제도의 전면실시를 기획한 바 있다. 유능하고 현명한 관리를 선발하는 것이 목적이면 그것에 합당한 평가의 타당도가 높은 선발방식을 취해야 하며, 양성과 선발이 연계되도록 하자는 이 주장은 오늘날의 학교교육정상화 논의와 유사한 관점에 터해 있다고 할 수 있다. 물론 이러한 제도는 현실화되지 못했다. 과거시험은 지필시험이기는 했지만 체계적이고 오랜 학습을 전제로 한 서술형시험이고 적어도 당시의 표준적 지식이라는 한계 안에서는 상당한 논리력과 지식응용력을 가져야만 최고 수준의 답안을 쓸 수 있는 시험이기는 했다. 뛰어난 유력 문인관료들인 성삼문이나 이이 등이 문과 장원급제자들이었던 점을 고려하면 우수한 자들을 가려내는 시험의 기능을 했다고 볼 수도 있을 것이다.

서양 근대의 산물인 능력주의가 중국의 전통적 과거시험과 같은 이질적인 시험형태를 취함으로써 관료선발시험과 같은 능력주의 기제를 형성시킨 점은, 동아시아 근대화과정에서 과거시험이 폐지되는 것(한국에서 1894년, 중국에서 1905년)에 비추어 보면 상당히 역설적이다. 19세기 말까지 존속한 과거시험의 문화적 효과는 보다 심각하게 재고할 필요가 있을 것이다. 전통사회에서 양천의 신분제약이 있었지만 과거시험은 양인 이상에게 개방된 시험이었고 그 시험의 결과에 의해서 관직이 서열적으로 부여되는 제도였기 때문에 근대 능력주의 원리와 기본적으로 통하는 부분이 있는 것이다.

식민지시기 고등문관시험이나 대학입학시험도 일본어를 모국어로 쓰는 자와 그렇지 않은 자들 간의 원천적인 차별은 있었지만, 개인의 능력으로 해소될 수도 있는 문제였다. 1924년 경성제국대학 1회 입학자인 유진오와 같은 인물이 일본인과의 경쟁에서도 수석을 할 수 있었던 것은 이 모든 장애를 극복한 결과였던 것이다. 오히려 1930년대 이후 지식위주 입학시험에 대한 비판으로 학생부, 추천서를 강화하는 전형으로 전환하면서 한국인 입학에 장애가 생긴 사례들이 많다(광주학생운동 이후 해당 학교 학생들 입학 거부 등). 이러한 전환은 당시 시험위주문화를 바꾸기 위한 노력이기는 했지만 일제의 군국주의적 통제논리와 맞물려 악용된 것이었다. 흥미로운 것은 식민지시기 경성제일고보(현 경기고)에서 교사들 간에 큰 논란거리가 된 사건 중 경성제대 입학자이지만 출석미달자인 학생에게 졸업장을 줘야 하는가 하는 문제가 있었다. 결국 논란 끝에, 불성실하기는 해도 제대입학을 한 이상 앞길을 막을 수 없으니 이 학생에게 졸업장을 부여하게 된다(이충우, 1980). 학교교육보다 상급학교진학이 우선시되는 관행이 여실히 드러나는 대목이고, 이러한 양상은 오늘날까지도 크게 바뀌고 있지 않다.

이러한 평가문화는 여러 줄 세우기를 위한 다양한 전형의 도입에도 불구하고 근본적으로 변화하지 못하고 있다. 이를 위해서는 노동시장구조, 학벌구조 등이 변화해야 하지만 무엇보다 평가의 객관성과 공정성을 대하는 시각이 문화적으로 변화해야 할 것이다. 적어도 교육 장면에

서 가르치고 배우는 맥락이 중시되고, 개개인의 다양한 능력이 존중되기 위한 다양성 기준에 대한 신념이 형성되고 강화될 필요가 있을 것이다.

나. 대안적 평가문화의 형성

여러 나라에서 표준화시험은 대학입학선발, 학교책무성강화 등을 위해 활용되고 있으며 '빠른 언어'가 지니는 장점도 있다. 그러나 여러 비판적 논의에서 제기되는 것처럼 이러한 시험에 대한 의존도가 지나치게 높아서 교수학습과정을 지배하고 인간의 능력에 대한 홀리스틱한 접근을 대체하고자 한다면 문제가 클 것이다. 표준화시험은 빠르고 쉽게 정보를 제공하는 방식이기는 하지만 보다 깊이 있는 배움을 이끌어내지 못한다면 교육적 의미는 약화될 것이다.

첫째, 변별보다는 성장과 발달을 중심으로 한 평가를 지향해야 한다. 교수학습과 평가가 연계되면서도 분리되기 위해서는 교육에 기반한 평가철학이 정립될 필요가 있다. 이제까지 표준화시험의 과도한 지배를 보완, 대체하기 위해 교수학습과정의 맥락에 충실한 평가방법들이 제안되어 왔다. 수행평가, 포트폴리오식평가, 참평가 등은 표준화하고 비교하기보다는 개인의 성장과 발달을 중심으로 하여 접근하는 방식으로 주목되어 왔다. 또한 교사의 교실평가(classroom assessment)의 중요성도 부각되어 왔다. 교수학습이 실제 일어나는 장면에서 배움과 평가가 동시에 일어나도록 하는 접근이 교육적으로 왜 중요한가의 문제는 평가철학이기도 하지만 교육관의 문제이기도 하다. 교육을 도구적 관점에서 바라보지 않고 개인의 성장 발달관점에서 접근하는 교육적 고려가 필요하다. 이를 위해서는 다양한 교육외적 요소들이 교육활동에 개입하지 않도록 교실평가의 자율성이 충분히 보장될 필요가 있다. 변별 압박, 공정성에 대한 압박 등으로 인해 교육적 평가보다는 문제없는 평가를 지향하도록 하는 것이 현행 평가의 현실적 문제점이다. 객관성, 공정성에 집착하는 것을 비판하면서도 교사 평가의 전문적 자율성에 대해 불신하고 개개인의 발달에 대한 주관적 측면을 허용하는 것에 대해 불안함을 가지고 있는 것이 현실이다. 이러한 불안감은 북유럽과 같이 단위학교 자율성이 높은 나라에서도 나타나지만 이 불안이 외부적 모니터링에 의해서가 아니라 교사들의 전문적 협력공동체 안에서 해결될 수 있도록 방향을 잡는 것이 중요하다.

둘째, 표준화시험 그 자체보다도 표준화시험의 특정 형식과 유형이 교육과정을 협소화시키고 교육내용을 제한한다는 점을 고려할 필요가 있다. 유럽에서 실시하는 졸업입학자격시험인 아비투어나 바깔로레아의 경우 서술형평가로 이루어져 있어 그 성격상 표준화시험임에도 불구하고 자료분석능력, 지식적용능력, 이해 및 탐구능력 등을 종합적으로 확인할 수 있도록 이루어져 있다. 평가를 통해 깊이 있는 교수내용과 학습방법을 유도할 수 있다면 표준화에 따른 논란 그 자체의 문제는 상대적으로 크지 않을 수도 있다.

앞서 살펴 본 바 있는 바, 최악의 시험은 '상대평가로, 선다형에 의해서, 제한된 시간에, 반복적으로, 1학년부터 실시되는' 시험이라고 보는 알피 콘의 주장은 반드시 교육적으로 고려할 필요가 있다. 표준화 그 자체가 때때로 불가피하다 하더라도, '최악의' 형식을 피하는 방법은 교육적으로 고려될 필요가 있는 것이다. 특정 형식과 유형에 치우치다 보면 시험의 형식으로 인해 제한된 능력만을 확인할 수밖에 없고 그러다 보면 시험의 형식으로 인해 잘못된 교수학습방

식과 학습습관이 의도하지 않게 형성될 수 있는 것이다. 즉 하나의 정답을 찾는 선다형문항에 익숙한 학생들은 주어진 문제에 대해 스스로 사고하기보다는 추측을 하게 될 가능성이 높고 그러한 방식으로 사고가 반복적으로 훈련되다보면 다양한 관점에서 맥락적인 사고를 하는데 제한을 받는다. 또한 문제는 선다형시험 그 자체보다도 선다형시험을 반복적으로 준비하면서 갖게 되는 사고 습성이 문제라고 할 수 있다. 실생활에서 문제해결력을 가진 주체가 되기 위해서는 학생들이 열린 질문들에 대해 스스로 사고할 수 있는 능력을 갖추는 것이 중요하다. 고부담 선다형 표준화시험 준비를 위한 반복훈련에 익숙한 교육문화에서 이러한 능력은 키워지기 어렵다. 평가문화가 형성하는 사고 습관은 고등사고를 훈련하기 위한 각종 교육과정과 충돌할 수밖에 없다.

셋째, 표준화시험이 고부담평가와 연계되면서 나타나는 교육적 폐해는 매우 직접적이다. 미국의 NCLB의 경우 학력부진학생비율을 학교의 책무성과 직접적으로 연결하다 보니 교육과정협소화양상이 문제시되어 왔다. 문제는 부유층 학교의 경우 이러한 현상이 심각하지 않은데 비해 저소득층 학교일수록 부진아비율을 낮추려다 보니 교육과정이 표준화시험에 맞춰지는 양상이 높아진다는 것이다. 표준화시험과 직접적으로 연계된 교과들의 비중이 높아지다 보니 학생의 전인교육 기회가 줄어들고, 결국 사회불평등이 심화된다는 지적을 받게 되는 것이다. 표준화시험에 의한 진단은 장기적 모니터링 등을 위해 행정적으로 필요할 수도 있지만, 이를 학교책무성과 직접적으로 연결시킬 경우 학습자개인의 필요에 대응하는 접근보다는 평균 수치를 높이기 위한 접근이 강화될 수밖에 없다. 교육과정을 협소화하지 않는 책무성 점검의 방식이 교육적으로 고민될 필요가 있다.

넷째 과학성, 객관성, 공정성 등을 표방하며 질 관리를 하고자 하는 평가문화(audit culture)가 형성됨으로써 얻는 것과 잃는 것이 무엇인지 엄밀히 따질 필요가 있다. 표준화시험의 역사에서 드러나듯이 표준화시험은 외부평가이며 교수학습의 맥락을 제한하거나 통제하는 효과를 지닌다. 표준화시험을 통해 학교의 내부를 단순화한 지표로 드러냄으로써 교수학습의 일반적 수준을 확인하고 학생들의 진보정도를 파악할 수 있다. 교사와 학교의 역할이 학생의 성취(평가결과)와 동일시되어야 하는가의 문제는 논란이 된다. 그러나 이러한 평가결과를 통해 학생들의 일반적이고 전체적 수준을 제한된 수준에서나마 파악하는 것이 가능하다.

문제는 이 결과를 어떻게 해석하고 교수학습이나 선발, 정책제도에 어떻게 반영할 것인가의 문제이다. 평가결과가 무엇을 측정한 것인가에 대해 섬세하게 접근할 필요가 있다는 것이고, 액면 점수만으로 결과를 이해했을 때 어떤 오류를 범할 수 있는지 종합적인 고려가 필요하다. 평가 결과를 통해 교수학습의 질을 어느 수준까지 설명할 수 있고 그 개선방안에 대해 무엇을 말할 수 있는지 질문해봐야 한다. 평가결과에 영향을 미친 다각적 요인들에 대한 엄밀하고 공정한 분석 없이 학교 간 차이 등이 단순 비교되는 오류들에 대해 문제제기를 하지 않는 것은 주어진 점수의 객관성과 과학성에 대한 일종의 신화적 믿음이 전제되기 때문이다. 대학입학상의 다양한 전형요소들이 변별을 위해 결국 표준화시험(수능)의 점수로 귀결되는 것, 표준화시험의 결과가 그대로 고부담 정책으로 귀결되는 것 등은 평가를 대하는 문화를 반영하는 것이다. 표준화시험의 결과가 전체를 보여줄 수 없는데도 전체를 본 것으로 간주한다면 그것은 객관적이지도, 공정하지도, 과학적이지도 않다. 총체성이 드러나야 전체에 대한 판단이 가능하게 되는

것이다. 이를 위해서는 '빠른 언어(quick language)'를 넘어서는 종합적 접근이 필요하고, 빠른 언어가 아니라 복합적이고 맥락적인 질적 특성을 포함하는 '치밀한 언어(thick language)'를 사용할 수 있어야 하는 것이다.

Ⅲ. 미국에서 표준화시험의 형성과 전개

1. 공교육의 확대와 표준화시험의 형성

▫ 1845년 미국 최초의 표준화 시험

미국에서 최초의 표준화시험은 1845년 매사추세츠주 보스톤에서 호레이스 만의 주도로 실시된 문법학교 대상의 시험이었던 것으로 알려져 있다. 이 시험은 학교별 시험을 벗어난 최초의 대규모 필기 시험으로 기존에 미국의 학교들에 실시되던 시험과 그 성격이 매우 달랐던 것으로 알려져 있다.

이 시험은 몇 가지 면에서 전통적 시험과 구분된다. 첫째는 필기시험으로 평가를 했다는 점이다. 둘째는 학교별로 이루어지는 평가가 아니라 여러 학교를 대상으로 한 대규모평가였다는 것이다. 셋째는 가르친 내용을 점검하는 것이기는 하나 교사가 가르친 것을 그대로 평가하는 것이 아니라 외부적 기준에 의거하여 평가가 이루어졌다는 것이다. 넷째는 평가의 결과를 기초로 하여 학교의 교육운영에 대한 책무성을 점검하는 평가라는 것이다.

이러한 특징들은 현재 논의되는 '표준화시험'의 전형적 특징들을 담고 있지만 당시만 해도 이러한 평가는 매우 새로운 것이었다. 놀라운 것은 1840년대 당시에는 매우 새로운 시도로 논란이 되었던 이러한 형태의 시험이 몇십년 사이에 빠르게 확산·적용되어 1870년대에는 이미 문제거리로 부각될 정도가 되었다는 것이다. 남북전쟁 이전에 이미 경쟁적 시험은 공립학교 시스템에서 핵심적 지위를 갖추어가고 있었다. 미국에서 표준화시험은 짧은 기간 동안 학교현장에 자리잡아 나가고 있었던 것이다.

미국에서 실시된 최초의 표준화시험을 분석한 교육사학자 리이스(Reese, 2013)는 찰스 파커(Charles I. Parker)가 1878년 스프링필드에서 열린 주교사협의회의 모임에서 당시의 교육이 '시험 매니아'에 빠지고 있고 학생들은 '걸어 다니는 백과사전'으로 만들어지고 있다고 비판한 것을 중요하게 언급한다. 파커는 퀸시 개혁(Quincy Reform)으로 진보주의교육운동의 역사에서 중요한 획을 그은 바 있다. 교사들은 학생들의 시험준비를 지도하느라 바쁘며 학생은 각종 지식(강의 길이, 미국 대통령 및 영국왕들의 계보, 문법규칙, 수학공식 등)을 암기해야 했다. 학교들은 학생의 진보를 추적하고 교사들의 책무성을 강화하기 위해 '통계국'으로 바뀌어 가고 있었다.

리이스에 의하면 파커가 지적하는 양상들은 1840년대와는 판이하게 달라진 것이었다. 그렇다면 1840년대 중반, 표준화시험은 어떻게 생겨나게 되었으며 기존의 교육은 어떠했는가? 미국에서 전통적 평가는 전시 혹은 발표(exhibition)와 동일시되었으며 학교에 대한 평판은 '인상'에 달려 있었다. 즉 교사들이 질서있는 학교를 유지하는가, 부모와 지역사회인사가 방문했을 때 학생들이 바르게 행동하는가 등과 같이 관찰된 내용을 토대로 하였다. 대부분의 학교가 연례 전시회를 열어 지역 공동체내에서 이벤트를 개최했다. 이때 동네의 어른들이 학생들에게 질문을 하게 되는데, 교사는 이를 몇 주간 학생들과 함께 준비를 했다. 학생들은 암기를 해서 연설

을 하고 노래나 찬송가를 부르는 것을 통해 관객을 즐겁게 하였다. 적어도 1800년대초의 학생은 시간이 정해지고 표준화된 필기시험을 보지 않았다.[10]

1840년대 미국의 대표적 개혁가들은 후일 표준화시험으로 알려진 시험을 도입하면서 이를 통해 진보적인 교육방법이 발전할 것을 기대했다. 왜냐 하면, 새로운 시험을 주도한 호레이스 만 등은 경쟁적인 시험과 필기시험이 교수법을 혁신할 것이라고 믿었고, 당시 발달한 신생학문인 통계학의 원리를 미국 공립학교에 적용하고자 했기 때문이다. 보스톤지역에서 시험이 실시되었던 것은 매사추세츠 교육위원회의 비서를 맡고 있던 호레이스 만의 주도에 의한 것이었다. 그는 당시 문법학교들에 대한 불만들을 청취하고 있었고 학교들에서 교육이 쇠퇴하고 있다고 믿었다. 따라서 학교조직 및 교수법의 변화를 위해 위로부터의 개혁을 시도하고자 했다. 호레이스 만은 자신과 뜻을 같이 하는 개혁적 인사들(대학교수, 법학, 의학 등 전문분야)로 구성된 위원회를 통해 시험을 출제하고 채점한 후 그 결과를 보고하도록 했다. 당시 개혁주도자들은 외국 경험(특히 프러시아 교육)을 가진 자들로 스스로 코스모폴리탄하다고 생각하는 자들이었으며, 자신들의 혁신적 아이디어가 미국교육을 개선할 것으로 확신했다. 이들은 유럽을 다녀온 후 보고서를 발표하여(1844년) 프러시아교사들을 극찬했다. 프러시아교육이 전통적 교과교육 및 아동중심적 교육에서도 우위라고 강조했다. 교사들은 호레이스 만이 주도하는 시험이 학교현장의 자율성(전통적 교수, 평가 방식)을 저해하고 폄하하기 위한 다분히 의도된 기획이라는 점을 부각하며, 강하게 반발했다. 당시 보스톤시내의 중등교사협의회인 마스터연합회(Association of Masters) 소속 31명의 교사는 1844년 9월 호레이스 만을 '교육경험이 없는 아마추어'라고 비난했다.

사무엘 호우(Howe)를 대표자로 하여 실시된 시험의 결과에서 학교들의 '실패'가 '점수'로 드러났다. 시험은 문법학교 최고학년인 전체 530명의 학생이 대상이었다. 이 시험은 대규모적으로 실시되고 체계적으로 분석되었다. 특히 성취를 비교하고 학교조직과 교수법 개선을 위해 통계를 사용했다는 점에서 특기할 만했다. 학생들은 공통의 절차, 동일한 조건, 동일한 문항으로 평가되었고 부정행위를 하지 못하도록 떨어져 앉아서 시험을 보았다.

채점자들은 보고서에서 학생들이 교과서지식을 언어적으로만 학습하고 있다고 지적하며, 학생의 독해나 이해능력이 떨어진다고 보았다. 또한 보스톤지역과 그 이외의 학교와 비교하여 볼 때 보스톤 학교들의 성취가 더 낮다고 지적했다. 이 평가결과 분석을 통해 교사간 비교가 이루어졌고 좋은 교수법 등이 발굴되었다.

시험의 결과를 통해 개혁안들이 제시되었다. 시스템을 바꾸기 위해서는 성취를 표준화할 수 있는 전업 행정가가 필요하다고 보아 교육장이 선출되어야 한다고 주장되었다. 또한 최고학년을 마스터(master)들이 가르치고 있는데 점수결과를 보면 도저히 정당화될 수 없다고 보았다. 표준을 설정하고 학생의 성취변화를 확인할 책임자가 필요하다고 본 것이다. 당시 보스톤에는

10) 서구에서 시험의 역사를 추적한 Teng Ssu-yu(1943)에 의하면 examination이라는 용어가 test의 의미로 사용되기 시작한 것은 1612년(Murray의 New English Dictionary 인용)이며, 고대 그리스 로마 시기나 중세시기까지도 형식적인 시험은 존재하지 않았다고 본다. 중세대학에서 신학 학위 개념이 생긴 것이 1215년 경인데 이것이 서구에서 시험의 시작이라고 보고 있다. 18세기까지도 시험은 일반적으로 구술시험이었고 필기시험은 생소한 것이었다. Rouse(History of the study of mathematics at Cambridge, 1835)라는 학자를 인용하여 1702년 캠브리지대학의 Bentley에 의해 실시된 시험이 기록으로 찾을 수 있는 서구 최초의 필기시험이라고 보고 있다. Teng, 1943: 271-273.

필요인원의 두 배 이상의 마스터들이 존재하고 있었고 이들이 최고학년을 가르쳤다. 그 아래 학년은 어셔(usher)들이 가르쳤으며 가장 어린 학생은 저임금 여교사들이 가르쳤다. 개혁가들은 마스터와 어셔의 수를 줄이고 여교사들을 더 많이 충원하는 것이 비용효율상 좋을 뿐 아니라 여교사들이 더 교육적이기도 해서 바람직하다고 보았다. 전통적 방식의 남자교사들은 체벌을 포함해서 악명높은 훈육방식을 사용하고 있었지만 여교사들은 그렇지 않았기 때문이다. 개혁가들은 기존 남자교사들의 행위를 기록하여 해고하고, 그 자리에 여교사를 충원하면 교실안의 몽둥이에 먼지가 앉을 것이라고 주장했다(Reese, 2013).

□ 외부의 관점 대 교육계의 관점: 표준화시험과 사회공학

이러한 시험 결과에 대해 휘그당(Whig) 개혁가들은 매우 반가와 했고, 보스톤에서의 시험결과를 전국적으로 홍보했다. 이들은 시험의 절차 및 결과가 투명하게 드러남으로써 교사와 학생이 어떻게 가르치고 배우는 지를 알 수 있다고 보았다. 이처럼 전문가들에 의해 공정하고 측정가능한 표준들이 설정되면 학교는 명백하게 개선될 것이라고 보았다. 1845년 여름과 가을에 있었던 시험과 채점의 결과는 큰 반향을 일으켰다.

1846년 보스톤시는 교육장을 임명하기로 발표했고 곧 로드아일랜드 교육장 출신인 비숍(Bishop)이 최초의 시교육행정가가 되었다. 1850년대에는 프러시아식 개혁모델에 따라 다수의 여교사를 배출할 보스톤사범학교가 설립되었다. 교육장의 장학지도권한이 증대되었다. 마스터들은 계속 반발했지만, 그들의 대부분이 퇴임이나 사망 등으로 교육현장을 떠남으로써 세력은 현저히 약해졌다. 향후 10년간 보스톤시 공립학교는 학교조직, 학년제 등에서 꾸준한 변화가 이루어졌다. 이후 모든 학교에 대한 공통적 표준의 제시가 필요하다는 주장이 설득력을 얻어갔고 보스톤이외의 지역에서도 필기시험 등이 도입되었다.

이상의 과정을 통해 볼 때 보스톤에서 실시된 최초의 표준화시험은 우연히 등장했다기보다는 위로부터의 개혁을 뒷받침하기 위한 목적에서 실시된 것임을 알 수 있다. 이러한 외부자에 의한 진단 및 개혁 유도로 인해 가장 타격을 받은 대상은 전통적 권위를 유지하고 있었던 엘리트 문법교사들이었다. 이들은 교과서를 직접 제작했고 자신들의 훈육에 대한 강한 확신을 가지고 있었다. 이로 인해 새로운 개혁의 가장 큰 걸림돌이기도 했던 것이다. 참혹한 시험 결과는 이들의 교육활동의 정당성을 크게 훼손시켰고 개혁가들의 입지를 강화시키게 되었다.

시험은 다른 분야의 개혁과 광범하게 연결되어 있었다. 즉 행정의 집중화, 선출직 행정가에 의한 지도, 표준 설정, 교과서 및 교육과정의 통일 등은 개혁의 목적과 연결되어 있었다. 시험을 통해 학생의 성취가 드러날 뿐 아니라 효과적 교사와 무능한 교사가 가려지고 문서화되기 시작했으며, 학교의 실패가 드러났다. 한편, 이 시기에 표준화시험이 실시될 수 있었던 배경에는 다음과 같은 조건도 발달해 있었다. 교육전문가 단체의 출현과 인쇄 매체 등의 발달로 의견교류 및 공론화가 가능했다는 점과 통계학의 발달이 현실을 분석하는 접근방식을 제공한 점이 그것이다. 이러한 인프라는 위로부터의 개혁 추진을 용이하게 한 요소이기도 했다(Reese, 2013).

보스톤에서 실시된 최초의 표준화시험은 시험 형식에 있어서의 큰 변화였을 뿐 아니라 사회통합을 위한 사회적 공학의 발달을 상징하는 것이기도 하다(Garrison, 2009). 게리슨에 의하면 호

레이스 만은 표준화시험을 통해 공립학교의 실패를 드러냄으로써 새로운 시민형성을 위한 공통의 기준을 도입하고자 했다. 이 기준은 공화주의 이론에 기반한 것으로 자연적 귀족주의를 지향하는 것이었다. 이를 통해 미국식 메리토크라시가 실현되기 시작했다고 본다. 즉 표준화 시험의 실시 이전에는 '지식의 확산'이 주요 모토였다면, 이제는 공정성이 중요하게 부각된다.

호레이스 만은 표준화시험의 의의를 다음과 같이 보고 있다. (Kandel의 요약, 1936; Garrison, 2009 재인용)

1) 균등성(impartial): 모든 학교와 학생에게 균등하게 실시
2) 공정성(just): 구술시험은 2분내외로 질문하는데 반해 1시간동안 충분한 시간 사용
3) 철저함: 넓은 범위, 충분한 문항수
4) 교사의 개입 제한: 학생 대답을 중단하거나 제안을 하는 등
5) 교사의 편견, 선호 방지
6) 학생들이 신뢰할만하고 유능하게 가르쳐졌는지 확인 가능: 자신의 사고를 전개시키고 사실간의 관계를 드러낼 수 있음
7) 성적표는 영구적 기록으로 남음: 학교간 비교가 가능하고 학교 내부에서 개선상황의 확인이 가능

흥미로운 것은 표준(standard)의 문제를 공정성과 연결하여 논의하는 점이다. 문제는, 공정하고 중립적이라고 할 때, 표준은 누가 정하냐가 쟁점이 될 수 있다. 물론, 표준은 당시의 지배적인 휘그당의 관점을 반영했고 그들이 정하는 것이었다. 이렇듯, 미국에서 표준화시험은 휘그당의 공화주의 이념을 반영하면서 사회공학적으로 도입된 측면이 있으며 획일적 적용에 기반한 '공정성'을 통해 정당화되고 있었다.

1840-50년대 이후 남북전쟁시기까지, 적어도 미국의 동북부에서 표준화된 필기시험은 상당한 정도로 보편화된 것으로 나타나며 이는 전통적 학교를 개혁하려는 움직임과 연관되어 있었다. 미국에서 20세기 이후 심리측정 및 통계학이 발달함에 따라 교육에서 계량심리학의 영향이 커지고, 표준화된 각종 심리검사들이 지배적 영향을 미치는 것의 배경에는 이미 미국사회에 표준화시험문화가 19세기말에 만연해 있었다는 점이 지적된다(Reese, 2013).

이러한 표준화시험문화는 당시 교육계 외부로부터의 학교개혁운동과 밀접히 관련되어 있음을 주목할 수 있다. 표준화시험은 외부의 주도에 의해 이루어진 경향이 있고, 외부의 관점을 반영한다. 그리고 그 갈등 대상이 교육계 내부라는 것이 특징이기도 하다.

1890년대초 언론을 통해 (The Forum, 1892년 10월부터 1893년 6월) 미국의 학교교육을 비판함으로써 큰 반향을 일으킨 조세프 메이어 라이스(Joseph Mayer Rice)는 36개 도시 학교를 탐방하여 1200명의 교사를 면담한 후 교육에 대한 대중적 무관심, 정치적 개입, 부패와 무능력 등을 폭로했다. 개혁에 대한 일반 언론을 환기시키고 교육개혁을 요청한 라이스의 주장은 교육계로부터 '교육현실을 모른다'고 비판을 받기도 했다. 라이스도 미국의 시험운동(testing movement)의 역사에서 기억되는 인물이라는 점은 '외부로부터의 개혁요구 대 교육계 내부'의 갈등구도에서 '시험'이 차지하는 위치를 확인하게 해준다(Cremin, 1964).

2. 계량심리학, 과학적 경영과 표준화시험

▫ 인간의 능력을 “재다”: 계량심리학의 발달과 시험

19세기에서 20세기초로 넘어오는 과정에서 교육에 가장 크게 영향을 준 것이 계량심리학의 발달이라고 해도 과언이 아니다. 18세기 계몽주의사상이 인간의 계몽가능성을 신뢰하며 공교육 체제의 기획에 힘을 실어주었다면 대중교육이 확대되는 19세기후반에 와서는 인간의 교육불가능성(ineducability)에 대한 생각들이 발달하게 된다(Garrison, 2009). 프랑스 제3공화국 하에서 교육이 불가능한 자들을 가려내려는 정책적 관심에서 탄생한 소위 IQ검사는 인간에 대한 분류를 위한 기제로 발달해왔다. 지능검사는 인간의 정신 능력을 재고 수치화함으로써 객관화하기 위한 시도이다. 프랑스에서 비네(Binet)가 지능검사의 창시자로 알려져 있다면, 영국에서도 갈톤(Galton), 스피어맨(Spearman) 등이 이와 관련되며 우생학주의(Eugenics)에 영향을 주게 된다. 지능은 인간의 능력을 분류하는 기준이 되었을 뿐 아니라, 유사 과학적인 객관성을 통해 엘리트들이 지배적 특권을 누릴 수 있도록 하는 근거로 작용하기도 했다. 지능은 성취가 아니라 잠재성을 의미하는 것으로 능력을 자연적으로 이해하도록 한다. 이는 두 가지 의미를 갖는다. 하나는 사회경제적 환경으로 인해 생긴 불평등을 자연적 능력을 통해 교정할 수 있다는 것이다. 다른 하나는 능력을 자연적인 것으로 보게 함으로써 인간에 대한 분류 자체와 능력의 차이에서 발생하는 불평등을 정당화한다는 것이다.

무엇보다 IQ검사의 특징은 능력이 숫자로 표시된다는 점에서 능력의 측정 지표, 측정 방법, 표기 방법 등에 대한 논란을 일으켜왔다. 인간의 능력을 단일한 숫자로 표시하여 분류하는 것이 미국에서 본격적으로 실용화한 것은 제1차대전 전후라고 할 수 있다. IQ검사지를 1905년에 처음 고안한 비네는 1911년에 사망했지만, 미국 스탠포드대학 교수 루이스 터먼(Lewis Terman)에 의해 비네-스탠포드 척도가 만들어지고 IQ(Intelligence Quotient, 정신연령 대 신체연령)라고 명명되었다. 미국에서 IQ검사는 군대에서의 직책구분을 위해 실시되었고, 이후 학교에도 적용되어 잠재적 능력을 재는 검사로서 미국적인 시험문화 형성에 기여했다.

미국에서 IQ 검사의 원리는 학업성취의 예언을 위해 SAT, GRE 등의 시험에 적용되었다. 프린스턴대학의 심리학자인 브릭험(Brigham)은 1차대전동안 군대 IQ검사에 참여한 본인의 경험을 토대로 1926년 군대검사를 변환하여 SAT시험을 개발하였다. 그는 유제닉스에 관심을 갖고 미국인의 지능을 종족별로 구분하는 연구를 수행했다(A study of American intelligence). 흥미로운 것은 1932년 <오류에 대한 연구(A study of error)>에서 그의 전작을 비판하고 SAT나 IQ로부터 거리를 유지하기 시작했다는 것이다. 브릭험은 IQ검사에 대한 신뢰가 아무런 생물학적 근거도 갖지 않는 위험한 것이라고 보았고 이의 실용화에 대해 여러 차례 경고했다. 아이러니한 것은 SAT 개발자가 회의를 갖는 동안 SAT 산업은 지속적으로 확대되고 있었다는 것이다(Lemann, 2000).

한편 대학원입학 전형자료로 사용되는 GRE를 개발한 우드(Wood)의 경우는 회의적인 브릭험과 달리 표준화시험의 강한 지지자였다. 심리학자 손다이크(Thorndike)의 제자였던 우드는 학교교육이 비논리적이고 비조직적으로 이루어지는 점을 비판했다. 그는 국가수준의 시험체제를 통해 두 가지 목적을 달성하고자 했다. 첫째는 고등교육을 위한 가장 뛰어난 학생을 선발하는

것이었으며, 둘째는 교사들의 임의적 권력을 박탈하고자 하는 것이었다. 그는 학생들의 교과숙달도를 드러냄으로써 이를 이룰 수 있다고 보았다.

1930년대 하버드대학의 총장이었던 제임스 브라이언트 코넌트(James Bryant Conant)는 자연적이고 잠재적인 능력에 대한 신봉자로 알려져 있는데, 지능검사의 원리에 기반한 SAT를 신뢰했다. 코넌트는 토마스 제퍼슨의 자연적 귀족주의(natural aristocracy)를 하버드대학에서 실현하고자 한 인물로 알려져 있다. 당시 하버드대학은 전통적인 종교적 세력의 영향력이 강했고 미국 동부지역의 부유한 학생들이 많이 다니고 있었다. 코넌트는 가난하지만 학구적인 학생을 위한 전액장학금을 지원함으로써 자연적 능력에 기반한 (능력주의에 기반한) 새로운 엘리트를 키워내고자 SAT를 선발과정에서 중요한 지표로 간주했다. 그는 선천적 능력인가, 성취(업적)인가라는 문제에서 전자를 선호하여, 개인의 능력을 성취가 아닌 선천적 능력(IQ)에서 찾고자 했다. 성취에는 다양한 사회문화적 요인들이 작용하기 때문이다. 그는 하버드대학에서 기존의 교회세력이 아닌 민주적으로 선출된, '지능검사로 검증된' 탁월한 엘리트가 주도권을 갖기를 바랬다. 그는 새로운 SAT 엘리트("new SAT elite")를 형성하고자 한 것이다(Lemann, 2000).

▫ 중등교육 확산과 표준화시험의 확대

제1차대전후 25년간 중등교육이 급속히 확대되었다. 이에 따라 행정의 중앙집권화 경향이 강화되어야 한다는 주장이 교육자 단체인 NEA 등에서 제기되었다. 보다 많은 학생이 교육을 받게 되면서 교육체제는 보다 효율적으로 관리될 필요가 있었고 이러한 필요는 과학적 효율성을 사회에 적용하는 운동에 기반하여 이루어졌다(Cremin, 1964).

표준화시험은 지능이나 성취에 따른 학생 분류 및 진로 지도의 기초가 되었다. SAT나 GRE와 같은 지능검사에 기반한 시험 외에도 19세기말부터 미국에는 다양한 종류의 시험들이 학교에서 사용되고 있었다. 이미 1920년에 50만 종류의 시험들이 존재했으며, 1930년대 1300종의 성취시험과 400여종의 심리검사가 있었다(Sacks, 2000: 70). 심리검사 및 표준화시험은 교육가능성을 진단하고 진로를 확인하고 분류, 배치하는데 이용되었으며, 과학주의로 정당화되었다.

중등교육의 확대는 한편으로는 엘리트주의자들로 하여금 너무 많은 학생들이 학교에 가고 있다는 주장을 하게 했고, 교육이 불가능한 자(uneducable)들이 분류되어야 한다고 보게 했다. 이에 대해 컬럼비아대학 티처스칼리지의 베이글리(Bagley) 등은 엘리트주의자들의 교육적 제한주의(educational restrictionism)를 공격했다. 민주주의사회에서 교육기회에 제한은 없으며 시험의 기능은 다만 교육자들로 하여금 어디서 시작해야 하는지를 알게 해주는 것이라고 주장했다. 따라서 IQ에 대한 제대로 된 해석은 교육에 대한 제한이 아닌 확대로 나아가야 한다고 주장했다. 존 듀이(John Dewey)도 과학주의와 시험의 용도에 대해 언급하면서 개인들을 분류하려는 목적으로 사용되는 것은 민주적 사회정책에 반한다고 지적했다(Cremin, 1964: 191).

▫ 제2차 세계대전후 평가산업의 조직화와 발달

2차대전이 끝난 후 미국의 대표적 평가기구인 ETS(Educational Testing Service)가 출범하게

된다. ETS는 College Board출신의 초온시(Chauncey)가 대표를 맡고, 하버드대학교 총장 출신 코넌트(Conant)가 이사장을 맡아 1948년 이후 프린스턴에서 업무를 개시했다. ETS의 설립으로 평가산업은 본격적으로 성장했고 각종 민간 평가기구들이 출범했다. 평가산업은 엄청난 자본이 집중되는 산업으로 활성화된 것이다.

표준화시험의 본격적인 발전은 채점의 기계화가 가능하면서부터였다고 할 수 있다. 대규모시험의 경우 채점의 효율성이 떨어질 경우 자주 실시하기 어렵기 때문이다. 이미 1931년에 IBM은 채점의 기계화를 도입하여 대규모 표준화시험의 실시를 용이하게 했다. 채점 기계는 연필의 흑연을 인식하는 기계를 제작한 고교 교사 출신인 존슨(Johnson)의 발명기술을 IBM이 인수함으로써 시작되었다. 정교화한 통계적 방법과 사회로부터의 책무성 요구가 결합하여 평가산업은 계속 발전하게 되었으며, 성공적 학교의 징표는 측정가능한 점수로 인식되어갔다.

ETS의 대표였던 초온시는 시험이 인류에게 유익함을 준다는 확신을 가지고 있었다. 그의 꿈은 인간들의 '다양한 능력의 센서스'를 구축하는 것이었다. 이러한 능력의 센서스(the Census of abilities)는 성공하지 못한 것으로 평가된다. 결국 학문적 능력(scholastic aptitude)이라는 하나의 능력에 대한 센서스(census of one ability)에 그쳤기 때문이다(Lemann, 2000).

제2차 대전후 고등교육인구의 증가로 SAT 응시율이 높아졌다. SAT가 자연적 능력에 관심이 있다면, 이와 달리 성취(achievement)를 중시하는 ACT도 경쟁적 시험으로 대학입시에 활용되었다. ACT(American College Testing)는 1956년에 만들어진 평가기구로, 주로 덜 경쟁적인 주립대학들에서의 학생선발을 목적으로 한 것이다. 이 시험을 개발한 린퀴스트(Lindquist)는 중서부 주립대학을 중심으로 활용되던 아이오와 테스트(Iowa Test)의 개발자로 대중교육적 관점을 취했다. 그는 애초에 ETS의 부대표 자리를 제의받았지만 거절한 바있다. 당시로서는 평가분야에서 ETS에 소속되어 있지 않은 가장 중요한 인물이었다고도 평가된다.

1950년대는 교육계에서 시험의 과도함에 대한 논의가 증가한 시기라고 할 수 있다. 1957년 카네기재단의 대표인 존 가드너와의 공조로 ETS가 이미지 제고에 성공하면서 교육계의 신뢰를 얻고자 했는데, 가장 큰 걸림돌은 ACT였다. SAT와 ACT의 차별화 지점은 지능이냐 성취냐의 문제였다고 할 수 있으며, 대학들에 따라 입학전형에서 이들을 반영하는 기준은 달랐다. 전통적으로 SAT를 적용해온 동부사립대학과 달리 ACT는 주로 중서부 주립대학들에서 발달해왔다. 1950년대에는 주립대학들의 경우도 ETS 시험(텍사스, 미시건, 조지아, 콜로라도 등)을 취하는 주와 ACT 시험(일리노이, 오하이오)을 취하는 주가 나뉘었다. ETS는 미국 속의 미국이라고 불리는 캘리포니아를 기회로 인식하여 UC계열에 College Board의 첫 주립 지회를 설립하게 되었다. 이렇듯 미국의 대부분 사립대학과 주립대학은 ETS, ACT와 같은 대규모 민간 평가기구에서 실시하는 시험을 통해 신입생을 선발하게 된 것이다(Lemann, 2000).

3. 표준화시험에 대한 반대와 재강화

▫ 1960-80년대 표준화시험에 대한 반대운동

미국에서 표준화시험에 대한 반대는 1920년대부터 제기되어온 것으로 알려져 있다. 1920년대

초 언론인 리프만(Lippmann)이 <새로운 공화국(The new republic)>에 연속적으로 기고하면서 IQ의 남용에 대해 비판한 바 있다. 그는 IQ검사의 남용은 지능카스트로 갈 것이라고 경고한 바 있다(Sacks, 2000). 이미 1920년에 50만 종류의 시험들이 존재했으며, 1930년대 1300종의 성취시험과 400여종의 심리검사가 있었음은 앞에서 언급한 바 있다. 듀이도 통계적 수치(평균, 퍼센트 등)에 대한 과도한 관심을 비판하며 이러한 범주들(평균, 열등성, 우월성 등) 속에 교육과 사회체제가 종속될 수 있음을 경고한 바 있다(Sacks, 2000: 73 재인용).

1950년대 중반에는 교육계에서 시험의 과도함에 대한 논의가 증가했다. 특히 1962년 물리학자 호프만(Banesh Hoffmann)의 <시험의 압제(The Tyranny of testing)> 이후 표준화시험에 대한 반대운동은 매우 활발하게 전개되었다. 호프만은 객관식시험의 객관성과 선다형시험을 비판했으며, 특히 SAT 등에 나타난 문항의 오류를 지적했다. 그는 표준화시험이 오히려 깊이 생각하는 천재를 가려내기보다는 차별한다고 비판한다(Hoffmann, 1962).

1970년대와 80년대에는 표준화시험에 대한 반대 분위기가 강했고 시험반대정서(Anti-test sentiment)도 강하게 나타나 지능검사에 대한 불신과 대학입시에서 표준화시험을 사용하는 것을 기피하는 경향도 나타났다.

반시험운동을 주도한 논의들은 다양한 저서나 보고서로(The mismeasurement of man, The case against SAT, The reign of ETS, Ralph Nader Report) 나타났다. 또한 시험을 반대하는 활동가들이 페어테스트(FairTest)라는 표준화시험반대를 위한 단체를 구성하게 된 점도 특기할 만하다.

표준화시험을 반대하는 논의들은 다양하다. 주요 비판 관점으로는 다음과 같은 것이 있다(Sacks, 2000). 첫째, 표준화시험은 학문적 성공에 대한 예측력이 의심된다. 둘째, 표준화시험은 사회경제적 지위와 상관이 높다. 셋째, 표준화시험은 수동적이고 피상적인 학습을 보상한다. 넷째, 교육을 바람직하지 않은 방향으로 유도한다. 다섯째, 의미있는 교육개혁을 방해한다.

이러한 표준화시험에 대한 비판은 1980년대 이후 대안적 평가(참평가, 수행평가, 포트폴리오 등)에 대한 강조로 나타났다. 선다형시험에 대한 대안이 모색되기 시작했으며 일부 주에서는 (켄터키, 버몬트, 캘리포니아) 선다형시험을 일시적으로 폐지하기도 했다. 일부 대학들도 1990년대 중반에 표준화시험을 입학사정에서 제외하는 등 표준화시험에 대해 반대하거나 경계하며 신중하게 접근하는 태도들이 강화되었다.

▫ 반동: 1990년대 이후 표준화시험의 재강화

1920년대 리프만(Lippmann)의 비판 이후 70년이 지난 후, <종형분포(The Bell Curve)>(Murray & Herrnstein)라는 책이 미국에서 논란을 야기시킨 바 있다. 이 책에 의하면 미국에서의 불평등은 서로 다른 지능을 타고난 결과라는 것이다. 이는 신자유주의 이후 미국사회에서 자유경쟁이 강조됨에 따라 불평등의 자연적 기원을 찾으려는 시도가 다시 발흥하는 것을 보여주는 것이다. 한편, 누가 능력이 있고 누가 그것을 가졌는가 하는 능력주의 관점에서 SAT 등의 표준화시험은 여러 비판에도 불구하고 핵심적 위치를 차지해왔다.

1990년대 이후의 반동적 상황은 차별철폐조치(Affirmative Action)에 대한 반대운동으로 전개되었다. 여러 주(California, Texas, Mississippi, Washington 등)에서 이러한 경향이 두드러졌다.

주요 주장은 인간은 능력에 의해 평가되어야지 젠더, 인종 등의 범주에 의해 평가되어서는 안 되며, 공정한 평가기준은 학점과 시험점수라는 것이다. 이러한 분위기를 배경으로 1990년대 여러 주에서 표준기반 시험이 도입되었다. 1982년에서 1994년까지 표준화시험지는 대학교재보다 훨씬 많이 팔린 것으로 보고된다(Sacks, 2000).

표준화시험에 대한 비판론자인 삭스는 미국사회에서 여러 비판에도 불구하고 왜 표준화시험이 계속 되는가라는 질문에 대해 효율성(비용)뿐 아니라 다음과 같은 이유들이 있다고 본다(Sacks, 2000).

첫째, 미국이 유난히 심리측정을 선호하는 경향이 있다는 것이다. 유럽에서는 수업에서 배운 것을 기초로 한 에세이시험이 일반적인 데 비해 미국인들은 IQ에 집착(논리분석력, 단일한 숫자)하는 경향이 있다. 둘째, 성취보다는 잠재성에 훨씬 비중을 두고 있다. 셋째, 다른 사회들에서는 사람들의 성취를 본다면, 미국은 지능검사에 비중을 두는 이례적 사회라는 것이다. 넷째, 능력과 성취는 곧 높은 점수와 등치된다는 인식이 대중문화 속에 있다고 본다. 이것이 표준화시험을 정당화하여 학생, 교사, 학교, 대학 등을 평가하게 한다. 다섯째, 대학들의 경제적 이익을 유지시켜준다. 대학랭킹과 SAT 점수 등을 연결함으로써 높은 점수는 학생의 수준을 재는 척도이고 대학의 질을 드러내준다고 믿게 한다. 여섯째, 엘리트들이 이러한 유사과학적 객관성을 통해 자신의 계급적 특권을 유지하고 있다.

4. 표준화시험과 책무성의 결합

▫ 1960년대 이후 대규모 표준화시험과 책무성의 연결

2000년대 이후 미국에서는 NCLB 등의 전격 도입으로 표준화시험과 학교에 대한 책무성(accountability)을 연결시키는 정책이 강한 힘을 발휘하고 있다. 대규모 표준화시험과 학교의 책무성을 연결시키는 시도는 이미 1960년대부터 시작되었다고도 볼 수 있다. 1960-70년대를 거치면서 시험이 정보제공으로부터 교사, 학생의 책무성을 묻는 것으로 변화한 것은 반세기동안의 가장 중요한 변화라고 평가된다(Koretz, 2008).

1960년대에 있었던 연방정부의 정책이 이를 변화시킨 것이다. 흑인인권운동의 결과로 시민권법(Civil Rights Act)이 개정되고 소득불평등시정에 대한 요구가 높아지면서 1965년 초중등교육법(ESEA)이 의회에서 통과되었다. 이로부터 타이틀원(Title I)으로 불리는, 저소득층 학생의 학력 향상을 위한 보상교육프로그램이 시작되었다. 이에 대한 지원을 위해 연방정부는 아이오와 테스트(Iowa test)와 같은 시험에서 전국기준에 의한 상대평가를 도입하도록 주들에 권고했다. 타이틀원 정책 자체는 시험정책도 아니고, 책무성정책도 아니었다. 그러나 이 정책의 시행에서 시험과 책무성정책의 연결이 이루어지기 시작한 것이다.

1974년에는 타이틀원에 대한 평가보고체제(TITLE 1 evaluation and reporting system, TIERS)를 만들어 표준화시험과 상대평가에 의한 성취평가로 학생점수를 평가하도록 하였다. 1960년대에 이미 연방정부는 표집평가인 '학업발달에 대한 국가수준 평가제도(National assessment of

educational pogress)'를 실시했다. 1990년대에는 학업성취도 결과가 상대평가로 언론에 발표됨으로써 주별 비교의 빈도가 높아졌다. 그러나 초기에는 주나 학군단위의 정보를 제공할 수 없었고 대부분의 결정이 평가 단위 내부에서 이루어졌기 때문에 책무성에 큰 영향을 주지는 못했다. 다만 정책결정자들에게 학생성취와 시기별 변화에 대한 정보를 제공하는 것에 목적을 둘 뿐이었다.

국가의 재정지원정책과 연동된 표준화시험의 실시는 책무성강화정책이기는 했으나 NAEP나 TIERS 모두 교사, 학생에게 평가결과를 가지고 직접 영향을 줄 수 있는 '고부담 평가'는 아니었다. 그럼에도 불구하고 이 두 가지가 변화의 기폭제가 된 것은 분명하다. 즉, 시험의 목적을 진단평가·지역단위평가로부터 대규모 모니터링·시험기반 책무성으로 전환하는데 기여했다는 점이다(Koretz, 2008).

▫ 최소역량평가운동의 영향

1970년대초에는 주단위 최소역량평가운동(minimum-competency testing movement)이 실시되기 시작하여 확대된 바 있다. 이는 1971년에 시작되어, 70년대말에 이르러서는 37개 주에서 실시될 정도로 확산되었다. 이는 학생들이 수용할만한 최저수준의 기본능력을 갖추었는가에 대한 평가였다.

이 평가는 대부분은 졸업시험(exit exam)으로 정해진 일정 점수(cut score)를 넘어야 고등학교 졸업장을 주는 것에 적용됐다. 일부에서는 이를 학년별 진급시험(promotional gates)으로 사용하기도 했다. 이 시험은 난이도도 낮고 커트점수도 낮았다. 아직도 일부 존속하지만 (New York City, Chicago 등은 진급시험) 1980년대 동안 대부분 사라졌다.

그럼에도 불구하고 최소역량평가운동은 4가지 정도의 영향을 미친 것으로 평가된다(Koretz, 2008).

첫째, 책무성평가에 시험을 활용하는 것이 일반화된 점이다. 둘째, 주단위로 이루어지는 시험이 전국적으로 확산되었다는 것이다. 1970년대 말, 60%의 주에서 주단위 시험프로그램을 실시하게 되고, 20년 후에는 거의 모든 주가 이를 갖추게 된다. 셋째, 상대평가가 아닌 목표치에 준한 평가(혹은 준거지향 절대평가, expectation. criterion-referenced tests)를 실시한 점이다. 넷째, 교수역량 향상에 활용되었다는 점 등이다. 이 시험의 결과가 교사의 책무성을 강화하기 위한 측정기반교수로 나아가는 기반이 되었다.

▫ 1983년 <위기에 처한 국가(Nation at Risk)> 이후 연방개입의 심화

1970년대 이후의 경제위기로 사회 재구조화에 대한 담론들이 형성되었고, 신자유주의와 신보수주의세력이 결합하는 신우파연합을 형성하게 된다. 경제사회적 위기에 대응하여 기업, 정치, 대중들로부터 학교교육을 비판하는 목소리들이 높아지고 이는 <위기에 처한 국가(Nation at Risk)>라는 보고서로 나타나게 된다.

이에 따라 교육예산이 팽창하고 대대적인 자유주의적 개혁이 시작되었다. 학교교육의 성과는 학교에서의 시험점수로 인식되었다. 사회는 표준화시험의 결과에서 실패의 증거를 찾고자 했으

며, 학교에게 책임질 것을 요구했다. 이후 시험이 매우 강조되는데, 단지 시험을 더 보라는 것이 아니라, 시험에 대한 책임을 지도록 하는 '책무성 운동'이 본격적으로 시작된 것이다.

교육에서 표준이 강화되었고 표준에 기반한 교육(standard-based education)이 시작되었다. 이러한 책무성운동의 직접적 대상이 된 교육계(AFT 등의 교원단체)는 크게 반발했고, 연방정부의 교육 개입을 우려하는 보수주의자들도 반대의 목소리를 드러냈다.

그럼에도 불구하고 일련의 논의들을 거쳐 책무성기제들의 법제화가 이루어졌다. 1994년 The Goals 2000이 의회를 통과했고, 미국 초중등교육법(ESEA)을 개정한 낙오자방지법(No Child Left Behind, NCLB)가 2002년 1월에 법제화했다. 이를 통해 공립학교에서 시험은 2배로 증가하게 된다.

NCLB는 초등 3-8학년과 중등 1개 학년을 대상으로 표준기반시험에 의한 수행표준을 정하고, 모든 학생이 12학년 동안 성취수준에 도달하게끔 하려는 것이다. 이를 위해 학생들이 수행목표를 향해 적정연간진보(adequate yearly progress, AYP)를 달성하도록 한다. 각 주별로 단일한 연간 목표를 정하도록 하며, AYP에 도달하지 못한 학교는 제재를 받게 되고, 미도달이 지속되면 강한 제재를 받는다. 2008년 이후 오바마정부에서도 'Race To The Top' 정책을 통해 공통핵심(common core)을 강조하고, 학생점수를 이용한 교사평가를 강화하며, 차터스쿨 등을 확대하는 등 책무성 정책을 이어가고 있다.

이러한 양상은 전반적으로 볼 때 학교내에서 시험에 대한 의존을 강화하는 것이다. 학생의 학업능력제고에 있어 최저역량이 아니라 보다 힘든 시험을 부과하게 된 점도 특징적이다. 교사와 학교에 대한 제재가 가해졌고 학교에 대한 재정적 차등보상이 초기에는 일부 주에 국한되었으나 10년을 지나면서 일반화되기 시작했다. 책무성강화와 함께 나타난 이러한 고부담평가의 영향으로 1980년대말부터 점수 인플레 현상이 심각한 현상으로 나타났고 각종 시험부정 사례도 빈번하게 지적되어왔다. 이는 책무성강화와 함께 나타난 가장 큰 문제라고 평가되기도 한다(Koretz, 2008).

□ 대규모시험, 고부담평가, 책무성

평가학자인 코레츠(Koretz)에 의하면 미국의 대규모시험들은 다음과 같은 특징을 공유한다. 첫째, 외부기관에 의해 부과된 외부 시험이라는 점이다. 둘째, 대부분 표준화시험(내용, 행정, 점수 공통적 적용)이라는 점이다. 셋째, 학교나 주의 성과를 모니터링하고 향상 추이를 확인한다는 점이다. 넷째, 교사, 학생 모두에게 근본적으로 '고부담'이라는 점이다. 이러한 변화는 최근 몇십년간 나타난 것으로 대규모시험의 기능에서 큰 변화가 나타난 것이라고 지적된다. 즉, 평가가 학생개인의 학습 강약점을 확인하는 것으로부터 책무성 점검을 하는 것을 목적으로 바뀌어졌다는 점이다. 이는 시험성적을 해석하는 결론에 대한 차이에도 반영된다. 첫째, 학생개인에 대한 추론의 비중이 약해진다. 현재도 그 중요성이 없어진 것은 아니나, 과거에 학생개인의 '형성'에 대한 관점이 더 강하게 자리잡고 있었다. 둘째, 현재는 개인의 형성보다는 학교, 학군이 어떻게 수행하는가에 대한 관심이 높다는 점이다. 셋째, 시험을 본 당시에 관한 것보다 시간적 경과를 중시한다는 점이다. 넷째, 학생간의 비교가 아니라 정책결정자들의 기대에 학생성적결과가 얼마나 부합하는지 비교하는 것이 목적이라는 점이다.

평가의 목적이 학생 개인보다는 모니터링을 위한 것으로 변화하는 과정에서 작동하는 가정들을 삭스는 '신화'라고 표현하며 다음과 같이 주목하고 있다(Sacks, 2000).

- 신화1 미국의 학교들은 위기에 처해 있다
- 신화2 미국경제는 열등한 학교들로 인해 위기에 처해 있다
- 신화3 책무성이 높은 교육은 학업성취가 높을 것이다

삭스는 1990년대(1995-96) 주별 8학년 학업성취도를 비교한 데이터를 통해 고부담시험을 적용하는 주, 중약부담을 적용하는 주, 주단위 시험이 없는 주(Colorado, Iowa, Minnesota, Nebraska)를 비교한 바 있다. 결과는 흥미로운데, 고부담시험을 적용하는 주일수록 전국 평균 미달비율이 높았고, 시험이 없는 주는 모두 평균 이상이었다. 이는 학업성취가 낮은 주들이 고부담시험을 적용하는 경향이 있지만, 고부담시험 자체가 학업성취를 높이는데 별로 기여하지 못하고 있다는 점을 보여주는 것이다.

□ 절충과 평가 경향의 변화

20세기 초 이래 심리측정의 발달, 객관적 시험에 대한 의존 강화는 1960년대 이후 진보적 신사회운동의 흐름과 맞물려 시험반대운동을 야기할 정도로 문제가 되어 왔다. 1960년대에서 1980년대까지 반시험운동이 교육계 내외부에서 전개되는 동안 정책적으로는 오히려 책무성 기제로 대규모시험을 보다 적극적으로 강화해온 역설이 나타났다.

한편으로는 대규모시험의 활용도가 높아지고 다른 한편에서는 평가의 비교육적 성격이 비판받는 가운데 평가에서는 다양한 절충과 변화가 나타났음을 주목할 수 있다(Koretz, 2008).

첫째는 선다형 문제에 대한 지적이 높아지면서 선다형시험을 대체한 시험형식이 도입되어 왔다. 이는 학교 시험과 표준화시험에 모두 적용되어 왔다. 학교시험에서는 1980년대말 이후 수행평가, 참평가 등을 도입하여 외부 평가의 한계를 극복하고자 하였다. 수행평가와 참평가는 현재까지도 표준화시험에 대응하는 가장 강력한 교육적 방안으로 논의되는 경향이 있다.

또한 시험과 교수를 의미있게 연결하려는 시도도 이루어져왔다. 시험을 위한 교수와는 또 다른 의미로, 의미있게 평가될 수 있는 것을 가르치고(test worth teaching to) 단편적 문제들을 벗어난 교수로 유도(what you test is what you teach)하는 형식들이 논의되어 왔다.

표준화시험에서도 수행평가의 요소를 도입하는 문제들이 포함되어왔다. 그러나 이를 전면적으로 실시하는데 따른 난점을 극복하기 위해 NAEP (TIMSS등에서도 사용) 등은 매트릭스표집평가(matrix-sample assessment) 등을 도입하여 시험을 몇 개 영역으로 나누어 서로 다른 과제를 부과하기도 한다. 그러나 이는 결과해석에 있어 학생들을 비교하기 위한 표준화가 아니라 학교간, 주간 비교를 위한 것이 되기 때문에, 절충형이 도입되기도 한다. 예컨대 매사추세츠 종합평가체제(Massachusetts comprehensive assessment system)에서는 일부는 공통, 일부는 매트릭스 샘플을 사용한 시험을 본다.

둘째는 표준의 강화이다. 수행평가 시대와 동시기에 이루어진 평가 개혁 중 하나는 표준기반

혹은 표준준거 평가의 증가이다. 즉 평가가 내용표준에 부합해야 하고 학생의 수행표준 도달도를 확인해야 한다는 것이 강조된다. 이는 전통적 시험으로부터 크게 벗어나는데, 시험(test)이라는 말의 사용을 거부하고 평가(assessment)를 선호한다. 그러나 내용표준(Content standard)이 전통적 시험에서 교육과정준거(curriculum framework)와 실제로 유사하다는 점에서 그 차별성이 의문시되기도 한다. 이러한 내용표준은 각 주에 고유하게 설정되어 있다. 이로 인해 각 주들은 서로 다른 시험을 적용해야 한다는 부담을 가지며, 이는 때때로 질 낮은 시험문제로 나타나기도 한다. 각 주별로 표준기반평가를 위한 작은 평가회사들이 다양하게 존재하고 있다.

셋째, 가치부가접근(value-added approach)이 도입되어 왔다. 이는 학생들의 진급에 따라 평가하는 것이다. 1990년대에 가장 보편적 방식은 동학년 학생들을 전 해와 비교하는 것이었다. 그러나 얼마나 향상되어야 '적정' 한가의 문제가 있다. 이 수준을 NAEP(National Assessment of Educational Progress)의 능숙 수준(Proficient level)에 도달하는 것으로 정해야 한다는 의견도 있지만 8학년 수학의 경우 1/4의 학생만이 도달할 뿐이다.

넷째 1990년대초 이후 배려대상학생을 지원하려는 노력이 강화되고 있다. 이들은 기존에는 시험대상에서 제외되었으나, 현재는 연방차원에서 법제화되어 있다(Individuals with Disabilities Education Act Amendments of 1997, No Child Left Behind Act 2001).

다른 한편, 책무성에 대해 다시 봐야 한다는 논의도 제기된다. 즉 이를 긍정적으로 참평가 관점에서 재규정하여 학생의 이해, 배움, 참여를 독려하는 책무성으로 전환해야 한다는 입장도 있다(Sacks, 2000).

5. 교육과정과 학생평가 패러다임

평가에 대한 관점은 교육과정이론과 밀접하게 연관되어 있다. 표준화시험을 둘러싼 논란 중 하나는 교육과정 및 교수이론은 새로운 방향을 추구하고 있으나 시험은 과거의 방식을 답습하는 경향이 있다는 것이다(Shepard, 2000: 4). 셰퍼드에 의하면 20세기 동안 지배적 교육과정 패러다임은 사회적 효율성주의에 입각한 과학적 경영론이었으며 이는 개인차에 대한 유전이론(IQ)과 연합주의 및 행동주의학습이론과 긴밀하게 연관되어 있었다. 이러한 심리학적 이론들은 능력과 성취에 대한 과학적 측명과 연결되었다.

20세기초 미국에서 사회적 효율성운동은 과학이 산업화와 도시화로 인한 문제를 해결해줄 것이라는 신념에서 힘을 얻었다. 산업현장의 효율성을 높이기 위해 사용된 근대과학의 원리들이 학교에도 적용되어야 한다는 관점들이 나타났다. 학생의 진로에 직접 대응한 교육(Taylor), 기능 숙달을 위한 정확한 측정 표준 설정(연합주의 심리학) 등이 교육에서 중시되었다. 사회적 효율성운동의 대표자인 보빗(John R. Bobbitt)은 낭비의 제거(1912)를 강조했는데, 그는 모든 사람에게 각자가 평생동안 전혀 쓰지 않을 것을 가르치는 교육과정은 낭비라고 보았다. 따라서 각자가 자신의 능력에 따라 교육을 받아야 한다는 것이었다. 이러한 관점은 고도로 분화된 교육과정을 낳게 했으며 학문적 교과는 대학을 가는 학생에게만 가르쳐지도록 하는 실용적 접근을 주장했다.

한편 손다이크(Thorndike) 등의 연합주의(1922), 헐(Hull)(1943), 스키너(Skinner)(1938, 1954), 가네(Gagne)(1965) 등의 행동주의는 학습을 자극-반응의 연합이라고 보았고 이에 따라 행동주의적 모델의 교수-평가를 개념화했다.

이러한 관점에 따르면, 1) 학습은 작은 단위로 분화된 지식의 누적에 의해 발생한다. 2) 학습은 긴밀하게 순차적이고 위계적으로 이루어진다. 3) 전이는 제한적이기 때문에 각 목표는 명시적으로 가르쳐져야 한다. 4) 다음 단계 목표로 나아가기 전에 숙달도 확인을 위해 시험의 활용 빈도를 높여야 한다. 5) 시험은 학습과 동형적이다 (시험=학습). 6) 동기는 외재적인 것이며, 다수의 작은 단계들로 구성된 긍정적 강화에 기초한다(Shepard, 2000: 5).

연합주의학습이론의 주창자이면서 과학적 측정주의의 아버지라고 불리는 손다이크는 미국 성취평가의 특징인 소위 객관식 시험의 발달에 기여했다. 초기 객관식 문항들은 단순 기억에 의존하고 있었다. 셰퍼드에 의하면 이는 시험이 교과지식을 왜곡하는 것이라기보다는 무엇이 학습되어야 하는가에 대한 전통적 생각들을 반영하고 있는 것이라고 봐야 한다. 객관식 시험 문화는 교사들의 생각을 지배하게 되어 홀리스틱하게 학생의 활동을 평가하는데 있어서의 주관성에 대해 부정적 인식을 주는데 기여한 측면이 있다(Shepard, 2000: 5-6). 아래는 20세기 초 미국에서 실시된 객관식 시험의 사례이다.

<그림 Ⅲ-1> 20세기초 미국 객관식 시험의 사례

New Stone Reasoning Tests in Arithmetic (1908)

1. James had 5 cents. He earned 13 cents more and then bought a top for 10 cents. How much money did he have left? *Answer*: _____

Sones-Harry High School Achievement Test, Part II (1929)

1. Write "25% of" as "a decimal times." .....(_____)
2. Write in figures: one thousand seven and four hundredths.....(_____)

The Modern School Achievement Tests, Language Usage

1. I borrowed a pen a. off b. off of c. from my brother. ____

The Barrett-Ryan Literature Test: Silas Marner

1. Dolly Winthrop is:
   a. an ambitious society woman.
   b. a frivolous girl.
   c. a haughty lady.
   d. a kind, helpful neighbor.

Examples of True-False Objective Test (Ruch, 1929)

1. Tetanus (lockjaw) germs usually enter the body through open wounds. *True* *False*

American History Examination, East High School (Sam Everett and Effey Riley, 1928)

I. Below is a list of statements. Indicate by a cross (X) after it, each statement that expresses a social heritage of the present-day American nation.
Place a (0) after each statement that is not a present-day social heritage of the American nation.
1. Americans believe in the ideal of religious toleration. ___
2. Property in land should be inherited by a man's eldest son. ___
3. Citizens should have the right to say what taxes should be put upon them. ___

II. To test your ability to see how an intelligent knowledge of past events helps us to understand present-day situations and tendencies.
(Note: Write your answer in essay form on a separate sheet of paper.)

State your reasons for every position assumed.

4. Take some *economic* fact or group of facts in American History about which we have studied and briefly show what seems to you to be the actual significance of this fact in the past, present and future of America.
5. Show this same *three-fold relationship* using some *political* fact or facts.
6. Show this same *three-fold relationship* using a *religious* fact or facts.

FIGURE 3. *Examples from some of the earliest 20th-century "standard" tests and objective-type classroom tests. Note: The first four examples are borrowed from Ross (1941); the last two, including the Everett-Riley American History Examination, appeared in Ruch (1929).*

(출처: Shepard, 2000: 7)

인지적 혁명을 통해 정신(mind)이라는 개념이 다시 도입됨으로써 정신적 구성과 의미 생성의 절차를 통해 지식이 획득되는 것을 이해할 수 있게 되었다. 비고츠키(Vygotsky)의 재발견은 정신이 사회문화적으로 결정된다는 것을 확인해주었다. 따라서 기존의 유전이론적 지능에 대한 접근은 인지능력이 사회적 상호작용을 통해 발달된다는 것으로 대체되었다.

이러한 학습이론의 변화는 교육과정변화에도 영향을 주었다. 모든 학생이 학습가능하다는 생각, 학습은 실생활 영역에서의 지식활용과 연계되어야 한다는 생각 등으로 관심이 전환되었다. 또한 인지능력뿐 아니라 학교의 기대나 규범 등을 통해 학생들이 어려운 과제에 대해 도전하고자 하는 의지를 갖게 하는 것이 중요하다는 점 등이 강조되었다.

이러한 흐름에 따라 평가는 다음과 같이 변화되어야 한다. 첫째, 시험의 내용과 형식이 교과에서 필요로 하는 사고 및 문제해결능력을 반영해야 한다. 둘째, 평가가 교실에서 활용되고 인식되는 방식이 바뀌어야 한다. 후자를 위해서는 교사들이 외부평가에 의해 교수과정이 왜곡되고 탈동기화되는 것들을 방어해야 한다(Shepard, 2000: 7).

학습을 위한 평가를 하기 위해서는 책무성평가의 부정적 효과를 직시해야 한다. 고부담책무성 평가는 교사들을 탈숙련·탈전문화한다. 시험의 긍정적 효과에 대한 논의도 있지만, 모든 시험의 효과가 부정적일 수 있으며 교수에 악영향을 미칠 수 있다는 논의도 있다(Whitford & Jones, 2000).

셰퍼드는 20세기의 지배적 패러다임이 사회적 효율성 교육과정, IQ 유전이론, 과학적 측정을 중심으로 구축되어 오다가 1980-90년대를 거치면서 새로운 교육과정 관점, 학습이론(인지이론, 구성주의) 및 평가이론(교실평가)에 의해 도전받고 있다고 본다. 이를 도식화한 것이 아래 <그림 III-2>, <그림 III-3>, <그림 III-4> 등이다.

<그림 III-2> 20세기 교육과정, 학습이론, 평가이론의 지배적 패러다임 변화

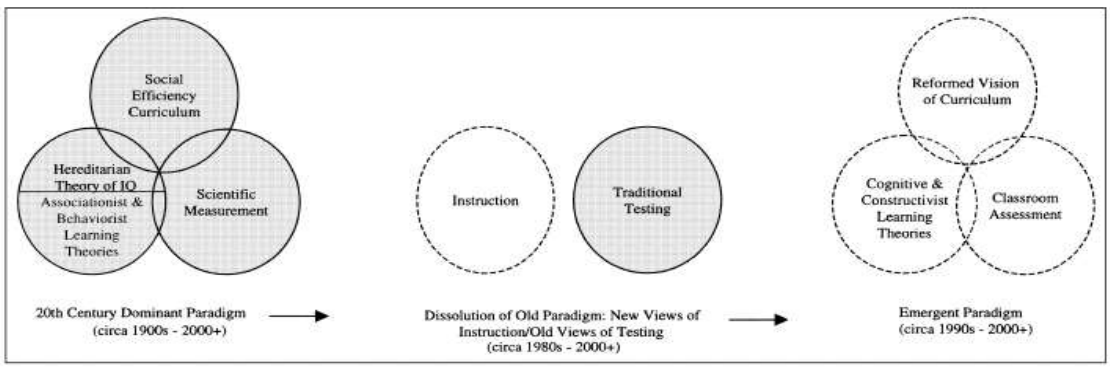


FIGURE 1. *An historical overview illustrating how changing conceptions of curriculum, learning theory, and measurement explain the current incompatibility between new views of instruction and traditional views of testing.*

(출처: Sphepard, 2000: 5)

<그림 III-3> 20세기의 지배적 패러다임

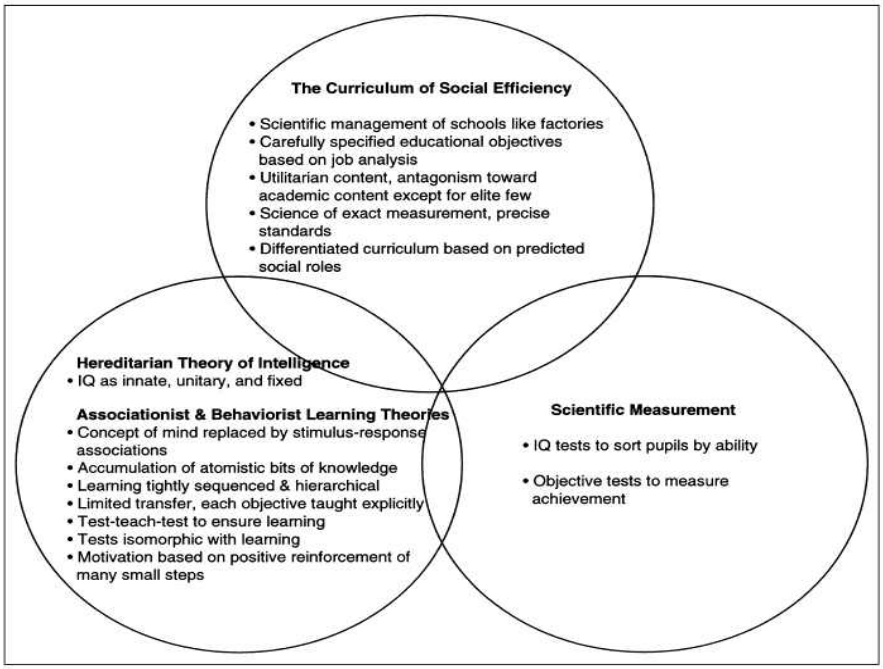


FIGURE 2. *Interlocking tenets of curriculum theory, psychological theories, and measurement theory characterizing the dominant 20th-century paradigm.*

(출처: Sphepard, 2000: 6)

<그림 III-4> 새로운 구성주의적 패러다임

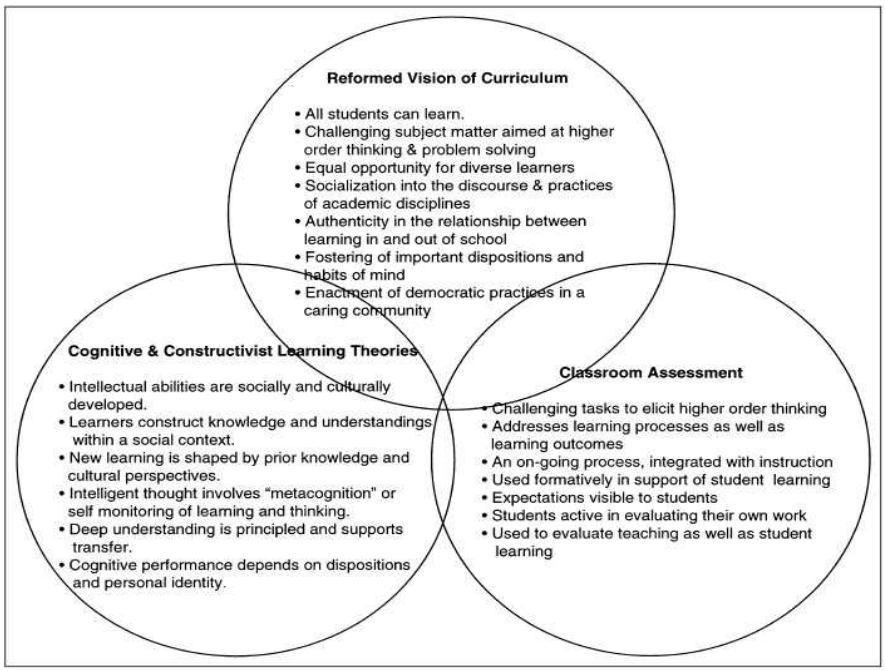


FIGURE 4. *Shared principles of curriculum theories, psychological theories and assessment theory characterizing an emergent, constructivist paradigm.*

(출처: Shepard, 2000: 8)

셰퍼드는 학습을 돕는 평가를 위한 평가의 활용과 전략에 대해 다음과 같이 제안한다(Shepard, 2000: 10-12):

1) 학습문화를 형성해야 한다. 교수학습과정과 분리된 평가가 아니라 교사와 학생이 다음 단계 학습 이전에 공통적이고 가치로운 프로젝트가 될 수 있도록 학습문화를 형성하는 것이 중요하다.
2) 역동적 평가가 이루어져야 한다. 학생이 스스로 할 수 있는 것과 성인의 도움을 필요로 하는 것이 무엇인지 확인하는 과정에서 평가와 동시에 지원이 이루어지도록 한다.
3) 개방적 토론이나 교육적 대화 등을 통해 선행지식을 활용한다. 이를 통해 학생들의 추론과 경험에 대해 알 수 있고, 다양한 문화로부터 어떤 자원들을 가지고 왔는지를 알 수 있다.
4) 학습자에게 제공되는 피드백은 자기교정과 개선을 유도할 수 있다. 피드백의 방식도 행동주의적 방식보다는 비계설정과 전문 튜터링을 통해 이루어질 필요가 있으며, 학습자의 동기를 유지하고 오류의 자기교정을 중심으로 해야 한다.
5) 개념에 대한 진정한 이해와 지식의 전이는 긴밀한 관계가 있다. 학습을 통해 적용하는 기회가 많을 경우 지식에 대한 전이능력이 높아진다는 연구가 있다. 좋은 교수는 탄탄한 이해를 위하여 기존의 이해에 대해 새로운 방식으로 질문함으로써, 새롭게 적용하고 새로운 연결점을 찾는 것이다. 좋은 평가도 마찬가지이다. 익숙하고 연습된 시험만이 공정한 시험이라고 보는 입장을 넘어서야 한다.
6) 학습자들은 평가에 대해 명확한 이해를 가지고 있어야 한다. 교사가 하는 평가와 동일한 방식으로 학생이 자신의 학업에 대해 평가를 할 수 있을 정도로 투명한 기준이 필요하다는 것이다. 자신의 활동이 어떻게 평가되는지의 기준을 아는 것이 공정성의 원리에 부합하는 것이다.
7) 학생의 자기평가는 인지적 목적도 있고 학생이 자신의 학습에 대한 책임을 갖도록 함으로써 교사와 학생간의 보다 협력적 관계를 형성할 수 있다.
8) 학생에 대한 평가는 교수행위를 개선하는데도 활용되어야 한다. 학생의 이해를 지속적이고 비공식적으로 평가하는 것을 통해 수업계획을 조정할 수 있으며 비판적 실천연구의 기반으로 삼을 수 있다. 수업활동에 대한 교사의 탐구를 학생에게 가시화시킴으로써 교수방향을 전환하는데 있어서 의사결정을 함께 할 수 있다. 학습의 문화를 형성하기 위해서는 학생이 피드백을 구하고 자신의 학업에 대해 평가해야 함과 동시에 교사도 같은 취지로 자신의 교육활동을 돌아보고 데이터들을 체계적으로 적용해야 한다.

6. 평가일반과 평가학의 전개

학교에서의 학생평가(testing, assessment)와 달리 평가일반(evaluation)을 사회공학의 기제로 사용하는 경향은 19세기부터 등장해왔다. 19세기 초부터 미국정부는 감옥, 학교, 병원, 고아원 등의 공적 시설에 대해 외부감사를 실시하도록 해왔다(Mertens, 2012). 제1차 대전 이전, 사회과학자들은 교육, 공공보건, 고용 등과 관련한 주요 사회프로그램 평가를 수행했다. 1950년대에

는 사회과학에 기반한 프로그램평가가 확대되어 일탈방지, 공공주거, 국제활동 등을 포괄했고, 1960년대까지 빈곤과의 전쟁은 당시 사회프로그램을 평가하기 위한 체계적 접근에 기여했다. 1980년대 이후 본격화된 책무성평가는 신자유주의정책 출현에 의한 평가적 국가(evaluative state)로의 전환을 반영하며 이러한 평가적 문화(audit culture)가 일반화되는 것을 보여준다.

평가가 전문화된 것은 1960년대부터라고 알려져 있는데, 이는 앞서 언급한 초중등교육법(ESEA), 헤드스타트(Head start) 등 연방지원 프로그램에서 평가를 프로그램의 일부로 보게 되면서부터라고 인식되고 있다. 평가(evaluation)는 이원적인 학문적 기원을 지니고 있는데, 교육적 평가(testing, assessment, objective-based evaluation background) 및 심리학 전통(응용사회과학)이 그것이다(Mertens, 2012).

이러한 평가학의 흐름은 그 패러다임 변천과정을 중심으로 구분된다. 구바와 링컨(Guba & Lincoln, 1989)은 <제4세대 평가(Fourth generation evaluation)>에서 평가의 역사를 4단계로 구분한다. 제1세대가 측정을 중심으로 하고 학생에 대한 시험 평가 위주였다면, 제2세대는 기술(description)적 단계로 타일러의 목표 모형이 이를 대표한다. 제3세대는 판정(judgment) 중심의 의사기반모델(Stake, 1983; Scriven, 1967a, Stufflebeam 1982)로 발전되어왔다. 제4세대 평가는 구성주의적이고 해석적 패러다임에 기초한 평가이다. 4세대평가에서의 패러다임변화는 양적 평가에서 질적 평가로, 평가자 중심에서 소통에 기반한 평가를 반영하는 것으로, 평가학 내부에 다양한 관점들이 존재하고 있음을 보여준다(Guba & Lincoln, 1989).

평가학 및 평가문화의 발달로 평가관련 단체, 학회, 국제네트워크도 상당히 활성화되어 있다. 1976년 미국에서 평가와 관련한 2개의 협회가 등장했는데, 평가연구학회(Evaluation Research Society)와 평가네트워크(Evaluation Network)가 그것이다. 이들은 1986년 AEA로 통합되어(www.eval.org) 6000명의 회원을 확보하고 있다. 이 조직에는 미국 50개주와 전세계 80개국 회원이 망라되어 있으며, 회원은 대학, 기업, 컨설팅회사, 정부기구, 교육 및 사회조직 등 다양한 활동영역에 소속되어 있다. 이들은 연례회의 및 온라인상의 리스트서브(listserv- EvalTalk)를 운영하며, 2개의 저널을 발간(The American Journal of Evaluation, New Directions for Evaluation)하고 있다. 전 세계적으로도, 1995년 이전까지 전세계적으로 5개의 광역(지역, 국가수준) 평가조직이 존재할 뿐이었으나 2003년 3월 이후에는 국제조직이 창설(International Organization for Cooperation in Evaluation (IOCE: ioce.net))되어 70개가 넘는 전세계 협회들의 리스트를 제공하고 있다(Mertens, 2012).

Ⅳ. 미국에서 표준화시험을 둘러싼 논란: 옹호와 저항

1. The tyranny of testing(1962) – 표준화시험 비판의 고전

미국 표준화시험의 역사에서 일종의 고전처럼 언급되는 책이 호프만(Hoffmann)의 <시험의 압제(The tyranny of testing, 1962)>이다. 1962년에 쓰여진 이 책은 직접적으로는 ETS에 의해 실시되던 시험들을 비판한 것이다. 저자인 호프만은 물리학자로 언론 매체인 Harper, The American Scholar 등에 기고하면서 객관식 시험(objective tests)에 대해 강하게 반대했다.

제2차 대전 이후 미국에서 시험은 하나의 산업으로 발전했다. 진위를 체크하는 시험이 전통적인 에세이시험을 대체하여 사회의 전 분야에 적용되었다. 그러다 보니 제2차 대전 이전 몇몇 논자들(Jacques Barzun 등)이 시험에 대해 반대하는 글을 쓸 때는 낡은 주장이라고 아무도 관심을 갖지 않던 것이 1960년대에 이르러서는 시험제공자들이 수세에 몰릴 정도로 시험에 대한 비판적 관점이 확산되었다. 당시 외국(영국, 독일, 프랑스 등)은 모두 아동중심적인 교육을 추구하는 반면 미국에서만 시험과 학력이 강조되고 있다는 비판이 제기됐다(Hoffmann, 1962).

호프만은 표준화시험이 개성을 억압하고 학생의 수행에 대해 오독하게 한다고 비판한다. 그리고 더 큰 문제는 정말 우수한 학생을 가려내지도 못할 뿐 아니라 독창성을 가진 학생을 잘못된 시험의 형식으로 인해 실패시키는 것이라고 본다.

그는 1959년에 장학기금인 National Merit Scholarship Corporation (J M Stalnaker가 대표)이 장학생 선발(지원자의 2%)을 하면서 객관식 시험점수만으로 적임자를 가려낸 방법에 대해 비판하며 논쟁을 제기한 바 있다. 객관식 시험을 통해서는 시험에 유리한 능력만을 가려내지 심오함이나 독창성에 대한 평가를 하기 어렵고, 그렇기 때문에 그런 특성을 가진 학생들에게 불리한 점이 지적되었다. 호프만에 의하면 아인스타인(Einstein)처럼 너무 깊이 생각하는 사람은 최종 선발의 대상이 되지 못한다. 또한 그는 객관식 시험의 결과로 아인스타인같은 천재가 100명 중 10위를 하는 경우보다 더 문제가 되는 경우는 잘못된 선다형 문항으로 인해 정신박약아(moron)가 80위를 하게 되는 경우이다. 후자는 더 잘 못된 배치라는 것이다.

호프만은 통계에 대한 맹목적 신뢰가 경계되어야 한다고 본다. 통계를 통해 밝혀내지 못하는 질적인 부분들이 있다는 것이다. 시험은 자신의 머리로 독창적으로 사고하고 행동하는 창의적 사람보다는 약삭빠르고 평이하며 순응적인 사람에게 유리하다고 말한다. 보다 치밀하게 사고하는 사람들에게는 오히려 불리하다는 것이다. 특히 선다형 문항은 고르기만 하고 왜 그 답을 골랐는지 설명을 할 수 없게 되어 있어 문제라고 지적한다.

호프만은 그의 책에서 SAT등 각종 표준화시험의 문항에 나타난 문제들을 지적하는 것에 많은 지면을 할애한다. 그는 문항들이 인위적으로 난해하게 만들어져 있고 의미를 모호하게 하여 풀기 어렵게 되어있다고 비판한다. 그는 잘못되거나 좋지 않은 문항의 사례들을 예시하면서, 이에 대한 시험출제자들의 변명을 정면으로 비판한다.

호프만은 문항의 문제점을 지적할 뿐 아니라 테스트회사인 ETS와 싸워서 해명을 받아내기도 하고 방송출연 등을 통해 문제를 공공연하게 확산하는데 기여했다. 시험 문항의 문제점을 지적

하기 위해 동일한 문항(외교관시험)을 여러 전문가들에게 테스트해보고 서로 다른 답이 나온 사례를 제시하기도 한다. 시험들이 객관성을 주장하고 있지만 모호한 문제들 투성이로 전문가들도 서로 다른 응답을 할 수밖에 없다는 점을 폭로한 것이다.

호프만은 그의 책 말미에서 테스트의 옹호자가 되어서는 안 되고 저항해야 한다는 말을 '테스트(test)'라는 용어를 넣어 재치있게 제시한다. "시험을 옹호하지 마라(Don't be pro-test), 저항하라(protest)." 그는 당시 ETS의 권위가 막강하다는 점을 지적하며, 이 기구를 집중적으로 공격했다. 특히 고등사고를 재지 못하는 객관식시험과 이에 연관된 통계기법에 대해 비판을 가했다. 그는 객관식문제의 출제자들이 거만하며, 과학적인 양하고, 통계기법을 연막으로 사용하고 있다고 지적한다. 이들이 객관성이나 과학성을 가정하면서 특권적이고 배타적으사용하지 못하도록 경계해야 한다고 주장한다.

호프만에 의하면 선다형문제는 아무리 잘 만들어도 본질적으로 결함이 있기 때문에 대안적 평가체제가 만들어져야 한다고 주장한다. 고등능력을 가려내는 방법은 객관식 · 선다형 방식이 아닌 보다 개별적인 방식으로 접근돼야 하며, 평가에서 사고의 깊이(depth)와 정교함(subtlety)이 고려돼야 한다고 강조한다. 다른 나라들(덴마크, 영국, 프랑스, 러시아)에서 에세이시험을 광범하게 활용하는 사례들을 제시하며 대안을 찾을 것을 제안한다.

2. 찬반 양론의 대립

표준화시험에 대한 찬반 양론은 매우 팽팽하다. 심지어 동일 사안에 대해 상반된 관점이 팽팽히 대립하고 있기도 하다. 빌리(Bily, 2011) 등은 학습, 교사, 시험의 미래 학업 예언능력(대학입학 관련) 등의 유사한 문제에 대한 찬반 양론의 대립적 시각들을 대비하여 다루고 있다. 즉 '표준화시험은 학습을 돕는가?,' '표준화시험은 교사들에게 유익한가?,' '표준화시험은 대학입학을 위해 가치가 있는가' 등에 대해 서로 다른 관점들이 어떻게 대비되는지 드러내고 있다. 각각의 시각을 도표화해보면 아래 <표 Ⅳ-1>, <표 Ⅳ-2>, <표 Ⅳ-3>과 같다.

<표 Ⅳ-1> '표준화시험은 학습을 돕는가'에 대한 찬반 양론

영역	관점	주장 및 근거
학습: 표준화시험이 학습을 돕는가?	표준화시험은 학생의 필요를 파악하는 것을 도와준다 (Karen Tankersley, 아리조나주립대 교수, 교사, 행정가 출신)	- 사람들의 상식과 달리 표준화시험은 복합적 사고를 요구함 - 좋은 교사라면 사실의 암기가 아닌 복합사고기술을 가르칠 수 있음 - 교사에게서 잘 배운 학생이라면 시험에서도 성공적 - 학생의 낮은 성취는 교사들에게 학생의 부족한 부분에 대해 진단 가능하게 함
	표준화시험은 학생의 학습을 제한한다	- 표준화시험은 대부분 한 개의 정답을 요구하기 때문에 교사들로 하여금 복합사고보다는 지식암기에

	(Helen Cole 외, 테네시 링컨메모리얼대학의 교사교육과정 교수진)	중점을 두도록 함 – 표준화시험은 교육과정을 협소화시키고 표준화시킴. 학교가 가르칠 수 있는 범위를 제한시킴 – 학교가 학생의 수행에 대해 보다 정교하게 평가할 수 있는 전략을 개발할 수 있도록 해야 함
	표준화시험은 저소득층의 교육을 개선했다 (Education Equality Project: 인종적 격차해소를 위해 관료, 시민운동가, 교육개혁운동가 등으로 구성된 단체)	– 표준화시험은 미국에서 새로운 수준의 공정성과 평등성을 제공하고 있음 – NCLB 법으로 인해 교사, 학교, 지역, 주 등에서 평등교육이 가능해짐. 실제 개선이 나타남(격차 줄고, 중도탈락율 낮아짐) – 1994년 이후 책무성 조치의 불명료성 제거됨. *1994년 ESEA개정(표준설정, 연례평가도입, 연간 진보확인), 2002년 NCLB (지역별 집단별 성취데이타 분석, 2014년까지 교육격차해소에 대한 지역행정기관의 책임, 3–8학년 학생에 대한 일기 및 쓰기 시험 매해 실시)
	표준화시험은 저소득, 소수층 학생에게 해롭다 (Monty Neill, National Center for Fair and Open Testing의 대표, 1987년 이래 공립학교에서의 시험에 대해 저술활동)	– 저소득층 학생들에게 고부담평가는 해롭다고 지적함 – 창의적이고 비판적 사고를 갖도록 교육하지 않으며, 학생 개개인의 필요를 반영하지 못함 – 저소득층 지역에는 교사수준이나 교육환경이 전반적으로 취약함 – 부유한 학생들과 동등하게 시험을 보도록 요구되지만 시험을 위한 준비과정이 동등하지 않음. 저소득층지역에서는 점수 높이기를 위해 다양한 활동시간을 줄이지만 부유층지역에서는 시험준비자체를 위한 교육과정협소화는 적게 나타남 – 이들의 낮은 점수는 대학이나 취업에 성공하기 어려움을 보여줌
	표준화시험은 장애학생의 교육을 개선한다 (Advocacy Institute, 장애인 삶개선을 위한 비영리단체)	– NCLB로 인해 장애인들의 교육에 대한 책무성규정이 강화됨 – 장애학생들도 표준화시험을 통과해야 하기 때문에 동등한 교사, 교육과정을 누릴 권리 갖게 됨
	고부담시험은 장애학생에게 걸림돌이 된다 (Candace Cortiella, 장애아동부모 컨설팅)	– 고교졸업의 전제로 치러야 하는 고부담평가는 장애학생에게 해로움 – 이 시험 통과를 위한 과정에서 유급되거나 중도탈락하는 사례 이어질 수 있음 – 학교들은 이 학생들이 시험에 대해 준비할 수 있도록 도와주어야 함

<표 Ⅳ-2> '표준화시험은 교사들에게 유익한가'에 대한 찬반 양론

교사: 표준화시험은 교사들에게 유익한가?	좋은 교사들은 시험을 준비시키는 교사들이다 (Walt Gardner, 고교 영어교사출신, 대학강사, 칼럼니스트)	- 표준화시험에서 요구하는 기능과 지식을 가르치는 것은 교사들에게 중요함 - 시험과 관련되지 않은 내용으로 수업의 상당 부분을 할애한다면 교사로서 무책임 - 교사는 시간과 주의를 어디에 기울여야 하는가에 있어 많은 요구에 직면하게 되는데, 이를 조정할 수 있는 가장 적절한 방법이 시험을 준비하는 것
	시험을 위한 교육은 교육을 약화시킨다 (Judy Willis, 신경학자, 중학교교사, 뇌과학기반 교수전략 권위자)	- NCLB로 인한 시험준비교육은 맥락과 개연성을 벗어난 지식/개념에 대한 단순 암기를 증가시킨 경향이 있다고 지적 - 그러한 교수방식은 뇌가 원하는 새로움이나 즐거움의 단서를 제공해주지 못하고 주의를 이끌지 못함 - 따라서 학생은 흥미를 잃고 지루해짐. 뇌는 새로운 자극의 원천을 찾게 됨 - 이러한 해독적 효과는 표준화시험을 통해 심화됨. 이러한 해독성은 학생의 삶, 흥미, 경험 등과 연관된 교육을 통해 대체되어야 함
	교사들은 학생들의 시험점수를 통해 평가되어야 한다 (Arne Duncan, 미국 교육장관, 2009년 연설)	- 학생에 대한 데이터의 수집과 분석이 교육의 개선에 중요하다는 점을 설명 - 교사들은 어떤 학생이 제대로 배우고 있으며 어떤 교사가 잘 가르치는가에 대한 데이터를 활용하는 것에 대해 저항해서는 안됨 - 보다 많은 주에서 학생의 성취진보와 교사의 성공적 교수에 대해 확인하기 위해 데이터를 수집해야 한다고 주장
	현행 시험은 교사의 효과성을 재기에 좋은 척도가 아니다 (W. James Popham, 미국교육학회 AERA 전 회장)	- 시험성적은 어느 교사가 일을 잘 하는지를 판단하기 위한 공정한 자료가 아님 - 대부분의 주별 표준화시험은 느슨하게 규정된 내용표준에 기초하고 있으며 교사나 학생 모두 시험이 어느 범위에서 나올지 알지 못함 - 시험은 학생이 한 순간 알고 있는 것에 대해 보여줄 뿐이며 그들이 잘 교육을 받은 것인지를 알려주지는 못함
	전국단위 시험은 엄격한 Proficiency 표준에 도달하도록 할 것이다 (Michael K. Smith, 테네시 녹스빌 대학 교수, Testprep Experts in Knoxville의 대표)	- 국가수준 컴퓨터기반 수학시험을 통해 유치원부터 대학원학생수준까지 적용가능하다고 제안 - 이를 통해 시험과정을 보다 표준적 준거에 의해 활용할 수 있을 것이라고 봄 - 읽기, 과학 등에도 적용가능함

	표준화시험은 Proficiency를 정확하게 재지 못한다(Finn Jr, Pertilli, Thomas B. Fordham Institute 대표 및 부대표, 수월성교육 주창자)	– 표준화시험은 학생의 지식이나 기능적 능력에 대해 사람들이 일반적으로 생각하는 것보다 훨씬 적은 정보를 제공함 – 능숙하다는 것에 대한 기준이 주별로 서로 다르게 정의되기 때문에 주별로 시행되는 표준화시험은 학생의 학업성취에 대한 객관적 정보를 제공하지 못함 – 데이터를 수집하고 처리하는 것에 지나치게 의존하면서 데이터가 의미하는 것이 무엇인가라는 중요한 질문을 하지 않는 것이 문제임

<표 IV-3>'표준화시험은 대학입학을 위해 가치가 있는가'에 대한 찬반 양론

대학입학: 표준화시험은 대학입학을 위해 가치가 있는가?	성취시험은 대학입시의 좋은 도구이다 (Richard C. Atkinson, UC 전 총장 & Sau Geiser, Center for studies in higher education의 객원연구원)	– 대학입학에 가장 유용한 표준화시험은 고등학교에서 학생들이 배우는 내용과 긴밀히 연관된 것임. 교육과정과 연계된 성취도평가는 대학입시에서 가장 유용함. 고등학교성적이 가장 중요한 지표이며 표준화시험은 보완자료임. – 대학에서의 성공가능성을 예측하려는 시도는 실패했지만 표준화시험은 다른 목적을 실현할 수 있음 – 잘 만들어진 성취도평가문항은 고등학생으로 하여금 학교에서 열심히 공부하는 것이 중요하고, 대학들로 하여금 입학가능한 학생의 예비군을 확대하는데 도움이 된다고 봄
	표준화시험은 대학입시에서 지나치게 강조된다 (Elizabeth Landau, CNN 리포터)	– 대학들이 시험점수의 비중을 낮추기 위해 노력하는 이유가 있음 – 저소득층과 소수자들이 대학교육을 받을 수 있는 수준임에도 불구하고 시험성적이 낮게 나오는 경향이 있음 – 대학들은 다양성을 존중함으로써 학교의 이미지도 높일 수 있음
	SAT는 대학에서의 성공을 잘 예측해준다 (Peter Salins, 뉴욕주립대 스토니브룩대학의 교수)	– 학생의 SAT 성적이 고등학교학점보다 대학에서의 성공을 더 잘 예측해준다고 봄 – 대규모 주립대학 차원에서 SAT점수가 높은 학생을 더 많이 선발한 결과 대학졸업율이 높아짐 – 표준화점수가 학교성적보다 대학졸업여부를 예측하는데 훨씬 가치가 있다는 것을 드러냄
	SAT는 대학에서의 성공에 대한 취약한 예측도구이다 (FairTest, National center for fair and open testing,	– SAT1(추론시험)이 대학에서의 성공을 예측하는데 거의 영향을 미치지 못하는 시험이라고 봄 – 고등학교 학점(내신성적), 석차, 학생태도 등과 같은 정보들이 학생들의 수행에 대해 보다 명확한

	표준화시험에 반대하는 대표적 단체)	자료 제공 - 시험은 예측력이 낮은 지표이고 여성, 유색인종, 성인학생 등을 평가절하하기 때문에 대학들은 시험점수제출을 선택화해야 한다고 주장
	시험준비과정은 좋은 투자이다 (Scott Jaschik, 고등교육에 대한 온라인 정보사이트 운영자)	- National Association for College Admission Counseling(NACAC)에서 시험준비회사(사교육업체)들의 효과에 대해 낮게 평가하는 문제를 지적 - 시험준비회사들에 의해 성적이 오른다는 것은 대학들에 윤리적 문제를 제기하지만 점수가 오르는 것은 분명함
	시험준비과정은 과평가되고 있다 (Derek C. Briggs, 콜로라도보울더대학의 연구 및 평가방법론 프로그램 대표)	- SAT 및 ACT 시험준비과정의 가치에 대해 연구한 결과를 기초로 논의 - 최근연구들은 개인교습 등은 성적을 올리는 효과가 있다고 지적하고 있으나 사교육업체들이 대규모연구를 할 이유가 없으므로 학부모나 학생들은 사교육업체의 성적상승에 대한 보장에 대해 의심할 필요가 있다고 지적

3. 비판 입장

미국에서 표준화시험에 대한 반대운동은 1960-80년대에 걸쳐 비교적 활발히 전개되었다. 표준화시험에 대한 비판은 1920년대 월터 리프만(Walter Lippman)에 의해 가장 먼저 제시되고, 1962년 호프만(Hoffmann)에 의해 활성화되었다고 알려져 있다. 최근의 비판론자들에는 삭스(Peter Sacks), 콘(Alfie Kohn), 폽햄(W. James Popham), 코레츠(Daniel Koretz), 브레이시(Gerald Bracey) 등이 있다. 시민단체인 <FairTest and its assessment reform network> 등은 시험을 합리적으로 사용해야 한다고 주장하고, 표준화시험의 한계를 드러내기 위한 노력을 경주해왔다.

표준화시험에 대한 반대는 앞 절에서 살펴본 것처럼 그 논점에 따라 다양하게 나타난다.

첫째, 표준화시험이 교육을 저해한다는 문제가 있다. 표준화시험이 교육과정을 협소화함으로써 학생의 학습을 제한시키고, 표준의 설정이 취약계층에게 불리하게 작용하며, 시험준비교수는 교육을 약화시킨다는 주장 등이 그것이다.

둘째, 표준화시험을 통해 책무성을 묻는 고부담평가가 지니는 문제가 있다. 즉 현행 시험이 교사의 효과성을 재기에 적합한 척도가 아니며, 학생의 성취를 파악할 수 있는 시험이 되지 못할 뿐 아니라 긴장감이나 스트레스를 유발하여 교육에서의 흥미유발을 저해한다는 문제가 있다. 또한 저소득층지역의 경우 점수를 올려야 한다는 부담이 교육과정을 협소하게 하고 암기중심의 교육방식을 강화함으로써 교육의 질을 더 떨어뜨리고 있다는 지적도 있다.

셋째, 표준화시험이 가지고 있는 방법론적 결함이나 부수적 문제들이 있다는 점이다. 표준화

시험이 미래학업 예측에 효과가 없다는 주장, 점수를 올리기 위해 사교육을 유발한다는 주장, 이를 통해 사회경제적 불평등이 강화될 수 있다는 주장 등이 있다.

표준화시험에 대한 비판적 내용을 담고 있는 논의들은 많지만, 이 절에서는 해리스와 스미스(Harris & Smith, 2011)의 논의와 반시험활동가인 콘(Kohn, 2000)의 주장을 중점적으로 살펴보고자 한다.

□ 표준화시험에 대한 지배적 가정들의 문제

해리스와 스미스(Harris & Smith, 2011)에 의하면, 표준화시험에 대해 다양한 문제들이 지적되어 왔음에도 불구하고, 아직 대중에게 이 문제들은 생소하다. 따라서 이에 대한 비판이 충분히 이루어졌다고 보기 어렵다. 특히, 이들은 표준화시험에 대한 지배적 가정들이 문제시될 필요가 있다고 본다.

이들이 제시하는 문제들은 다음과 같다:

- 학생의 지식과 기능이 제한된 샘플의 문제들 안에서 확인될 수 있다고 보는가
- 평균 점수 이외의 정보는 알지 못하면서 특정 학교의 성취에 대해 얘기해본 적이 있는가
- 학생의 성취에 대한 정보를 객관적으로 확인해야 한다고 할 때 객관성은 곧 표준화시험이며 그 이상이 아니라는 것에 대한 논쟁을 들어본 일이 있는가
- 성취점수에 기초해서 교사나 학생에게 상벌을 준 적이 있는가. 이는 동기유발에 어떤 영향을 주는가
- 고부담시험에서의 점수 향상이 학습 개선의 지표라고 생각한 적이 있는가
- 시험이 교육과정과 교실 일상에 영향을 줄 것이라고 본 적이 있는가. 시험을 준비하기 위해 잃어버리고 있는 것이 있다고 생각해봤는가
- 학생의 성취에 대한 직접적 평가가 아닌 간접적 측정(표준화시험)에 더 비중을 둔 적이 있는가(직접적 평가는 교사들의 판단이나 학교에서의 직접 관찰에 근거한 것을 의미함)
- 성취점수가 높은 학교나 지역으로 이사하는 것이 자녀의 높은 성취에 성공에 도움이 될 것이라고 생각해보았는가. 표준화시험은 학교에서의 학업은 물론이고 생애전체를 통한 성공을 얼마나 잘 예견하는가?

이외에도 문제시해야 할 가정들로는 다음과 같은 것들이 지적된다.

- 시험시간은 정해져 있어야 하는가?
- 빠른 것은 반드시 우수한 것인가?
- 능숙도(Proficiency)란 무엇을 의미하며 그것은 누가 정한 것인가?
- 문제에 답한다는 것은 어떤 문제를 맥락적으로 발견하고 제시하고 해결하는 것과 동등한가?

– 표준화시험은 왜 특정한 기능들을 평가하는가? 누가 그것을 결정하며, 어떻게 결정되어지는가?

해리스와 스미스(Harris & Smith, 2011)에 의하면, 경험·계량심리학의 결과들은 표준화시험의 효과를 지지하지 않는다. 미국심리학회(APA), 미국교육학회(AERA), 미국교육측정학회(National Council on Measurement in Education) 등은 이미 시험을 위한 윤리적 표준을 만든 바 있다(1999 version of standards). 그럼에도 불구하고 적용의 과정에서 이 표준은 숱하게 위반되어 왔다는 것이다. 위반의 사례들을 보면, 다음과 같은 것들이 있다:

- 집단성취를 보도록 설계된 검사결과를 기반으로 개인성취를 판단하는 문제
- 시험점수를 근거로 고교졸업장 수여를 거부하는 것
- 우수학생에 대한 프로그램 배치를 시험점수로 결정하는 것

해리스와 스미스에 의하면, 표준을 오용하고 그 오용을 지지하는 자들은 대부분 교육자들이라기보다는 정책결정자, 기업대표들, 정치인들이다. 표준화시험이 교육적으로 적용되었을 때 나타날 수 있는 부작용에 대한 우려들이 있고 그것을 경계하기 위한 전문가집단의 개입이 있음에도 불구하고 교육외적 요인들에 의한 비교육적 오용이 계속되고 있다는 것이다.

□ 표준화시험은 교육에서 절대악이다.

알피 콘(Alfie Kohn)은 미국에서 표준화시험에 대한 반대 입장을 일관되게 강경한 어조로 비판해온 활동가이다. 표준화시험을 다루는 논의들이 균형적 관점을 제시하거나 근본적 문제와 활용상의 문제 등 다루는 초점에 따라 강조점에 차이가 있는 것과 다르게 콘은 표준화시험은 어떠한 장점도 없다고 말한다. 그의 주요 주장은 다음과 같다(Kohn, 2000).

1) 표준화시험은 불필요할 뿐 아니라 매우 위험하다

- 상관없는 것을 측정함
- 미국적 특수성: 역사적으로 어느 때보다 시험과 점수가 학생의 판정에 중요한 시대. 다른 어느 나라보다도 일찍, 모든 연령의 아이들에게 선다형시험을 실시하는 나라라는 점
- 왜곡된 정보에 의해 갑자기 거세어진 책무성에 대한 요구: <위기에 처한 국가(Nation at risk)> 발표 이후 학교에 대한 우려가 거세지고, 시험을 더 많이 실시할 것을 요구하게 됨
- 시험과 교육의 질: 가장 질 낮은 시험이 가장 많이 활용. 빠르고, 쉽게 실시가능, 실시하기에 비용도 낮음(어디선가 들여와서 실시하고 되돌려서 순식간에 결과 처리). 효율적인 시험은 덜 효율적 시험을 몰아내며, 중요한 능력들은 평가되지 않거나 가르쳐지지 못하게 함
- 시험의 배후에서 이익을 보는 집단: 시험을 통해 막대한 영리를 추구하는 회사들, 정치가들에게 유용함
- 숫자는 과학적이라는 문화적 경향(penchant): 내적 동기, 지적 탐구 등은 잘 안 보이지만 시험성적은 판매실적과 유사하게 계산, 추적되어 성공과 실패를 판가름하는데 사용될 수 있음

– 학습과정의 특성 및 교육의 본질 미고려: 교실에서 이루어지는 것들 모두가 양화될 수 있는 것은 아님. 어떻게 하느냐가 아니라 뭘하느냐, 효율적이냐보다 효과적이냐가 재기가 어려움. 따라서 측정가능한 결과들은 학습에 있어 가장 중요성이 낮은 것들일 뿐임

2) 시험의 객관성에 대한 가정은 오류이다.

– 평가가 판단이 아니라고 생각하지만 이는 오류이며 매우 주관적임
– 객관적이기 위해서는 규범, 표준에 대한 동의가 있어야 하는데 현실에서는 의견의 불일치만이 존재함
– 시험문제의 타당성은 문제시될 수 있으며, 결코 객관적인 것이 아님
– 시험불안은 학생이 자신이 알거나 할 줄 아는 것을 드러내는데 방해가 된다는 연구결과가 있음. 중요한 시험일수록 학생의 시험불안은 높아지는데, 시험성적의 타당성은 낮아진다는 점이 지적됨
– 사회는 정해진 시간 안에 기억을 통해 즉각적으로 정답을 찾는 능력을 요구하지 않음. 실제 생활을 정말 걱정한다면 표준화시험이라는 인위적 연습을 그만둬야 함
– 대학입학시험(SAT, ACT)들도 대학 1학년 성적에 대한 예측능력이 약하며, 특히 직업세계 성공에 대해서는 더욱 낮음. 사고력이나 적성을 재지 못하고, 언어섹션은 그저 어휘력 테스트일 뿐임. 이는 누가 대학을 가야하는지를 결정하는데 불필요함. 캐나다에서는 실시하지 않으며 수백개의 대학이 이를 요구하지 않음

3) 시험결과가 말해주는 것은 학생의 집이 얼마나 큰가에 대한 것 뿐이다.

– 성적은 학교교육의 질을 알려주지 않으며, 학생의 사회경제적 지위와 동원가능한 자원에 대해 알려줄 뿐임
– 성적을 결정하는 진짜 요인이 무엇인지를 알면 고부담결정에 그것을 이용하는 것을 변호하지 못할 것임
– 시험자체를 봐도 단기 기억에 의존한 시험준비 결과로 고차원능력을 잴 수 없음. 이 시험들은 학생이 정답을 선택했는지에만 관심이 있을 뿐임
– 정답이 없다는 것이 아니라, 정답을 맞췄다고 이해를 했고 정답이 아니라서 이해를 못한게 아니라는 것임
– 과정에 대해서는 관심이 없기 때문에 사소한 계산 실수를 추론능력의 결함이라고 해석하게 됨. 이는 모든 교과에서 유사하게 나타남
– 시험이 가소로울 정도로(말도 안 되게) 어려운 것도 문제이지만, 문제는 시험 자체가 그냥 가소롭다는 사실에 있음. 시험이 잘 교육된 사람을 가려내지 않는다는 것임
– 최상위 학생들의 경우도 누가 최상위학생이냐의 문제가 있음. 배우는 것을 좋아하고 깊이 생각하는 학생이 최상급학생이면 답은 다름. 물론 깊은 사고의 학생으로 시험을 잘 보는 학생도 있으나 시험은 깊은 이해보다는 학습에 대한 얄팍한 접근을 위주로 함

4) 최악의 시험들

- 전체적으로 선다형인 문제들
  - 답에 대해 설명할 수 없음
  - 다른 사고기능은 배제함(정보의 조직, 논의 구축)
  - 복잡한 문제해결인지과정을 알고자 하는 경우 문제가 큼

- 개방형 문제의 경우도 장치를 많이 걸어두기 때문에 에세이시험을 준비시키면 글쓰기 능력이 오히려 떨어짐
  - 가르친 사람이 직접 채점하지 않고, 외부 채점자들이 수박겉핥기식으로 한 두 구절을 가지고 평가함

- 문제유형이외의 문제
  - 시험시간이 정해져 있다는 것: 관건이 스피드이지 사려깊음(thoughtfulness)이나 철저함(thoroughness) 아니라는 의미임
  - 시험을 자주 실시하는 것: 특히 학년별 변화를 추적하는 경우 모든 학생이 같은 페이스로 학습하는 것을 가정하기 때문에 느린 학습자에게 비교육적임
  - 지나치게 어린 학생에게 시험을 보게 하는 것: 4학년 미만의 학생에게는 표준화시험을 보게 하지 말아야 함. 이들이 시험을 통해 자신의 이해정도에 대해 의사소통할 수 있다고 보면 안 됨. 결과는 그들의 인지적 능력이 아니라 동일한 장소에 일정 시간 앉아 있는 능력을 재는 것 이상의 의미가 없음
  - 상대평가시험의 문제: 중요하지도 않은 지식을 분류만을 위해 시험한다면 시험을 위한 교육은 해악이며 교육의 질을 개선하지 못함. 상대평가는 우수성을 평가하기 위한 것이 아니라 분류에 목적이 있음

- 최악의 시험은 상대평가로, 선다형에 의해서, 정해진 시간에, 반복적으로, 1학년부터 실시되는 시험임
  - 위의 요건을 차치하고라도 암기력 테스트이거나 시험보는 기술을 보는 것이문제임
  - 모든 표준화시험은 좋은 학습자의 특성을 무시하며, 좋은 인간에 대해서는 말할 것도 없음
  - 잴 수 없는 것들이 많다는 점을 간과함: 동기, 창의성, 상상력, 개념적 사고, 호기심, 노력, 아이러니, 판단, 의지, 뉘앙스, 선의, 윤리적 성찰 등
  - 학습에 대해 시험을 통해 가르쳐지는 것들: 사실에 대한 암기가 매우 중요하구나, 많은 것을 아는 것이 똑똑한거구나, 빠른 시간안에 하는 것이 중요하구나, 모든 것에는 정답이 있고 누군가가 답을 가지고 있구나 (따라서 독창적인 생각을 하지 않음. 해석적인 활동을 하기보다 정답을 찾거나 추측하려고 함)
  - 또 다른 두 가지 교훈: 개인에게 부과(협력하는 것 경계), 사전에 문제를 알지 못함(스트

레스 레벨이 높고, 암기해야 함)
- 이 모든 것들은 표준화시험에 내재된 문제들로, 어떻게 하면 학생들이 얼마나 잘 학습했고 교사들이 잘 가르치는지를 알 수 있는지를 고민해봐야 함

5) 고부담 시험의 문제

- 인센티브를 통해 동기화하는 방법은 중요한 목표 도달에 도움이 되지 못함
- 무능한 교사들에게 새로운 동기를 주기보다 그들이 해온 무능한 행태를 더 강화함
- 일부 학교의 성공스토리 등은 교육의 결과가 아니라 시험의 논리일 뿐: 성공적이라는 것이 무엇인가를 다시 따져 봐야 함. 단기적 시험의 결과로 효과성 거론해서는 안 됨
- 고부담평가는 학업성취도제고에 기여하지 못했음
- 학업성취도 제고 외의 문제들: 좋은 교사와 행정가들을 학교에서 떠나게 하거나 떠나지 않더라도 방어적이고 경쟁적으로 만듦; 부정행위(cheating)가 만연하며 교사와 학생이 적대화됨(낮은 성취학생을 부담거리로 간주); 과도한 분과화가 이루어짐(초4부터 분절적 교육); 교육에 대한 대화를 협소화시켜 시험을 잘 보는 것에 집중시킴(즉 성적만 좋으면 교실에서 무슨 일이 일어나든 상관없다는 것).
- 가장 심각하게는 교사들로 하여금 시험을 위한 교육을 하게 하는 것(teaching to the test); 시험에 잘 나올만한 내용 위주로 교육내용을 선택하게 함
- 위로부터의 통제적 압박은 교사들로 하여금 학생들의 학습에 대한 의욕을 높이고 성취를 위한 동기를 부여하기보다는 오히려 질낮은 교육을 하게 함
- 시험을 준비하는 시간에 학생들이 보다 비판적, 창의적, 탐구적(호기심 가진) 학습자가 되도록 해야 함
- 높은 점수가 높은 질의 학습을 의미하는 것이 아니며, 오히려 높은 점수는 깊은 사고와는 부적 상관이 있음
- 학교에서 배운 것이나 교육과정과 무관하게 시험보는 기술을 가르치게 됨

4. (재)옹호 입장

미국에서 표준화시험의 역사적 기원은 19세기로 올라간다. 그러나 표준화시험이 본격적으로 시행된 것은 1980년대 이후 교육에서의 책무성을 높이고자 하는 정책결정자들과 그 지지자들에 의해 이루어진 것이다. 교육에서의 책무성정책이나 고부담평가의 폐해에 대한 비판 여론이 강해지고, 교육현장이 받는 압박도 커지면서 표준화시험을 옹호하는 입장과 반대하는 입장이 팽팽하게 맞서 왔음은 앞의 사례에서도 확인할 수 있었다.

펠프스(Phelps, 2006)등의 논의는 마치 콘이 표준화시험을 절대악이라고 보는 것과 유사하게, 반대의 입장에서 표준화시험의 가치를 전적으로 옹호하는 태도를 취하고 있어 주목된다. 주로 전문 심리학자 및 통계학자인 이들은 책무성, 고부담평가를 포함하여 표준화시험이 교육적으로

지닌 가치와 그 활용에 대해 전적으로 긍정적인 관점을 취한다.

이들 논의의 핵심은 대중이 표준화시험에 대해 잘 모르고 있다는 것이고, 잘 개발된 시험이 지닌 교육적 의의는 매우 높고 교육적 가치가 옹호되어야 한다는 것이다. 이들은 표준화시험이 유익하고, 어떤 다른 대안보다 우수한 테크놀로지를 갖추고 있다고 본다.

펠프스(Richard Phelps)에 의하면, 1960년대부터 40여년 동안 여론조사 등에서 대중이나 교육자들로부터 표준화시험에 대한 반대는 없었다고 주장된다. 최근 들어 교사들의 비판은 증가했으나 교육소비자들의 지지도는 반대보다 3배 이상 높다고 언급한다. 표준화시험의 장점은 개선된 진단능력, 예측과 선발(대학, 장학금, 고용), 개선된 성취(가장 논쟁적) 등이라고 보며, 표준화시험이 있는 학교체제를 다닌 학생이 그렇지 않은 학생보다 더 많이 배우고 있다는 증거들이 보고된다고 주장한다.

이들에 의하면, 대규모 교육평가에 대해서는 오해가 있었다. 이러한 오해에 대해 국가수준평가(National Assessment of Educational Progress)의 지역별 데이터를 통해 반박하고 있다. (예. 너무 비싸고 시간이 많이 듬, 소수자에 대해 불리, 가르친 것에 대한 평가가 아님, 표준이 너무 높게 설정, 고등학교졸업자격을 위해 단일 시험에 근거 등)

이 저자들은 표준화시험에 대한 신화를 벗기고자 한다. 사람들이 표준화시험을 비판하면서도 정작 기본적인 것들에 대해 질문하고 있지 않다고 지적한다. 즉, 무엇이 표준화시험인가? 신뢰도가 무엇인가? 타당도가 무엇인가? 통과기준은 어떻게 설정되는가? 시험에 대한 추가적 정보를 어디서 구할 수 있는가? 등에 대해 원론적으로 짚어야 한다고 본다. 이들은 미국 50개 중 48개 주가 고부담평가를 실시하고 있고 주 표준을 가지고 있지만 시험의 기술적 질에 대해 제대로 평가가 이루어지지 않았다고 보면서 주단위 고부담평가를 실시하는데 있어서의 긍정사례, 부정사례들을 밝힌다.[11]

또한, 표준화시험에 대해 이루어져온 각종 비판들을 재반박한다. 예컨대, 크로커(Crocker) 등은 시험을 준비하는 교육(teaching to the test)이 교육과정의 필요를 충족한다는 주장을 하고 있고, 통과점수 설정의 임의성 문제를 반박하면서 표준설정의 절차가 잘 고안되어 있다는 것을 강조한다. 또한 (언어) 소수자, 장애인들에게 유익한 사례들을 제시하고, 책무성평가에 대해 비판하는 반대자들의 논거를 반박한다. 이들은 논란이 되어온 NCLB를 지지하며, 그 역사와 종합적 의의를 재확인하고 있다.

5. 대안의 모색

교육현실에 대한 해석의 차이에 따라 찬성과 반대가 팽팽히 맞서는 가운데 문제점들을 개선하고 평가의 폐해를 교육적으로 극복하기 위한 대안들도 제시되고 있다. 다음은 대안의 모색에

11) 또한 교육평가프로그램들이 지난 30년간 직면해온 법적 도전들에 대해 다룬다. 법적 쟁점은 두 가지 차원(법률적, 헌법적)에서 계량심리학적 요소들이 어떻게 변호되었는지를 밝히고 있다. 1) 법률적(statutory) 차원: title VI + title VII – Civiil Rights Act 1964, 2) 헌법적(constitutional) 차원: 14th Amendment(Debra P. vs. Turlington 재판과 GI Forum vs. Texas Education Agency 재판)

있어서 논의되고 있는 사안들이다.

첫째는 참평가방식의 실천이다.

표준화시험에 대한 반대론자들은 시험이 교육에 도움이 되어야 하며 교사에 의한 교육을 반영하는 것이어야 한다고 본다. 학교에 대한 책무성 확인을 위해서는 학교가 얼마나 좋은지, 학생들이 얼마나 잘 배우는지를 알 수 있어야 한다는 것이다(Kohn, 2000).

이를 위해서는 다음과 같은 사항들이 반영되어야 한다고 본다.

- 교사에 의한 서술형 통지표가 작성되어야 하며, 이는 관찰에 기반한 것이어야 한다.
- 수행평가, 포트폴리오 등의 방법을 통한 것이어야 한다.
- 학생이 무엇을 할 수 있고, 어떤 도움을 필요로 하는지 알 수 있어야 한다.
- 교사들이 평가한 것을 보고하는 방식과 관련한 체제가 구축되어야 한다. The Learning Record, Work Sampling (수행 기록, 변화 관찰) 등 참평가 과정을 반영하는 보고 형식이 활용되어야 한다.

둘째는 평가나 성취에 있어 비인지적인 측면과 다양성을 존중하는 방향으로 나아가야 한다는 것이다. 특히, 다중지능의 개념으로 접근이 이루어져야 하고, 다양성을 보장하기 위한 전략 및 평가 기준이 도입됨으로써 대학입시라는 관문에서 SAT라는 "거대 시험(big test, 능력과 잠재성의 심볼)"을 넘어설 수 있어야 한다고 본다(Wolf et al., 1991; Zessoules & Gardner, 1991; Gardner, 1992; Sedlacek, 2004).

셋째, (표준화)시험에 대해 조직적인 저항이 필요하다는 것이다(Kohn, 2000).

표준화시험의 폐해가 교육현장에서 큰 문제가 되고 있다면, 현실이 변화될 수 없다는 패배주의가 극복되어야 한다는 것이다. 나쁜 것은 알지만 할 수 없다는 인식이 극복되어야 한다. 현실의 학생과 학교가 고통받고 있는 점을 주목해야 하며, 이를 위한 교사의 개인적 실천이 중요하다고 본다. 또한 행동의 조직화가 이루어져야 제도적 부조리에 맞설 수 있다고 본다.

넷째, 평가가 실시되는 국가사회의 가정(혹은 표준)을 근본적으로 재검토하여 재조직하는 것이 필요하다는 근본적인 주장도 제기된다. 교육계 내부의 보다 급진적 관점들은 표준화시험을 책무성과 강하게 결합한 NCLB는 공화당, 민주당의 공동작품이라고 보고, 보다 근본적이고 진보적 논의들을 제기해야 한다고 본다(Spring, 2014).

예컨대 녹색당은 NCLB를 없애고 학교를 다시 지역의 공동체 통제하에 두라고 주장하는데, 학교에서 시민의식과 환경교육이 강조되어야 한다고 본다. 이를 위해 예술교육을 통해 억압을 폭로하고 경제사회문제의 해법을 찾아야 한다고 주장한다. 한편, 보수적 소수당인 자유당은 학교와 국가의 분리시켜 교육을 자유시장에 맡기고 정부의 개입을 없애야 한다고 주장하기도 한다.

교육사학자이자 활동가인 스프링(Spring, 2014)에 의하면, 학교교육의 목적을 전면적으로 바꾸고 공적 지원을 강화해야 한다고 주장한다. 스프링은 미국 헌법을 개정하여 공립학교에 평등

한 재정지원을 보장하고 교사의 학문적 자유를 보장해야 한다고 말한다. 또한 학교의 지역자치와 다문화적 문화언어를 보장해야 한다고 본다. 교육의 패러다임을 근본적으로 바꾸어, 인간자본의 강조가 아닌 노동을 위한 교육을 해야 하고, 모든 사람의 길고 만족스런 삶을 실현하기 위한 경제 교육과 환경교육을 해야 한다고 주장한다.

Ⅴ. 표준화시험 적용의 국가별 사례: 프랑스, 독일, 일본

1. 표준화시험 적용의 맥락

Ⅲ-Ⅳ장에서는 미국 사례를 통해 19세기이후 적용되어온 표준화시험의 형성과정과 이에 대한 찬반 논의를 살펴보았다. 미국은 측정에 대한 과학적 관심에 기반하여 대규모 표준화시험을 발달시켜왔고, 최근에는 표준화시험과 책무성을 강하게 결합시키는 정책을 실시하고 있다. 1990년대 이후에는 대부분의 국가들에서 국가수준의 평가를 통해 학력을 진단하고 교육체제를 모니터링하는 정책을 시행하고 있다. 이러한 국가수준의 평가는 각국의 평가체제와 조응하여 운용되고 있다.

이 장에서는 프랑스, 독일, 일본의 사례를 중심으로 유럽과 동아시아에서 표준화시험이 실시되어온 과정들을 살펴본다. 일반적으로 표준화시험은 대학입학과정과 국가적 학력평가에 적용되는 경향이 있다. 따라서 각 사례별로 대학입학에서 표준화시험이 적용되어온 과정을 살펴보고, 국가수준 학력평가의 운용 실태를 검토하고자 한다.

2. 각국별 사례

가. 프랑스

□ 대학입시의 기본 체제

프랑스의 고등교육체제는[12] 일반대학과 그랑제꼴(grandes écoles)로 구분되어 있고, 이들 학교간에는 입학을 위한 절차에 차이가 있다. 일반 대학의 경우 고등학교(리세) 졸업시험인 바깔로레아만으로 입학이 이루어지며, 그랑제꼴은 별도의 준비과정을 거쳐야 한다. 그랑제꼴은 전문 고등교육기관으로 바깔로레아 이후에 CPGE(prépas)라는 2년의 추가적 준비과정을 요구하며, 입학 경쟁도 높다. 대학 진학 이후에도 CPGE 과정에 동시 수강할 수 있다.

바깔로레아(baccalauréat)는 18세에 응시하며, 일반(자연과학S, 경제 및 사회ES, 문학L), 기술, 전문(직업)의 3종으로 이루어져 있다. 바깔로레아는 논리적 사고력, 정확한 표현력을 재는 논술형 시험으로 전 계열에서 철학과목이 필수과목인 점이 특징적이다. 이밖에 불어, 사회, 외국어, 수학, 물리, 자연과학, 체육 등 공통필수과목이 7개이다. 바깔로레아는 리세 최종학년 교사들이 평가문항을 개발하며 국가수준에서 채점 기준이 설정된다.

프랑스에서는 1968년 이후 사회변혁 요구에 따른 대학교육 대중화 개혁의 일환으로 소르본느 대학 등 각 대학의 고유 명칭을 없애고 '파리1대학,' '파리2대학' 등으로 '대학평준화'를 단

12) http://www.diplomatie.gouv.fr/en/france/studying-in-france/presentation-1988/the-french-education-system/article/the-higher-education-system

행했다. 이에 따라 대학 간의 지위 경쟁은 사라졌고 대학 진학은 거주지 위주로 이루어지게 되었다. 즉 파리 지역의 대학으로 진학하기 위해서는 파리지역의 고등학교를 졸업해야 한다. 따라서 프랑스대학들에서의 입학 기준은 바깔로레아만으로 한정되며 별도의 자격을 요구하지 않는다.

한편 그랑제꼴은 전문분야의 고등교육기관으로 프랑스에서 엘리트 교육기관으로 간주된다. 고등사범학교 등 명문 그랑제꼴에 들어가기 위한 입시 경쟁은 상당히 치열한 것으로 알려져 있다. 그랑제꼴을 엘리트교육기관으로 볼 수 있지만, 일반 대학과 그랑제꼴을 반드시 위계적으로 봐야 하는가에 대해서는 논란이 있다. 입학절차와 사회적 지위의 차이는 있지만, 대학은 일반적인 학문을 하는 곳이고 그랑제꼴은 실용적 전문교육기관이기 때문에 그 기능이 다르다고 볼 수 있다.

그랑제꼴 입학을 위한 CPGE(prépas)과정의 입학을 위해서는 고등학교 최종 2개 학년 성적이 요구된다. CPGE 과정은 고등학교 안에 소재하지만, 바깔로레아 통과 이후에 진학이 가능하므로 고등교육과정으로 간주된다. 이 과정은 대학 등록 이후에도 동시 이수가 가능하다. 그랑제꼴의 경우, 과학 및 경제 분야의 입학 경쟁은 낮은 편이지만 인문학은 매우 높은데, 해당 분야에는 고등사범밖에 없기 때문이다. CPGE 과정은 매주 시험과 작문, 구술발표시험을 치른다.

□ 바깔로레아의 역사와 현황[13)]

프랑스에서 대표적인 표준화시험은 바깔로레아이다. 바깔로레아는 선다형이 아닌 에세이문항을 위주로 구성되어 있고 교사들이 출제와 채점과정에 참여하기 때문에 선다형 객관식시험을 위주로 한 일반적 표준화시험과는 다른 유형의 시험으로 이해되기도 한다. 그러나 바깔로레아도 국가주도에 의해 표준화되고 대규모적으로 실시되는 평가이다.

바깔로레아는 1808년 3월 17일 나폴레옹치하에서 처음으로 실시된 시험이다. 이는 기본적으로 중등학교 졸업자격시험이지만 고등교육의 수학능력과 깊이 관련되어 있다. 이 시험이 처음 실시될 때의 목적은 중등교육에서의 성취를 확인하여 오직 시험결과만으로 고등교육에 배치하고자한 것이었다. 즉 프랑스 최초의 능력에 기반한(merit-based) 국가시험으로 시작되었다.[14)]

바깔로레아는 졸업자격과 대학입학의 두 가지 기능(dual function)을 동시에 수행하기 때문에 지속적으로 논란이 되어 왔다. 즉 입학시험으로 간주하여 소수만 통과하게 할 것인가, 졸업시험으로 간주하여 다수를 통과하게 할 것인가가 그것이다. 바깔로레아 취득비율은 1960년대 학생운동과정에서 86%까지 높아졌고, 1980년대에는 미테랑정부가 모든 시민이 고등학교를 졸업하고 바깔로레아를 취득하게 해야 한다는 정책을 추진했다. 2007년 보수파 사르코지의 집권으로 다시 바깔로레아를 강화하겠다는 입장을 표명하는 등, 바깔로레아는 사회적 평등성과 수월성 사이에서 중요한 지표로 기능해왔다.

13) Atia, Samir El (2008). From Napoleon to Sarkozy: two hundred years of the Baccalauréat exam. Language Assessment Quarterly, 5(2), 142-153에 기초함.

14) 프랑스학자인 브뤼네티에르는 프랑스의 교육제도와 시험에 의한 경쟁적 선발제도가 중국체제에서 영향을 받은 것이라고 본다. Brunetière, F., "L'Orient dans la littérature française" in Études critiques sur l'histoire de la littérature française, 8e serie, Paris, 1910: 197-201; Teng, 1943: 302 재인용.

현재 바깔로레아는 진로에 따라 일반, 기술(1968년부터), 전문(1985년부터) 프로그램으로 구분되어 있다. 프랑스에서 바깔로레아는 일종의 고교 졸업자격으로, 프랑스시민이 갖추어야 하는 필수적 요건으로 간주된다. 따라서 보다 많은 학생이 고교졸업자격을 갖추는 것, 즉 바깔로레아를 취득하는 것은 사회평등의 지표와 관련된다고 보는 것이다. 일반프로그램(General Program)은 4년제 대학 이상의 진학과 관련되며, 기술바깔로레아는 2년제 기술학교 진학, 전문바깔로레아는 취업과 연결되어 있다. 졸업자격이자 진로를 지시하는 두 가지 기능을 함께 갖고 있다.

바깔로레아라는 용어는 Bachelier에서 비롯되었는데 젊은 신사(young gentleman)라는 의미로 '기사가 되기 위해 훈련을 받는 사람'을 지칭하는 것에서 유래했다. 청년기에서 성인기로의 이행을 상징하는 것이기도 하며, 청년에서 성인으로 나아가는 통과의례로 간주되기도 한다.

프랑스에서 국가의 교육적 책임에 대한 논의는 18세기 대혁명 이전에 활발히 이루어지기 시작했고, 나폴레옹시기에 이르러 교육은 정부의 제도로 확고히 자리매김했다. 나폴레옹은 동일한 교수, 학습, 평가를 국가전체에 적용하고자 했다. 19세기초 교육장관 푸흐크호이(Fourcroy)는 초등, 중등, 고등단계의 학교체계를 세우고, 지역행정을 대표하는 아카데미도 세웠다. 교육행정의 집권화로 평가도 집중화되어, 1808년 3월 17일에는 제국 칙령으로 대학과 바깔로레아에 대한 규정이 만들어지게 된 것이다. 이렇게 출발초기부터 바깔로레아는 국가주도의 표준화시험이었고, 기존의 봉건적 귀속성에 영향을 받지 않는 능력주의를 표방했다.

1820년에는 바깔레로레아가 관직에 필수적인 시험이 되었고, 프랑스사회에서 주요 위치를 갖게 되었다. 이 자격 취득자도 1820년의 1,573명에서 1821년에는 4,503명으로 급속히 증가했다.

시험과목은 초기에는 그리스·로마 고전에 기반한 구술시험이었다. 500개의 주제에 대해 1년간 공부하여 시험을 보고, 시험당일은 심판관앞에서 질문에 대답하는 형식으로 치러졌다. 1830년 교육장관 빅토르 쿠젱(Cousin)이 불어시험에서 필기시험을 도입한 이후, 바깔로레아는 1905년에 이르러서는 순전히 필기만으로 된 시험으로 되었다.

현재는 구술시험이 일부 포함되고 있다. 1853년 최초의 외국어 시험이 도입(영어, 독어)된 후 100년을 경과하면서 다양한 과목이 추가되고 현대어 등이 강화되었다. 20세기초에는 프랑스교육과 바깔로레아의 대 전환이 이루어진 것으로 평가된다. 1902년 고전 바깔로레아를 대체하여 과학교과와 현대어가 주요과목이 되었고, 문학 및 고전어의 비중이 약화되었다.

이미 19세기 중반에 바깔로레아는 2개의 시험으로 분리되었는데, 언어와 과학(의학부 입학용)이 그것이다. 20세기 후반에 들어서는 1968년 기술 바깔로레아가 실시되기 시작했고, 1985년에는 전문 바깔로레아가 도입되었다. 이는 학생들을 그들의 강점에 따라 서로 다른 진로 프로그램에 배치하도록 하기 위한 것이었다.

이러한 진로 분화는 중학교말기에 시작하여 고등학교기간동안 계속된다. 이 배치에서 학생의 선호 및 선택이 일차적이고, 학과별 수행결과에 의해 종합 결정된다.

고등교육에 필요한 기능(일반)을 갖추지 못한 학생은 다른 바깔로레아(기술, 전문) 통해 1) 그들이 필수적 교육을 받았고 2) 적절한 직업전문세계에 나갈 수 있음을 의미하는 자격을 추구하게 된다.

대학진학이나 그랑제꼴 진학자는 일반 프로그램(General program)의 3개 영역 중 하나를 선

택한다(과학, 어문학, 경제 및 사회과학). 기술 바깔로레아는 2년제 기술학교 진학자를 위한 것이고 8개의 영역 중 선택한다. 전문 바깔로레아는 직업분야로 들어가게 하는 자격으로 교육부와 전문분야간의 협력에 의해 이루어진다. 전문 바깔로레아는 자격있는 노동자를 준비시키기 위한 것이며, 전문 바깔로레아를 선택하는 학생의 경우, 진학을 목적으로 하지 않기 때문에 학교에서의 교육은 시험 통과와 함께 종결되는 것이다. 2007년 교육부자료에 의하면, 고교졸업자의 절반 가량이 일반 바깔로레아(최소 4년제 고등교육)를 취득했고, 선택 분야는 어문학 19%, 과학 49%, 경제 사회과학 32% 등으로 나타났다. 나머지 절반은 졸업후 직업계로 가거나, 2년제 기술학교로 진학하고 있다(Atia, 2008 재인용).

바깔로레아의 시험개발 및 행정 절차는 상당히 관료적이고 복잡하며 장기적으로 이루어진다. 시험의 출제와 채점에 걸리는 시간이 총 12-15개월이다. 바깔로레아는 6월에 시험을 실시하게 되는데, 전년 5월에 교육부는 주제와 시험문제 개발을 의뢰하게 된다. 6월부터 분야별로 시험준비를 하게 되는데, 이는 지역별 아카데미에 맡겨진다. 아카데미에서는 위원회(시험분과)를 통해 제출된 시험문제 중 최종선택을 하는 선정위원회를 구성한다.

이 위원회의 구성원은 고교 교사(아카데미의 교육장이 임명)들로 지역장학관에 의해 추천된 교사들 중 아카데미의 장이 선발한다. 프랑스의 교사는 교육부 소속으로, 교수와 평가는 사범학교에서 교육부에 의해 지도된다. 선정위원회는 대학교수와 교육장학관이 공동위원장이 되며, 나머지 구성원은 모두 고교교사로서 시험문제를 출제하고 이 위원장들로부터 검토를 받는다. 9월과 12월에는 학과별로 두 개의 시험을 준비한다 (하나는 본시험이고, 다른 하나는 백업시험). 12월에서 1월 사이에 이에 대한 예비 검토가 이루어지고, 출제에 참여하지 않은 교사들이 이를 담당한다. 이들 검토교사들은 시험을 직접 본후, 분석하고 피드백 제공한다.

1월에서 2월 사이에 수정 보완이 이루어지고, 선정위원회의 두 공동위원장이 교육장이 제시한 가이드라인에 부합하는 지 판단한다. 이후 본시험과 백업시험의 최종본, 검토 위원들의 피드백보고서 및 선정위원회 자체 보고서를 교육부 시험국에 제출한다. 3월에 교육부가 당해년 시험을 선택하고 4월 및 5월에 최종 보완한다.

완성된 시험은 각 아카데미에 보관했다가 6월에 바깔로레아 실시하기 이전에 각 응시장으로 보내진다. 응시장에서는 고등학교와 중학교의 교사들이 시험을 집행하고 감독한다. 시험 후 교사들이 채점하며 결과를 이를 판정관에게 보낸다. 바깔로레아를 위해 총 4,000개의 문제가 만들어지고 이 중 40%는 백업으로 사용한다(Atia, 2008).

19세기에 출현한 바깔로레아는 능력기반시험이라고 하지만 엘리트들에 국한된 시험이었다. 1902년 교육개혁으로 초등교육이 세속, 무상, 의무화됨으로써 바깔로레아는 보다 많은 학생들에게 적용되어 왔다. 1940년대까지도 40% 수준의 합격률을 보였으나, 1960년대 개혁기 이후 많은 변화가 나타났다. 1968년 이후 최고 합격률 이 86%로 나타날 정도였다.

이 증가는 교육철학 및 시험 목표의 변화와 관련되는데, 보다 많은 사람이 필수 교육을 받도록 하려는 것이다. 일반 바깔로레아 외에 기술, 전문 바깔레로아로 내부 분화가 이루어진 것도 그러한 시민필수교육과 사회평등요구를 반영하는 것이라고 할 수 있다.

프랑스의 학생들은 고교 1학년때 진로계열을 선택해야 하고, 이에 따라 응시하는 과목들이 다르다. 인문학을 전공하려면 바깔로레아 L(문학)을, 사회과학은 바깔로레아 ES(경제-사회)를,

순수 자연 과학은 바깔로레아 S(과학), 산업 기술 분야는 바깔로레아 T(테크닉)를 통과해야 한다. 공통필수과목 외에 계열별로 지정 과목들을 추가적으로 응시해야 하며, 과목에 따른 응시 시기도 다르다.

필기시험의 경우 문제는 모두 주관식이다. 외국어(필기, 회화)나 수학(주관식 문제풀이)을 제외하고는 논술형 문제(논술, 텍스트논평)로 출제된다. 문제는 선택형으로 되어 있어 학생들이 불리하지 않게 하고 있으며 불합격할 경우 2차 시험을 볼 수 있다. 필기시험 이외에 구두시험과 실기시험도 있다. 계열별로 2-3과목에 대해 실시되는 구두시험의 경우 학생은 자신이 읽은 책들의 목록을 제출하고 그 안에서 문제가 출제되도록 하고 있다. 학생은 문제가 잘못 설정될 경우 이의제기를 할 수 있다.

주관적 시험이지만 공정성이나 객관성에 대한 시비는 없고 사회적으로 시험관들의 판단이 존중된다. 평가위원 간에 채점상의 신뢰도를 높이기 위한 절차들이 있는데, 학생의 내신성적이 주요한 판단 기준이 된다. 바깔로레아의 출제와 채점은 고교교사들에 의해 이루어지고 대학교수들은 관리 이외에는 관여하지 않는다. 고등학교 교사들이 직접 출제하고 채점하므로 고교 교육과정에서 다룬 내용의 범위를 넘어서지 않는다.

바깔로레아시험은 지방교육청이 주관하여 실시하며, 몇몇 지역이 공동으로 출제위원회를 구성하기도 한다. 각 과목은 20점 만점으로 매우 우수(16점이상), 우수(14-16점), 약간 우수(12-14점), 합격(낙제) 등으로 구분되며 10점 미만은 낙제이다. 바깔레로아 합격자는 1970년대에 20% 수준이었으나 1990년대 이후에는 50% 이상으로 늘어났다. 바깔로레아의 공통교과목에 철학과목이 있는 것은 특징적이며 시험문제들도 난해하고 높은 수준의 사고력과 표현력을 요구하는 것으로 잘 알려져 있다(이윤미 외, 2012: 53-56).[15]

□ 국가수준 표준화시험과 책무성[16]

한편, 바깔로레아 이외에도 프랑스에서는 1979년 이래 국가수준의 표준화시험이 실시되고 있다. 프랑스는 유럽연합에서도 국가수준시험의 종류가 많고 실시빈도도 높은 국가로 간주된다.

프랑스에서 실시되는 표준화시험을 목적에 따라 크게 세 가지로 분류하고 상세화해보면 아래 표와 같다.

<표 V-1> 프랑스 초등 및 전기중등 최종 시험(모니터링 주기 평가)

시험 종류	초등 및 전기중등 최종 시험(모니터링 주기 평가)
주요 목적	학교 및 교육시스템 모니터링
실시 학교급	초등학교, 전기중등(중학교)
목표와 활용	목표: - 국가수준에서 교육체제에 대한 모니터링

15) http://www.diplomatie.gouv.fr/en/france/studying-in-france/presentation-1988/the-french-education-system/article/the-higher-education-system

16) EURYDICE (2009)를 기초로 정리함.

	– 핵심교과에서 학생의 역량과 지식에 대한 객관적 보고 활용: – 국가수준에서 교육정책을 조정. 구체적으로는 교육과정 내용, 역량 진단, 교과 조직, 교사 조직, 특정학습자군 등과 관련
주요 대상	– 대표 표집 (공립, 사립) – 초등 졸업학년(10–11세) – 의무교육 최종학년(14–15세)
시험 과목	– 5년 주기로 초등과 전기중등 교과(예술, 체육 제외)를 순차적으로 실시 – 1차년도 불어, 2차년도 외국어(영어, 독일어, 스페인어), 3차년도 사회과, 4차년도 자연과학, 5차년도 수학

<표 Ⅴ–2> 프랑스 초등 및 전기중등 최종시험 (기초 역량)

시험 종류	초등 및 전기중등 최종시험 (불어와 수학에 대한 기초 역량)
주요 목적	학교 및 교육시스템 모니터링
실시 학교급	초등학교, 전기중등(중학교)
목표와 활용	목표: – 국가수준에서 교육체제에 대한 모니터링 – 불어와 수학의 기초역량에 대한 객관적 보고 활용: – 국가수준에서 교육정책을 조정. 구체적으로는 교육과정 내용, 역량 진단, 교과 조직, 교사 조직, 특정학습자군 등과 관련
주요 대상	– 학년 중간에 실시 – 대표 표집 (공립, 사립) – 초등 졸업학년(10–11세) – 의무교육 최종학년(14–15세)
시험 과목	불어, 수학

<표 Ⅴ–3> 프랑스 초등 및 중등 진단평가

시험 종류	진단평가
주요 목적	학생 개인의 학습 필요 확인
실시 학교급	초등 및 전기중등
목표와 활용	목표: – 학생과 학급의 성취수준 파악 (강점, 약점) 활용: – 교사는 학생의 학습과정을 돕기 위해 필요한 조치를 취해야 함 – 학급내 이질성과 학습속도의 다양성을 고려해야 함
주요 대상	– 초등학교(CE2) 학생에게 선택적 – 전기중등 첫 학년 학생에게 필수
시험 과목	불어, 수학

국가수준 시험의 실시와 관련하여 프랑스체제의 특징은 다음과 같다. 우선, 역사적 과정을 보면, 1979년 교육의 현대화와 민주화 진행 중에 처음으로 국가시험을 실시하게 되었다. 이를 통해 교육과정 목표에 대한 학생의 수행정도와 학교차이를 비교하기 위한 조치가 취해졌다. 1989년이후에는 중학교 진급시 진단평가가 의무시행 되었다(1989년 7월 교육법). 2007년 이후 초등학교에서의 진단평가는 선택으로 바뀌고, 중학교에서의 시험은 그대로 유지되었다. 여기에 5년단위 주기평가가 추가 되었다.

평가의 실시는 교육부전문기구가 외부전문가와 함께 한다. 평가시 시험문제 외에 다양한 배경 정보에 대한 설문조사를 통해 맥락적인 확인을 하고자 한다. 진단평가를 의무화한 이후 장애학생을 고려한 시설 등을 갖추고 있다. 채점의 경우, 모니터링시험은 평가기구(DEPP)가 하고 형성평가는 교사가 직접 한다. 모니터링평가를 학교의 책무성과 연결시키지 않으며 국가수준 정보에 국한한다.

나. 독일[17]

□ 대학입시의 기본 체제[18]

연방국가인 독일에서 교육의 권한은 각 주가 지니고 있기 때문에 주별로 교육과 관련한 각종 사항들에 다소간의 차이가 있다. 독일은 복선제체제의 영향이 강하게 남아 있어 복잡한 중등교육체제를 지니고 있으며 대학진학방식도 이와 관련되어 있다.

중등교육은 1단계와 2단계로 구분된다. 4년의 초등교육을 마친 후 대개 5-6학년이 진로탐색 기간으로 설정된다. 중등 1단계는 하우프트슐레(직업실무교육, 최종 9학년), 레알슐레(일반 및 직업 통합, 최종 10학년), 김나지움 하중급반(10학년이후 2단계 진급 혹은 직업교육)로 구성되며, 직업교육과 김나지움교육이 통합된 종합학교(최종 10학년)도 있다.

중등 2단계는 김나지움 상급반(11학년에서 13학년까지) 및 중등1단계 이후의 직업교육을 포함한다. 김나지움 상급반은 대학입학자격인 아비투어(abitur)와 연결되어 있다. 상급반의 교과목은 5분야인데 그 중 3분야(언어문학예술, 사회과학, 수학자연과학공학)는 필수이다.

아비투어는 학생이 선택한 2과목의 기초코스와 2과목의 심화코스로 치러진다. 시험과목은 세 가지 필수분야를 충족시켜야 한다. 상급반에서 받은 성적은 아비투어의 점수에 합산된다.

대학입학자격인 아비투어는 학교가 주관하여 치르는 시험이지만 주에 따라서는 주정부가 주관하여 동일하게 치르기도 한다. 대학진학시에는 학생이 이수한 기초코스, 심화코스, 아비투어 점수를 합산하여 성취등급을 매기고 이를 기초로 입학허가를 받게 된다.[19]

1970년대 이후에는 지원자가 많은 일부 전공(의학, 치의학, 수의학 등)을 중심으로 정원제를 실시하여 중앙 대학입학분배기관(ZVS)에서 배정하고 있다. 정원제에 해당하지 않는 대학에 대

17) 독일 교육사에 등장하는 학교명과 관청명은 의미를 살리기 위해 번역하지 않고 원어 그대로 사용함.

18) 정영근 (2010). 독일학교의 내신평가, 일제고사 그리고 대학 진학, 우리교육 평가제도를 평가한다, 제5회 원탁토론 학술심포지엄(서울·경기·강원교육청 위탁 전문과정 연수자료), 2010년 10월 23일, 91-100쪽

19) www.abi-nachholen.de/abituraufgaben.html 에서 베를린(부란덴부르크)과 바이에른의 사례 확인 가능함. 지역별로 일반의 접근을 차단하고 있는 경우 있음. 2014년부터 CAS(Computeralgebrasystem)를 이용한 시험이 실시되어 별도로 구분되어있음.

해서는 스스로 대학을 정할 수 있다. 학생선발기준은 아비투어 등급에 따라 정해지며, 입학자리를 기다린 대기기간 가산점을 적용하여 선발한다(정영근, 2010).

□ 아비투어의 역사와 현황[20]

독일의 아비투어의 경우도 프랑스의 바깔로레아와 유사하게 졸업과 입학의 두 가지 기능(dual function)을 수행하는 시험이다. 차이가 있다면 바깔로레아가 전문프로그램의 도입으로 졸업자격 및 진로결정의 의미를 지닌 확대된 방식으로 활용된다면, 아비투어는 대학입학자격과 보다 직접적으로 연결되어 있다는 것이다.

독일에서 아비투어는 1788년에 처음으로 실시된 시험으로, 이후 중등학교-대학 간 이행과정에서 주요 역할을 해왔다. 히퍼드(Heafford, 1995: 285-286)에 의하면, 아비투어의 주요 특징은 7가지로 요약될 수 있다.

1) 고교와 대학에서의 학습수준 설정
2) 국가에 의해 주관되며 국가수준의 통일된 표준 설정
3) 졸업자격시험이면서 대학입학시험이라는 이중 기능 수행
4) 학교단위에서 학교교사들에 의해 집행됨
5) 시험의 통과는 대학입학지원자격이 아닌 입학을 보장함
6) 특권화되어 있어 다른 하위단계 시험과 호환이 안 됨
7) 특정 교과목들에 대해 적정 수준을 요구하고 자유롭게 선택할 수는 없음

이러한 특징들은 19세기 프러시아에서 확립되어 왔고 독일의 학교와 대학의 특성을 형성해왔다. 아비투어의 출현과정에서는 두가지 요인이 핵심적이다. 하나는 프러시아의 학교체제와 대학에서의 개혁 필요와 다른 하나는 프러시아국가의 존재이다. 학교체제의 경우, 교육내부의 연계성이 부족하고, 독일대학의 쇠퇴(질낮은 강의)와 관련된다. 대학의 특권화로 학생에 대한 군입대가 면제되자 학습보다는 군대면제에 관심있는 사람이 증가했다. 또한 라틴학교들에서 많은 학생을 수용하게 되자 이는 상층계급의 특권에 대한 위협의식을 높이고, 상공업분야가 약화될 것이라는 우려를 높였다.

한편 대학에서는 질낮은 학생의 대학입학 증가, 질 낮은 강의, 빈곤층 학생의 학비지불능력 취약 등이 문제가 되고 있었고 음주, 도박, 무력충돌 폭동 등의 무질서 양상이 나타나고 있었다. 18세기초부터 이에 대응하기 위해 학생들을 단속하기 위한 일련의 조치들이 발표되었다(1712년, 1750년, 1798년 등).

18세기말에 이르러 이러한 문제들과 함께 유능한 관료 충원의 필요로 인해 교육개혁을 단행하게 된다. 1787년 교육부의 전신이 된 기구(Oberschulkollegium)의 설치는 중등문법학교(lateinschulen; gelehrten schulen)의 교육을 개선하기 위한 것이었다. 아울러 교사교육의 표준

20) 아비투어의 역사는 Heafford, Michael (1995). The early history of the abitur as an administrative device, German History, 13(3), 285-304의 내용에 기초함.

을 강화하고, 교사임금을 인상하는 등의 조치도 취해졌다. 학교들을 분류하여 우수학교는 gelehrten schulen로, 일반학교는 burgerschulen로 구분되었다. 이 기구가 설치된 후 수행한 가장 중요한 역할은 1788년에 아비투어 시험(abiturientenexamen)을 시작한 것이다.

앞서 보았듯이 18세기초부터 국가는 교육에 관심을 보여왔다. 그러나 18세기말 Oberschulkollegium을 설치하기 전에는 즉흥적으로 문제를 해결해왔다. 국가는 교육의 최종단계교육을 규제하는 것이 가장 효과적 통제라는 것을 알게 되었다. 전통적으로 대학입시는 해당 대학 학장에게 지원하는 것으로 별도의 형식적 절차가 없었다. 프러시아국가는 대학들로 하여금 입시를 강화하도록 하는 대신 중등졸업자격을 요구하고 그것을 대학이 확인하도록 조치하는 형식을 취하게 되었다. 학생이 졸업자격없이 오는 경우에만 대학의 학장이 별도의 시험을 실시했다. 이러한 조치는 18세기초부터 취하도록 되어 있었지만 규정들은 종종 무시되었다. 1787년 이전에도 이미 규정은 존재했으나 제대로 시행되지 않았던 것이다. 이러한 규정을 실질화하기 위해 국가가 Oberschulkollegium을 설치하여 대학입학을 관리하고자 했다. 졸업자격증에는 학생의 지식과 태도에 대해 명시하도록 했고, 필기시험과 구술로 이루어진 시험을 실시하도록 했다(Heafford, 1995: 291).

아비투어의 실시로 형식화된 시험이 시작되었으나 학생들의 대학수학능력이 쉽게 향상되었던 것은 아니며 학교들에서도 시험지도가 항상 엄격했던 것은 아니다. 시험문제를 미리 주고 준비시킨다든가 횟수, 시간, 내용 등 시험에 대한 사항들이 불명료하게 남아 있는 문제들이 있었다. 시험결과에 있어서도 도달(mature)과 미도달(immature) 사이의 편차는 학교들이 설정하는 표준의 차이에서 비롯되기도 했지만 일부 학교들이 순수한 학문적인 기준외의 기준을 적용하기 때문에 나타나기도 했다. 아비투어 시험은 일반적으로 대학시험보다는 어려웠다. 따라서 학교를 중퇴하고 대학에서 직접 시험을 치는 사례도 있었다. 대부분의 대학들은 거의 입학을 거부하지 않았다.

교육부의 맹아에 해당하는 관청을 통해 아비투어를 시작하게 되고, 이로써 중등학교에 대한 관리를 합리화할 뿐 아니라 직접 규제는 하지 않으면서도 대학입학을 통제할 수 있게 되었다. 또한 아비투어수준으로 가르치는 학교와 다른 중등학교들 간의 구분이 이루어지게 되었다. 대학들은 입학관리에 직접 나서지 않아도 되고 대학교육과정에 전념할 수 있었다. 아비투어 시험이 규범이 되면서 대학별로 실시하는 시험은 예외적인 것이 되었다.

나폴레옹전쟁으로 1806년 예나에서 프러시아군대가 패배하였다. 이는 프러시아에 큰 타격을 주게 되었고 18세기까지의 교육제도에도 변화를 초래했다. 그러나 엘리트 중등학교(gelehrten schulen)와 대학들은 오히려 더 강화되었다. 18세기말 프랑스어가 외교언어이자 문학언어로 독일에서 통용되어가면서, 독일의 문화적 독자성에 대한 요구가 높아지게 되었다. 나폴레옹 침공 이후 어문적 독립성에서 정치적 독립에 대한 요구로 나아가게 되었고, 새로운 건설에 대한 열망이 높아졌다. 즉 물리적으로 잃은 것을 지적, 정신적으로 되찾고자 했으며 이를 위해 교육이 중시되었다.

피히테는 1807년에서 1808년 사이에 독일민족의 생존을 위해서는 전통적 교육제도의 전면적 변화가 필요하다고 주장했다. 아이레르(Eilers)도 지적능력(intelligence)이 권력이며, 지적 역량이 높은 국가가 더 강대해질 수 있다고 주장했다. 교육의 책임은 국가에 있다고 주장되었다.

Oberschulkollegium을 대체한 기구(Sektion fur Kultus und den offentlichen unterricht)가 신설되었고, 훔볼트가 대표로 임명됨으로써 활기를 찾았다. 교육은 보다 넓은 목적을 충족하도록 기대 되었고, 인간성을 위한 교양(Bildung zur humanitat)이 시대의 철학 및 교육의 기본 원리가 되었다. 1808년에서 1810년까지 훔볼트가 재임한 짧은 기간동안 그의 철학이 교육의 기본체제에 반영되었고, 이는 19세기에서 20세기까지 독일교육의 근간이 되었다.

훔볼트는 초중등 및 고등교육이 하나의 계속적 전체를 이룬다고 보았다. 모든 사람에게 직업준비가 아닌 일반교육을 제공해야 한다고 주장했다. 중등교육은 기억훈련, 이해력, 판단력, 도덕적 감성을 키워야 한다고 보았고, 직업계열이나 대학으로 가는 학생에 대해 교육이 차별적으로 이루어지는 것을 문제시했다. 김나지움교육도 균형적 교육과정을 취해야 한다고 보았다. 훔볼트가 19세기에 실현하고자 한 것은 18세기적 인문주의 즉 고전을 통한 인간화였다.

훔볼트는 새롭게 실행된 교사임용시험에서 교사의 지위가 높아질 뿐 아니라, 기존처럼 신학부가 아닌 철학부에서 교사훈련이 이루어지기를 의도했다. 이러한 인문주의적 관점은 1812년 아비투어시험 규정에 반영되어 일반교육의 표준을 충분히 실현하도록 요구했고 유기적 통합체(organic unity)를 형성하도록 했다. 교과목은 언어, 역사, 수학적 지식을 중시했다.

1800년에서 1850년 사이의 기간은 중앙교육행정과 지역행정(김나지움과 중앙연결)이 발달한 시기로 김나지움은 아비투어를 준비하는 학교의 명칭으로 자리잡아갔다. 중앙행정부처는 다시 1817년에 명칭 개칭을 했다(Ministerium der giestlichen und unterrichtsangelegenheiten).

1830년대 이후에는 통제가 보다 엄격해졌다. 1834년 이후 김나지움 졸업자격시험은 개정되어 보다 엄격하게 되었고 통일성이 강화되었다. 1837년에는 모든 김나지움에 공통 실라버스를 요구했다. 9년으로 김나지움 교육과정을 규정하고 매해 진급을 평가하도록 했다. 이러한 과정을 거쳐 아비투어는 가장 특권적인 시험으로 자리잡아갔다. 이론적으로는, 계급적 한계 속에서도 능력있는 학생들에게 기회가 개방되어 있는 시험으로 위상지워졌다.

국가에 의한 외부통제는 시험의 권위를 높였다. 교사에 의한 내부 의사결정은 학생에게 압박으로 작용하고 경쟁을 유발했다. 이는 부모와 학생의 저항을 초래하기도 하여 지나친 경쟁과 단조로운 시험준비 등을 하지 않도록 조치들이 강구되기도 했다. 1856년까지 대략의 형식이 형성되었다. 전체적으로 볼 때 아비투어는 프러시아정부의 관심에 의해 1780-1915년 기간동안 상당한 합의하에 형성되었다. 시험이 사회적으로 특권화함으로써 보다 많은 학생이 학교에 남도록 하는 동인이 되기도 했다. 강한 통제는 훔볼트가 원하던 폭넓고 균형적 교육으로부터 멀어졌고, (선택된 소수를 대상으로 한) 협소하고 기계적인 시험준비교육을 실시하게 했다.

아비투어시험의 결과 판정에서 중등학교 교사의 의사결정권이 강하여 대학의 이익보다는 중등학교의 관심이 내용에 반영되는 경향이 있었다. 전면적 교양(Allgemeine Bildung)의 요소들을 유지함으로써 학생들이 다양한 진로를 선택하게 하는데 유용하게 작용했다. 또한 대학의 통제로부터 자유로운 시험이기 때문에 상대평가보다는 절대평가를 할 수 있게 했다.

아비투어는 능력주의를 통해 사회계급적 제한을 넘어설 수 있게 한다고 주장되기도 한다. 그러나 아비투어는 관료기제와 긴밀히 연관되어 있어 민주적 제도로만 조망되어서는 곤란하다. 또한 김나지움의 입학자체가 워낙 선별적이기 때문에 아비투어를 민주적이라고 보아서는 곤란한 것이다.

독일에서 아비투어 자격자가 지니는 전공 및 대학선택의 자유는 다른 국가들에서 쉽게 상상하기 어려운 제도이다. 독일의 대학은 입학자나 전공자격을 통제할 수 없다. 1960년대 베이비붐세대로 인한 아비투어 자격자의 증가는 대학이 관리 불가능한 수준으로 입학자가 늘어나게 했다. 이러한 증가는 튜터(tutor)제 운영을 어렵게 하고, 대학 재학기간을 증가시켜 대학인구의 폭발로 나타났다고 평가되기도 한다.

아비투어는 현재 4월말에서 5월 중순 사이에 실시되며 6월말에는 점수가 공개된다. 시험시간의 경우 기초 코스인 필기시험은 240분, 심화코스로 선택한 두 과목의 시험은 각기 300분이다. 구두시험의 경우 20분간의 준비시간을 둔 뒤에 20-30분간 실시된다. 시험의 응시자격은 김나지움의 필수 기초 및 심화코스를 모두 마친 자이며, 성적의 합이 수준에 도달했을 때이다. 4과목의 시험 중 3과목은 필기로 봐야 하며 그 중 2과목은 반드시 심화코스이고 나머지 한 과목은 기초코스로 본다. 네 번째 과목은 구술시험으로 본다. 4과목은 3개 필수분야를 모두 포함해야 하며, 한 과목은 반드시 언어(독일어 혹은 외국어) 혹은 수학이어야 한다. 김나지움 상급반의 성적은 아비투어점수를 합산하여 1.0-4.0까지의 성취등급(소수점 한 자리수까지 표시)으로 산출되며 이를 근거로 입학허가를 받게 된다(정영근, 2010).

□ 국가수준 표준화시험과 책무성[21)]

<표 V-4> 독일 전기중등 졸업자격 시험

시험 종류	하우프슐레 졸업자격 9학년/ 레알슐레 졸업자격 10학년
주요 목적	학생의 진로에 대한 결정
실시 학교급	전기 중등
목표와 활용	목표: - 질관리 및 학교와 교수활동의 개선 - 내부 및 외부 평가를 지원 활용: - 학생이 학교교육의 목적에 도달했음을 확인 - 자격증으로 활용 - 학점부여 및 상급학교의 진급 지원 - 지역학교행정청 및 Länd 교육부에 대해, 평가결과는 종합되어 지표와 함께 제시되며, 시간적 추이를 제공함
주요 대상	- 학교를 졸업하고자 하는 9학년 학생 모두에게 의무적 - 10학년까지 학교를 다니고자 하는 학생에게는 10학년에서 의무적
시험 과목	- 의무적 교과는 독일어, 수학, 제1외국어(대개 영어이며, 불어를 선택하기도 함) - Baden-Württemberg의 경우, 주어진 주제의 프로젝트 시험을 추가적으로 응시해야 함

21) EURYDICE (2009)를 기초로 정리함.

독일에서는 다른 나라에 비해 다소 늦은 2004년에 표준을 도입하고, 모든 Länder의 중등교육 단계에서 국가시험을 실시하고 있다(2004/5년부터). 이 시험은 전기중등 졸업자격시험으로 학생의 진로결정을 위한 것이다. 초등학교단계에서는 시험이 실시되고 있지 않다. 이 시험은 교육부가 아닌 평가전문기구가 주관하여 실시하고 있다. 독일의 평가전문기구인 Institute for educational progress(IQB)의 경우, 프랑스와 영국으로부터 도움을 받아 설치되었다(Huber & Gördel, 2006).

비교적 최근인 2009년에 진로결정 이외의 목적으로 실시하는 시험이 처음으로 실시되었다. 2009년 이후 표준에 기반한 주별 비교연구를 매해 실시하고 있는데, 3학년과 8학년 전체를 대상으로 하며 학력에 대한 정보 제공을 목적으로 한 것이다.

독일의 경우 연방체제이기 때문에 전문기구가 Länder등과 협의를 하여 시험을 운영한다. 시험에는 2004년에 도입된 Länder 교육문화부의 표준과 Länd의 교육과정을 함께 반영한다. 시험의 집행은 학급교사가 하며, 채점은 학급교사와 외부인이 함께 한다. 담당교사가 1차 채점을 하고 동일교과의 타교사가 2차채점을 한다. 2차 채점자는 학교장이 추천한 같은 학교의 교사일 수도 있고, 타학교의 교사가 될 수도 있다.

독일에서의 표준기반측정 및 책무성평가의 도입은 기존의 교육문화에 비추어 상당히 획기적인 것이라고 논의된다. 독일은 1990년대에 유럽에서 전개된 대단위평가운동의 흐름에 냉담했는데, 그 이유는 '교육에서 중요한 가치를 지니는 것은 측정될 수 없다'는 문화적 관점이 강했기 때문이다. 그러나 2000년대 이후 소위 'PISA 쇼크'라고 불리는 학력논란을 겪으면서 최근에는 표준에 기반한 측정에 대한 문화적 반감이 약화되었다는 평가를 받고 있다(Lundahl & Waldow, 2009: 376-377).

다. 일본[22]

▫ 대학입시의 기본 체제

일본의 대학입학제도는 국가 수준에서 출제하고 관리하는 대학입시센터시험과 대학별로 실시하는 대학별 고사로 이루어져 있다.

우리나라와 유사하게 일본도 대학입학경쟁이 심한 것으로 알려져 있는데, 이는 근대 초기 일본교육체제 성립과정과 관련되어 있다. 일본에서는 일찍부터 대학별 시험이 발달했는데 그 이유는 1877년 메이지 10년 최초의 대학인 동경대학이 설립되었지만 중등학교체제와 연관성이 낮았기 때문이다. 중등교육의 졸업여부만으로 전문교육의 기초학력 보유 여부를 판별하기 어려웠기 때문에 별도의 시험을 실시할 수밖에 없었다. 이후 일본에서는 중등학교와 대학 사이에 예비학교수준의 교육기관설립의 필요성을 느끼게 되었고, 이러한 필요에 의해 출현하게 된 것이 고등학교이다. 메이지 19년에서 20년까지 고등학교들이 만들어졌고 졸업생들은 자동적으로 동경제국대학에 입학을 하게 되었다. 이 결과 고교입시경쟁이 대학입학경쟁보다 더 치열했다.

22) 이양락 외 (2009). 일본대학입시센터시험 문항분석, 한국교육과정평가원, 연구자료 ORM 2009-46; 한국교육과정평가원 (2001). 일본, 중국, 대만 대학입시제도연구, 한국교육과정평가원, 연구보고 RRE 2001-12

그러나 1920년대 고등학교수가 급격히 증가하면서 무시험대학입학제도는 없어지게 되었다.

현재의 대학입시제도는 대학입시센터시험과 대학별 전형으로 구성된다. 대학입학센터시험은 실제 시험이 이루어지는 해의 2년 전부터 문제의 설계, 작성, 검토가 이루어진다. 대학 교수로 구성된 위원회와 고등학교 교사들로 구성된 위원회의 위원들이 출제를 기획하며 문제에 대한 출제, 검토, 수정을 한다. 센터 시험의 교과는 언어, 수학, 지리역사, 외국어, 이과, 공민으로 나뉘며 국어와 수학은 과목수와 이수단위수가 다른 여러 종류의 과목으로 구성된다. 각 대학은 임의로 과목을 선택하고 이용할 수 있고 과목별 배점비율까지도 대학이 정할 수 있도록 하고 있다. 시험은 2일에 나누어 본다. 문항은 다지선택형으로 논리적 사고력과 판단력을 평가할 수 있는 문제를 내고 있으며 평균점의 범위를 60점으로 하고 있다. 학생의 개인 점수는 지원 대학에 바로 보내지며 입시가 끝나기 전에 개별 통지되지 않는다.

대학별 시험에서는 학부나 학과에서 필요로 하는 학생을 선발하기 위해 필요한 적성과 능력을 측정한다. 대학별 시험에는 종합시험, 소논문, 면접, 실기시험, 듣기테스트 등의 형태가 있다.

일본의 대학별 입시는 국공립대와 사립대로 구분될 수 있다. 사립대에 비해 국공립대의 입시가 훨씬 단순한데, 그것은 객관성과 공정성에 대한 기대가 높기 때문이다. 즉 국민의 세금으로 운영되고 진학희망자가 집중된 국공립대에 비해 사립대학은 그러한 사회적 기대가 덜하기 때문에 훨씬 다양한 방법이 활용된다.

대학별 고사 이외에 추천입학, AO입학(입학사정관제) 등의 입학전형방식이 있으며, 특정 프로그램과 연계하여 학생을 특별선발하는 프로그램도 있다. 추천입학의 경우 학교를 지정해서 추천받는 지정교제 추천, 고등학교 성적평점 등으로 추천되는 공모제 추천(일반 추천, 자기 추천, 특별 추천) 등이 있다. 지정교제는 대학들에서 자기 학부 합격자를 많이 배출하거나 지역적 특성을 고려해서 결정하는데 대학들에서 매년 재검토하여 학교를 지정하며, 추천을 받은 학생은 특별한 사유가 없는 한 대부분 합격한다. 공모제 추천에는 학교장 추천, 자기 추천(지도교사 평가서, 조사서 등), 특별 추천(스포츠, 문화활동, 사회자원봉사, 한 가지 재주에 뛰어난 학생, 각종 자격 및 기능 취득 학생 등) 등이 있으며, 서류심사만으로 전형을 하거나 서류심사와 다른 방법(면접, 소논문, 학력시험)을 병행하기도 하며 센터시험을 요구하기도 한다. 이밖에 일본에는 대학부속학교로부터 곧장 대학진학이 가능한 제도도 운영된다.

▫ 대학입시센터시험의 역사와 현황

앞서 살펴본 프랑스나 독일의 경우와 달리 일본은 대학입학시험이 중등학교 교육과는 독립적으로 비중있게 발달해왔다. 일본의 경우 1860년대 메이지유신 이후 서구식 근대화를 추진하기 시작하고 학제가 학교급별로 체계화되는 것이 1890년대에 가서야 이루어지기 때문에 서구 국가들과는 다른 경로를 취해 왔다고 할 수 있다.

더욱이 1945년 제2차 대전에서의 패전을 계기로 미국식 제도의 영향을 받게 되어 전전 교육과는 다른 체제가 형성되기도 했다.

전후 일본의 교육은 한편으로는 중등 및 고등교육의 빠른 대중화로 특징지워지며, 다른 한편

으로는 중등 및 고등교육에 대한 '총서열화 체제'라고 표현되기도 한다. 9년의 의무교육시기를 제외한 고교 및 대학의 경우 서열적 위계가 매우 심한 체제이며 이에 따른 입시경쟁이 매우 심각하다고 지적되어 왔다(竹內洋, 1995).

일본은 유럽국가들로부터 영향을 받아 대학제도를 만들었지만, 유럽국가들에서처럼 중등교육 졸업자격을 대학입학에 반영할 수 없었다. 아직 학교체제가 정비되지 않았기 때문에 대학에서의 전문적 교육을 받기 위해서는 별도의 판별기준이 갖추어져야 했다. 따라서 초기에 형성된 일본 대학들은 대학이 마련한 독자적 시험을 통해 입학생을 선발해야 했다. 이러한 맥락에서 설치되었던 것이 전전 일본에 존속했던 구제(舊制) 고등학교제도이다. 당시 고등학교는 전후 고등학교와 달리 일반 중등학교를 지칭하던 중학교 이후에 진학하는 제국대학준비 예비학교였다.

고등학교를 통한 대학진학준비는 1890년대 이후 형성된 제국대학 입학을 위한 필수과정이었다. 일반 사립들의 경우 일반 중등교육과정인 중학교만 졸업해도 입학이 가능했다. 고등학교 입학자격이 엄격했기 때문에 졸업자에게는 무시험입학이 이루어졌으나 1920년대 이후 대학진학 희망자가 증가하면서 무시험제는 폐지되었다.

일본에서는 1890년대 이후 고등문관시험을 통해 사법, 행정, 외교 분야의 공직에 대해 시험을 통해 선발하는 제도를 갖추는 등 시험을 통한 선발제도를 유지하고 있었다. 대학입학이나 관료입직이 사회적으로 매우 특권적 지위를 보장해주기 때문에 이를 정점으로 한 경쟁이 형성되고 하급단계 교육기관들은 이를 위한 과정에서 도구적 가치를 갖게 되는 '수험문화' 중심의 평가 · 선발의 문화가 형성되었다.

제2차대전에서의 패배이후 미군정 지배를 받게 되면서 초기에는 미국식 대학입학제도 등이 도입되었다. 즉 대학입학에서 고등학교 재학중 성적(조사서), 진학적성검사, 대학주관의 입학시험 등으로 이루어진 입시체제가 1954년까지 실시되었다. 입학시험 성적을 중시해온 전전 일본의 방식에 비추어볼 때 3가지 요소를 종합적으로 평가하는 미국식 입시제도는 상당한 변화를 의미하는 것이었다. 고교성적은 과거의 학력, 진학적성검사는 미래의 학력, 대학주관 입학시험은 현재의 학력을 가름하는 것이라고 하여 매우 합리적이고 공평한 방식이라고 간주되기도 했다. 그러나 미국식 SAT와 유사한 진학적성검사는 1954년까지 실시된 후 미군정 종식 이후 제도를 재조정하는 과정에서 폐지되었다. 대학들이 입학사정과정에서 중시하지 않는데다가 고등학교에서도 이중부담으로 작용한다는 비판때문이었다.

전후 일본에서는 고등교육에 대한 수요가 높아지면서 진학경쟁도 격화되었다. 이 과정에서 모든 대학이 공통으로 이용할 수 있는 국가수준의 표준화시험에 대한 요구가 높아졌고, 1970년대에 들어 국공립대학을 중심으로 공통시험에 대한 개혁요구가 제기되었다. 입시에서 지나치게 어려운 문항과 너무 생소한 문항의 출제로 인한 문제들을 완화하기 위해 정부가 국립대학협회의 의견을 수렴하여 공통시험을 도입하게 된 것이다. 1977년 문부성은 대학입시센터를 창설하고 국공립대학의 공통시험을 개발하게 된다. 1979년에 처음 실시된 공통 1차 학력시험에서 (2단계는 대학별 시험) 국공립대학 응시자는 모두 고등학교 교육과정에 준거한 5교과(국어, 수학, 영어, 사회, 이과의 5개 교과) 7과목의 공통시험을 보도록 요구되었다. 이러한 공통 제1차 학력시험은 1979년부터 1989년까지 실시되었다.

이 시험은 1990년 이후 대학입시센터시험(일명 '센터시험')으로 개칭되어 현재에 이르고 있다. 공통시험의 도입이후 사교육의 개입으로 인한 대학 및 학과서열화가 나타났다는 지적이 나타나면서 이러한 문제 해결을 위해 1985년 국립대학협회가 개혁안을 결정하고 개선안을 발표했다. 새로운 시험('신테스트')에서는 고등학교 교육과정에 충실한 양질의 문제 확보, 대학입시의 개성과 다양성에 기여, 사립대학의 참여 유도로 입시개혁, 대학별 필요 교과목의 자율 지정('아라카르트 방식') 등을 핵심으로 하는 변화가 시도되었다. 센터시험은 국공립대학 학생에게는 필수이며 사립대학에도 확대되는 추세이다.

대학별로 센터시험의 점수를 이용하는 방법이 모두 다르지만 크게 3가지 유형으로 구분된다. 첫째는 센터시험만으로 선발하는 방식이다. 일부 사립대 등에서 특별전형의 일부로 높은 기준을 부과하기 때문에 상당히 높은 득점이 요구된다. 둘째는 센터시험과 2차시험(소논문, 면접 등)의 결과를 종합해서 판정하는 형식으로 모든 국공립대학에서 취하고 있는 형태이다. 셋째는 센터시험을 1차선발에만 사용하고 최종판정은 대학별 본고사로 하는 방식으로, 도쿄대학이나 교토대학에서 취하고 있다. 응시자의 학력수준이 높기 때문에 2차시험 응시자격을 제한하는 목적으로 센터시험을 활용하는 것이다.

센터시험의 출제과목은 2009년 현재 6개교과 28개 과목이며, 시험은 연1회 1월 중순 2일(토, 일)에 실시된다. 출제과목은 모두 28개이지만 응시자의 필요에 따라 과목의 선택이 이루어진다. 최대 응시과목은 9과목이다. 상위권대학은 7-8개과목, 하위권대학은 3-4과목을 선택한다.

센터시험의 출제 체제를 보면 대학, 고등학교, 교육위원회 등에서 파견된 위원으로 구성된 위원회가 있으며 시험의 집행은 대학입시센터와 대학입시센터시험 이용 대학간에 공동 실시된다. 출제는 출제 관련 3개의 위원회를 중심으로 이루어진다. 센터문항작성을 위한 교과과목 제1위원회, 문항타당도 검토를 위한 제2위원회, 과목간 내용 적합성과 중복을 점검하는 제3위원회가 있다. 교과과목 1위원회는 과목별 20명의 전문가로 구성되며 2년 임기로 재임 중 개인정보를 공개하지 않는다. 업무연계를 위해 위원들은 1년동안 50%만 교체된다. 출제는 한 달에 1-3일간 보안시설에서 출퇴근하는 형식으로 이루어진다.

□ 국가수준 표준화시험과 책무성

<표 V-5> 일본 전국 학력·학습상황조사

시험 종류	전국 학력·학습상황조사 (의무교육학생 대상, 초등 6학년, 중학교 3학년)
주요 목적	학습상황파악을 통한 교육시책의 검증
실시 학교급	초등학교 6학년, 중학교 3학년
목표와 활용	목표: - 의무교육의 기회균등과 그 수준의 유지향상의 관점에서 전국적인 아동생도의 학력과 학습상황을 차악하여 교육시책의 성과와 과제를 검증하여 그 개선을 도모함 활용: - 학교에서 아동생도에 대한 교육지도의 충실과 학습상황의 개선등에 기여 - 이를 통해 교육에 관한 계속적인 검증개선 사이클을 확립함

주요 대상	– 소학교 6학년 전생도 (실 응시율: 2014년 4월 22일 99.2%) – 중학교 3학년 전생도 (실 응시율: 2014년 4월 22일 95.5%)
시험 과목	– 국어 (지식, 활용) – 산수. 수학 (지식, 활용) – 질문지조사 (아동생도에 관한 조사, 학교에 관한 조사)

일본은 2007년부터 학습상황 확인 및 지도 충실성을 위해 초등 6학년 중학교 3학년을 대상으로 전국 학력·학습상황조사를 실시하고 있다. 국어, 수학 중심의 표집평가로 실시되고 있다.[23)]

일본의 전국 학력조사는 1956년에 도입되었다가 과도한 경쟁조장을 이유로 중단되었으나 2007년에 다시 47년만에 부활했다. 학교별 성적 공표는 학교서열화를 우려해 금지해왔으나 2013년 11월 29일에 일본 문부과학성이 학교별 성적을 각 교육위원회의 판단에 따라 2014년부터 공표할 수 있도록 했다.

문부성은 분석결과와 향후 지도 대책도 함께 공지하도록 했다. 이러한 발표는 학력향상을 목표로 하고 있으나 아베정권의 도덕교육강화 등과 관련된 것으로 추정되기도 한다(연합뉴스. 2013.11.29).

23) 일본 문부과학성 국립교육정책연구소, 平成２６年度 全国学力・学習状況調査 調査結果のポイント, 文部科学省 国立教育政策研究所, 2014.8.

Ⅵ. 주요 쟁점 및 정책적 시사점

1. 요약

□ 평가정책의 국제적 수렴 경향

이 연구에서는 표준화시험이 출현하고 보편화되며, 학생선발 및 교육에 대한 책무성 점검의 기제로 강화되어오는 과정을 살펴보았다. 특히 표준화시험산업이 발달해 있고 국가수준에서 책무성평가를 강조해온 미국의 사례를 중심적으로 살펴보면서 프랑스, 독일, 일본의 사례도 검토하였다. 이들 나라에서 시험의 유형은 다르지만 대규모 표준화시험이 교육의 질관리와 유능 인재선발을 위해 국가수준에서 활용되어 왔음을 살펴보았다.

미국은 19세기중반이래 표준화시험을 실시하기 시작했으며, 프랑스나 독일의 경우도 19세기 초부터 잘 알려진 바깔로레아, 아비투어 등을 통해 국가수준에서 교육에 대한 질관리를 해왔다. 물론 바깔로레아, 아비투어 등은 중등졸업자격시험이자 대학입학자격으로서 학력관리를 위해 실시되는 책무성평가와는 다르다. 일본의 경우 대학입시센터시험이라는 표준화시험이 실시되고 있는데, 주요 국공립대학 등 상위대학입학에서 이러한 시험이 미치는 영향이 매우 크다.

1990년대 이후 대부분의 국가들에서 국가수준 학업성취도평가를 통해 초중등교육에서의 질관리를 하는 경향이 나타나고 있다. 미국은 NCLB법안 등을 통해 매우 강력한 책무성기제를 작동시키고 있고, 프랑스, 독일, 일본 등에서도 표준화된 시험을 통한 진단 평가 및 교육시스템 모니터링을 위한 주기 평가를 실시하고 있다. 이 연구에서 살펴본 국가 중에는 독일이 독자적 책무성평가를 가장 늦게 실시한 사례이다. 독일은 2009년 이전까지는 아비투어 이외에 중등 전문교육계열 학교들(하우프슐레, 레알슐레)의 졸업자격 시험을 실시할 뿐 별도의 국가수준 평가를 실시하지 않았다. 그러나 최근에는 학력점검을 위한 표준화시험을 도입하여 운영하고 있다.

주요 국가들에서 교육문화상의 차이는 존재하지만 표준화시험을 실시하고 평가를 통해 교육시스템을 관리하는 점에서 유사한 흐름을 발견할 수 있다.

□ 교사주도 평가 대 외부 평가

표준화시험은 그것이 교육에 미치는 영향으로 인해 비판이 되기도 하지만 다른 한편으로는 교육시스템 운영을 위해서는 불가결한 것으로 간주되기도 한다. 표준화시험의 기능은 학생의 학습필요를 진단하고 교수를 과학화하는데 기여할 수 있다는 측면이 있지만, 주로 행정의 용이성이나 사회공학적 필요를 반영해왔다. 그 기능이 어느 범위에 미치고 어떠한 형식으로 실시되며 그 결과가 얼마나 고부담인가에 따라 그 교육적 효과는 매우 상이하게 나타난다.

표준화시험의 역사를 볼 때, 교사에 의한 평가가 아닌 외부적 시험으로서의 표준화시험은 교육현장 혹은 내부에 대한 외부의 통제와 밀접하게 관련되어온 것이 사실이다. 미국의 경우 1840년대 최초의 표준화시험부터 최근의 NCLB에 이르기까지 표준화시험은 교육계 외부의 요

구를 강하게 반영하고 있을 뿐 아니라 교육계 내부를 개혁의 대상으로 간주해 온 특징이 있다. 즉 “교사주도평가 대 외부평가”의 구도가 강했음을 알 수 있다.

표준화시험이 단지 교사주도평가를 보조하는 역할을 할 뿐 아니라 객관성, 공정성을 담보하는 중요한 장치로 간주되어 교사주도평가가 지니는 개별성, 주관성을 넘어서고자 하는 목적을 지니고 있었음을 지적할 필요가 있다. 이러한 외부 평가로서의 성격은 프랑스, 독일의 시험들에서도 나타나는데 이들 나라에서도 19세기초 이후 국가 수준에서 바깔로레아, 아비투어 시험 등이 제도화되었다. 그러나 시험이 중등교육에서의 교육내용에 기반한 졸업자격고사의 성격을 지니고 있고, 출제나 채점, 판정 등에 이르기까지 중등교사들의 역할이 크게 작용함으로써 교사들에 대한 외부적 통제로서의 요소가 있지만 그 양상은 다르다고 할 수 있을 것이다.

□ 표준화시험의 교육적 역기능과 대안

표준화시험을 매개로 한 행정적 편의의 추구와 사회공학적 관심 즉 관리주의적 필요는 1980-90년대 이후 신자유주의적 정책의 적용으로 더욱 강화된 경향이 있다. 이는 국제적으로 보편화되는 방식으로 이루어져, 유럽연합의 보고에서 알 수 있듯이 유럽 대부분의 국가들이 1990년대와 2000년대에 들어 평가를 중심으로 한 시스템 관리체제를 형성하고 정비해왔다. 이 평가 관리체제는 표준화시험을 통해 매개되고 있지만, 국가수준의 시험이 학교교육에 주는 영향은 각 국마다 운용형태에 따라 다양하게 나타난다고 봐야 한다. 표준화시험은 책무성평가, 특히 고부담평가와 연계될 때 교육적 부작용이 심한 것으로 비판받는다. 특히 교육과정을 협소화하고 교사들의 교수, 평가 자율성을 저해한다는 것이 문제시된다.

알피 콘 등 비판론자들이 지적하는 것처럼 표준화시험은 책무성평가와 결합하기 전에도 이미 교육적이지 못한 것으로 주장되기도 한다. 가르치는 맥락과 시간과 공간적으로 분리되어 실시되는 표준화시험은 교수나 학습 그 자체를 돕는 것이라기보다는 변별이나 모니터링 등 외부자들의 필요를 충족하는 경향이 있다. 이것이 선다형문항이나 제한된 시간에 이루어지고 어린 학습자들에게 빈번하게 실시될 경우 그 폐해는 더욱 크다. 표준화시험이 객관성이나 공정성의 이름으로 옹호되고 일반인들에게 지지될 때 교육적 자율성이나 개개인 학습자의 필요를 충족하는 교육은 어렵게 된다. 대학입학수학능력시험이나 일제고사(내부형, 외부형) 등의 표준화시험이 학교의 교육과정을 지배하는 양상이 강한 한국 교육에서는 이러한 문제들이 역사적으로 누적되어 왔고 그에 기반한 교육문화를 형성해온 경향이 있다.

표준화시험에 대한 대안으로 교사의 전문적 자율성에 기반하여 학생 개개인의 학습필요를 충족시키는 교육적 접근이 필요하다는 점이 그 효과를 우려하는 논자들에 의해 꾸준히 주장되어 왔다. 현실적 부작용 이외에도, 20세기후반 이후 학습이론, 교육과정이론 등이 표준화에 대립되는 인식론에 기초한 (구성주의적) 패러다임을 취하는 경향에 비추어 볼 때 표준화시험은 이러한 교수철학과도 괴리된다.

외부에 의해 실시되는 평가는 교사를 탈숙련, 탈전문화할 뿐 아니라 학습자의 학습이 일어나는 상황과 괴리되어 실시된다는 점에서 비판된다. 교사와 학생이 협력적으로 학습과정에 참여함으로써 고등사고 및 문제해결능력을 높이는 방향으로 나아가는 것이 지향되어야 할 방향이라

면, 표준화시험은 이에 대한 걸림돌이 되고 있다고 지적된다.

2. 주요 쟁점

□ 교육을 협소화하지 않는 책무성평가의 가능성

시험과 평가를 통한 관리는 이제는 너무도 일반화되어 있다. 이에 대한 비판이 끊임없이 제기됨에도 불구하고 현실에서 뚜렷한 대안이나 출구를 찾기 어려운 실정이다. 소위 평가적 국가(evaluative state)로의 전환이 전세계적으로 전개된 1990년대 이후, 결과관리를 통한 통제가 심화되어 왔고 성취평가, 교원평가, 학교평가, 각종 기관평가 등의 체제가 형성되었다. 이에 대한 비판도 함께 심화되어왔음에도 불구하고 평가가 줄어들기보다는 오히려 강화된 것이 현실이다. 공교육이 공공성을 지니고 운영되기 위해 책무성에 대한 점검은 필수적일 수 있다. 다만 관리주의의 우위로 인해 교육활동을 지원하기 위한 목적으로 시행되어야 할 평가가 교육을 주도하는 전도양상이 나타나는 것이 문제이다.

이러한 평가의 심화와 강화의 역사는 어떻게 형성되어왔는가? 오늘날의 교육에서 문제가 되는 것은 평가문화(audit culture)라고 할 수 있을 것이며, 이 때의 평가(evaluation)는 시험(testing), 학업평가(assessment) 등과는 맥락이 다소 구분된다. 그럼에도 불구하고 교육활동에서 이 모든 평가는 연동되어 강조되는 경향이 있다.

교육에서 평가문화의 형성에서 주목할 수 있는 시기는 19세기 초중반부터라고 할 수 있다. 1840년대 미국의 사례는 표준적, 대규모적 평가를 책무성 점검을 목표로 실시했다는 점에서 주목된다. 프랑스의 바깔로레아나 독일의 아비투어 등도 18세기 말, 19세기 초에 실시되기 시작했고 국가에 의한 교육통제의 목적이 내재되어 있었지만, 기본적으로 자격고사 및 선발시험의 성격을 가지고 있어 미국의 시험과 기능차이가 있다고 하겠다.

교육에서 평가정책은 외부기준과 내부기준의 대립 혹은 경쟁을 쟁점화한다. 현실에서, 관리주의의 우세로 인해 책무성평가의 비중이 높지만 이에 대한 개선의 목소리도 큰 만큼 내부기준과 외부기준의 관계를 재조정하는 대안의 모색이 필요하다. 책무성이 중요하다면, 외부에 의한 표준화시험에 의존하는 현재와 같은 방식으로만 가능한가? 학습과 발달의 논리에 맞는 평가를 통한 교육적 책무성확인은 가능하지 않은가? 교사 전문적 자율성이나 학습자의 전면적 발달에 배치되지 않으면서 교육적 책무성을 제고하는 것은 어떻게 이루어질 수 있는가?

책무성은 공교육의 의무와 책임과 관련한 준거이기 때문에 그 대상은 국가 및 국가로부터 교육 권한을 위임받은 학교들이다. 엄밀히 말해 책무성이 표준적 방식으로 동일하고 획일적 방법으로 확인되어야 할 이유는 없으며, 현재와 같은 관리방식이 본격화된 것은 20세기 후반 신자유주의원리의 전면화와 관련되어 있다. 즉 결과 혹은 산출 중심의 방식을 취하다보니 관리의 효율화를 위해 비교가 용이하고 측정가능한 지표들이 강화된 것이라고 볼 수 있다.

여기에는 측정될 수 있는 것이 객관적이고 과학적이라는 사회적 신념과 상식이 결합해 있다. 이러한 관리 및 효율중심의 원리가 교육의 일부를 확인할 수는 있지만 전체를 확인할 수 없고, 그 확인방식으로 인해 보다 의미있는 원칙과 활동을 제한시킬 수 있다는 점은 논란의 대상이

되어왔다. 즉 현행 관리 방식이 학교교육에 지속적인 역효과를 야기하고 있다면 이는 장기적으로는 사회적 비용으로 나타나게 될 것이며, 따라서 전적으로 효율적인 방식이라고 볼 수도 없다. 따라서 장기적 효용의 관점에서도 관리의 방식에 대한 조정이 필요할 수 있다.

관리의 효율을 위한 방식은 그 자체로서는 책무성의 유일한 점검방식이 아니며, 학습과 발달의 관점을 고려한 개념으로 책무성에 대한 이해가 재개념화할 필요가 있을 것이다. 또한 대규모, 획일적, 표준화방식과 고부담평가라는 형식이 책무성평가에서 반드시 필요한가에 대한 교육학적, 사회공학적 재검토가 필요할 것이다.

□ 공교육의 목적 및 대안적 평가

1990년대 이후 대부분의 국가에서 표준화시험을 통한 학습상황 진단 및 교육시스템 모니터링을 실시하고 있다. 국가별로 시험의 표준화정도나 시험의 학교교육에 대한 영향력은 교육문화에 따라 차이가 난다. 미국처럼 학교의 학력관리를 연도별로 고부담평가에 의해 엄격하게 관리하는 체제에서는 시험준비교육(teaching to the test)과 관련한 교육과정협소화 등 각종 부작용이 심각하게 제기된다. 표준화시험이나 국가수준학력평가를 고부담평가로 연결시키지 않는 경우 이에 대한 부담의 강도나 학교교육에의 부정적 영향력은 상대적으로 낮을 것이다. 그러나 이 경우에도 외부적 학력관리체제의 존재 그 자체로 인해 학교현장의 평가 자율성은 기본적으로 제한되거나 영향을 받을 수밖에 없다.

교육행정에서의 관리주의의 발달은 다른 한편으로는 교육학에서의 학습이론이나 교육과정원리 등의 발달 방향과는 상충되는 경향이 있다. 특히 개방적 교수목표를 설정하고 학습상황에서의 맥락성과 사회적 상호작용을 중시하는 교수학습원리에 비추어 볼 때, 외부기준에 의한 표준화시험은 기본적 교육철학이나 원리가 판이하게 다르다.

공교육이 자기 역할을 하고 있는가에 대한 점검과 지원은 이루어질 필요가 있지만 그 근본 목적에 충실하게 이루어져야 할 것이다. 외부평가는 확인과 지원의 목적으로 이루어지는 것이 타당하며, 교육의 방향을 좌우하는 수준으로 영향을 미쳐서는 곤란할 것이다. 학교에서의 평가를 표준화된 책무성평가와 지나치게 긴밀하게 연계하는 미국의 사례는 이러한 점에서 비판의 타깃이 되어왔다고 할 수 있다.

표준화시험의 폐해는 대안적 평가에 대한 논의를 활성화시키고 있다. 대안적 평가는 참평가를 지향하는 경향이 있으며, 수행평가, 포트폴리오 등의 형식을 선호한다. 참평가(authentic assessment)에서는 학생들이 핵심 지식이나 기능을 실제 세계의 과제들과 연관하여 적용하도록 한다. 참평가에서는 교수과정과 평가가 분리되는 것이 아니라 학생이 해결해야 할 과제를 위주로 교육과정이 구성된다. 이러한 평가방식은 전통적 평가(traditional assessment)와 구분된다. 전통적 평가에서는 학생들이 기억한 정보를 회상하도록 요구하며, 선다형, 진위문항 등을 포함한 필답시험의 형식을 취한다. 전통적 평가에서 학생의 학습이나 이해에 대한 간접적 증거들을 통해 평가가 이루어진다면 참평가에서는 학생들이 직접적으로 그 능력을 수행해보이도록 함으로써 직접적 증거에 기초한다.

특히 사회적, 협력적 상호작용과 발달과 이해과정에서 교사의 적절한 개입을 중시하는 사회

적 구성주의(비고츠키이론을 포함)에 기초해볼 때 평가는 수업이나 교육활동과 분리되어 실시되기보다 학습상황 속에서 진단, 피드백, 지원 등이 역동적으로 이루어지는 방식을 취할 필요가 있다. 또한 학습자에게도 자기평가와 성찰을 하도록 함으로써 교사-학생 간의 협력적 수업과 평가가 이루어지도록 하는 것도 학습과 발달에 중요하다.

수업의 특성과 교수학습의 목표에 따라 전통적 평가와 대안적 평가는 상호보완적으로 이루어질 수 있을 것이다. 교육 책무성도 이와 다르지 않다. 책무성 확인을 위한 유일한 방법은 없으며 교사의 학습지원능력을 제고하는 것이 책무성을 높이는 것이라는 점을 인식할 필요가 있다.

□ 선발에 있어 표준화시험의 역할

교육에서 평가는 학습상황에 대한 진단이나 책무성 점검 이외에도 상급학교진학시의 선발과도 관련이 된다. 상급학교가 직접 입학시험을 실시하는가, 하급단계학교의 자료를 전형에 반영하는가에 따라 교육문화는 다른 양상으로 나타날 수 있다. 또한 전형자료에서 표준화시험의 내용과 형식, 그리고 기능이 무엇인가에 따라서도 교육제도의 성격은 다양하게 나타난다.

많은 국가들의 대학입시에서 표준화시험이 활용되고 있는 경향이 있지만 시험의 성격에는 차이가 있다. 미국의 경우, 20세기초부터 IQ검사의 원리를 응용한 SAT시험이 실시되기 시작했고, 이는 학교교육과정에 따른 성취시험인 ACT와 대립해왔다. 미국의 경우 표준화시험개발 및 보급이 매우 일반화되고 활성화되어 있지만 국가가 아닌 민간기구가 이를 관장하고 있다는 점이 특징적이다. 또한 측정평가의 발달로 선다형 문항을 중심으로 시험이 발달해온 점도 주목된다. 이러한 표준화시험은 주별로 다양하게 활용되고 있다. 주에 따라 평가업체에서 개발된 표준화시험을 학력평가에 사용하기도 하고 졸업자격부여에 활용하기도 한다. 대학입시에 주로 인정되는 시험은 SAT, ACT 등인데 ACT의 경우 교육과정에 대한 성취평가의 성격을 지니고 있어 일부 주에서는 고교단계에서 응시를 의무화하기도 한다. 미국의 대학들에서는 매우 다양한 입학사정방식을 유지하고 있다. 고등학교 학점, 표준화시험, 에세이 및 교과외활동, 추천서 등을 다양하게 반영한다. 미국은 초중등 단선제학제와 높은 대학진학율로 인해 고등교육도 상대적으로 대중교육화된 체제로 이루어져 있다. 대학들은 지원자 대비 입학자수에 따라 선발경쟁의 정도가 구분되지만 고등교육기관의 수와 종류가 다양하고 입학기준도 다양하기 때문에 대학들에 대한 엄격한 범주화를 하기가 쉽지 않은 체제이다.

유럽의 경우에도 표준화시험은 19세기초부터 일찍이 실시되고 있음을 알 수 있다. 프랑스의 바깔로레아나 독일의 아비투어는 모두 국가가 주관하는 표준화시험이지만 그 형식이 서술형이고 고교교사들의 출제 및 채점 위원으로 참가한다는 점이 민간 전문평가업체 중심으로 시험을 개발하는 미국과 대비된다. 또한 기본적으로 고등학교까지의 성취를 기본으로 하여 대학에서의 수학능력을 연계적으로 확인하고자 하는 이중 기능을 지니고 있는 점도 주목된다. 졸업자격시험이면서 입학시험이며, 고교까지의 교육이 대학입학의 준비과정으로 연계되어 기능하고 있음을 알 수 있는 것이다. 유럽국가들에서는 전통적으로 복선형제도의 발달로 진로에 따른 조기분화가 이루어져 왔기 때문에 미국이나 아시아국가들에 비해 대학 진학자의 비율이 높지 않다. 1960년대말 이후 대학진학자의 수가 급증한 것으로 보고되지만, 인문계 중등학교 진학이 이미

엘리트교육적 성격을 지니고 있다고 할 수 있다. 따라서 졸업자격시험만으로 대학입학을 하는 체제를 유지하고 있지만 선발의 객관성이나 공정성에 따른 논란이 많지 않고, 입학을 위한 선발경쟁이 심하지 않다. 프랑스의 그랑제꼴처럼 바깔로레아취득 이후 몇년간의 부가적 준비가 필요한 고등교육기관도 있지만, 일반 대학들의 경우 비교적 균등한 수준(평준화)을 유지하고 있어 졸업자격시험의 결과가 대학입학자격과 무리없이 연동되고 있다.

일본의 경우 고교졸업과 대학입시가 분리 운영되는 체제를 취하고 있다. 일본에서는 고교와 대학이 총서열체제를 이루고 있다는 평가를 받는데, 19세기말부터 학교간 서열화가 심한 체제를 유지해왔다. 일본에서는 학생선발을 위한 대학별 입시가 발달해 있고, 유럽에서처럼 고등학교까지의 교육과 대학입학 간의 연속적 연계를 설정하고 있지 않다. 일본에서는 국공립대학과 사립대학에서의 입시에 차이가 나타나는데, 전통적으로 위상이 높은 국공립대학은 높은 경쟁률과 선발의 공정성 및 객관성을 내세워 표준화시험(대학센터시험)을 강조하며 사립대학의 경우 입학사정관전형(AO) 등 다양한 요소들을 중시한다. 일본에서 대학입학 표준화시험은 국가기구가 관장하고 있지만 국공립대학들간의 협의체를 중심으로 시행되고 있다는 점에서 특수성이 있다. 일본의 경우 패전후 중등교육이 빠르게 완전취학화, 대중교육화함으로써 선발경쟁이 매우 심하게 나타난 특징이 있다. 의무교육 이후 기간인 고교 이후 단계에서 학교간 서열화가 심하여 중학교부터 입학준비를 위한 경쟁과 사교육이 문제시되고 있다. 대학입시를 위한 표준화시험인 대학입시센터시험의 경우 고교 교육과정을 반영하지만 대학이 주관하는 외부시험이기 때문에 학교에서의 교육평가와 연계된 시험은 아니며 별도의 준비가 요구된다.

진로에 따른 분화가 어느 시기에 이루어지고, 사회적 선발의 결정적 시기가 언제인가에 따라 평가를 둘러싼 교육문화는 상당히 다르게 나타난다. 선발을 위한 표준화시험의 경우도 개발 및 시행의 주체의 성격, 시험의 내용과 형식, 대학입학에서의 사정방식 등에 따라 그 교육적 효과는 상이하게 나타난다.

3. 한국교육현실에 대한 시사점과 변화의 방향

□ 객관성과 공정성의 평가문화 재고

미국에서 발달한 표준화시험은 객관성과 공정성의 확보를 가장 중요한 정당화 근거로 갖고 있다. 이는 주관적이고 맥락내적인 평가와 대조된다. 평가는 그것이 교수학습의 직접적 지원을 목적으로 하는가, 변별과 선발의 도구로 사용하는가, 교육전반의 질에 대한 모니터링을 목적으로 하는가 등에 따라 보편적 기준(표준)에 의한 객관성이 강조될 수도 있고 교사의 전문적 관찰과 개입에 의한 주관성이 강조될 수도 있다.

미국의 경우 주별 자치에 의해 교육제도가 다양하게 운영되고 있기 때문에 일반적 수준에서 상대적으로 객관성과 공정성이 강조되어 왔다고 볼 수도 있다. 복잡한 현실을 단순화하고 빠른 정보를 얻어내고자 하며 즉각적이고 효과적인 처방을 구하고자 하는 실용주의적 문화와 이를 뒷받침하는 정교한 측정방법이 발달해온 것도 이러한 조건과 관련이 될 것이다. 또한 20세기초

에 중등교육이 빠르게 확산됨에 따라 학생들의 진로에 따른 교육과정차별화와 분류배치의 필요가 높았고, 대중교육체제를 관리하기 위한 효율적 행정의 필요 등이 표준화시험의 발달을 부추겼다고 볼 수 있을 것이다. 1980년대 이후 경제위기로 교육의 질 제고가 문제시되었을 때도 이를 국가 수준에서 효율적으로 관리하기 위한 방편으로 표준화시험에 의한 학업성취도와 학교의 책무성을 직접 연결시키는 조치들이 손쉽게 취해진 것도 이미 발달해 있는 표준화시험 문화가 뒷받침이 되었기 때문이라고 할 수도 있다.

다른 한편, 미국은 표준화를 반대하는 저항적 흐름도 거셌다고 할 수 있다. 19세기말 이후의 진보적 교육은 엘리트주의적 전통교육에 대립하는 공통 요소와 함께 표준화, 과학화, 효율성을 중시하는 하나의 흐름과 개성, 다양성, 시민성, 인권, 참여, 민주주의 등의 가치를 중시하는 다른 흐름 간에 차이를 드러내고 있었다. 표준화에 대한 요구가 한편에서 높아졌다면 다른 한편에서는 이에 대한 반대 흐름도 강하게 형성되어 왔다. 1960년대에서 1980년대까지의 기간은 표준화에 반대하는 경향이 상당히 지배적이었던 시기이다. 표준, 중립성, 객관성에 대한 근본적 문제제기들이 이루어지고, 공정성에 대한 기준들이 다양하게 논의되었다. 학력과 국가경쟁력의 문제를 심각하게 거론하면서 강한 책무성 정책을 사용하게 된 1980-90년대 이후, 일련의 담론과 정책들은 표준화시험의 활용을 극대화시켜왔고 이에 대한 비판도 매우 강렬하게 제기되고 있다고 하겠다.

1980년대 이후 미국과 영국을 중심으로 시작된 표준화시험과 책무성의 연계 강화 경향은 국제적으로 수렴되는 경향을 보이고 있다. 1990년대 이후 유럽연합에서는 대부분의 국가가 표준화시험을 학력진단과 교수학습지원의 목적 이외에 교육의 질관리 기제에 활용하고 있다. 이 과정에서 미국을 중심으로 이루어져온 선다형 문항유형을 도입하는 사례들도 늘어나고 있다.

한국의 교육문화는 미국 교육사와는 매우 다른 흐름을 가지고 있지만 외관상의 유사성도 있다. 한국의 경우 객관성, 공정성을 중시하는 평가문화를 가지고 있다는 점이 특징적인데 이는 선발이 매우 중요한 비중을 차지해온 것과 관련이 된다. 한국에서는 상급학교 입학 선발이 학교교육에 결정적인 영향을 미쳐왔을 뿐 아니라 취업 선발과정도 시험의 형식으로 이루어져 있어 일관적인 평가문화가 형성되는데 영향을 주었다고 할 수 있다. 객관성, 공정성을 표방하는 시험은 대체로 점수 등으로 표시되는 결과에 의존하게 되고 사고나 학습의 과정을 파악하기는 어렵다. 또한 특정 영역에 필요한 지식, 기능, 태도와는 무관하게 일반적인 능력 혹은 일부 능력을 중심으로 선발을 하게 되는 경우가 많다. 이는 표준화시험의 단점으로 지적된 문제들과 일치한다.

가드너는 형식적 시험 모형과 도제적 모형을 구분하면서 특정한 분야의 직능을 중심으로 선발을 해야 할 경우 도제적 모형이 타당하고, 형식적 시험모형은 맥락을 배제한 평가이기 때문에 타당도가 낮다고 보았다. 역사적으로 이러한 형식적 시험모형의 대표적 사례는 중국의 관리선발제도(과거시험)이다(Gardner, 1992). 앞서 언급했듯이 중국에서 천년 이상 유지되어온 이 제도를 근대 이후 서구 국가들이 이러한 제도를 수용하게 되었다는 논의들이 있다(Teng, 1943).

한국은 고려 광종 이후 시험에 의한 선발제도를 유지해왔으며 최종 선발자의 성적은 엄격하게 서열화되어 관직부여의 기준이 되었다. 갑오경장으로 과거시험은 제도적으로 폐지되었지만,

식민지시기에는 일본의 고등문관시험(사시, 외시, 행시)이 실시되어 고급 관리들에 의한 선발이 시험에 의해 이루어졌다. 이는 일본 관료 선발을 목적으로 한 것이지만 학벌과 무관하게 시험점수에 의한 당락으로 선발이 이루어지기 때문에 나름대로 객관성과 공정성을 표방하는 제도였던 것이다. 또한 교육기회가 제한되어 있는 조건에서 각급 학교의 입학시험이 존재했고, 고등교육단계에서도 대학별 입학시험이 있었다. 이러한 관례는 해방후까지 존속해왔다. 학교에서의 교수학습의 맥락과 선발에서 요구하는 요건의 차이는 진학이나 취업의 욕구가 강한 상태에서는 교육보다는 선발에 치중하도록 요청한다. 특히 해방후 교육기회가 급격하게 개방·확대되고 학력이 지위결정의 핵심적 요인이 되면서 상급학교 입시경쟁은 과열양상을 띠게 되었다. 중등교육 평준화, 대학정원자율화 등으로 경쟁요인을 완화하는 조치들을 취했음에도 불구하고 제한된 기회배분을 둘러싼 경쟁은 실질적으로 줄어들지 않고 경쟁으로 인한 각종 사회적 비용을 지속적으로 발생시키고 있다. 특히 과열경쟁 속에서 시험점수가 능력을 재는 데 있어 객관적이고 공정한 것이라는 생각은 강화되어 왔다. 각종 경쟁 완화 조치에도 불구하고 경쟁이 줄어들지 않는 것은 학교간 위계가 존재하고 학벌의 효과가 크기 때문이라고 할 수 있다. 한 줄 세우기가 아닌 여러 줄 세우기라는 표어가 등장하여 제도적으로 시행된지 10여년이 지났지만 여전히 점수가 실력이고 객관적인 것이며 공정성의 기준이라는 인식은 근본적으로 바뀌지 못하고 있다. 대학입학선발고사가 학교 교육과정을 지배하는 양상도 거의 개선되지 못하고 있다. 입학사정관제도, 학생부전형강화, 성취평가제 도입 등 학교교육을 정상화하려는 시도들은 꾸준히 제시되어 왔지만 이를 다시 '표준화' 하려는 문화가 함께 작동하고 있다. 사회의 경쟁, 변별, 선발 압박에 학교가 자유롭지 못하고 끊임없이 맞추게 되는 조건들이 형성되고 있다.

제도적 외피는 바뀌었지만 시험의 형식만을 놓고 볼 때 전통사회로부터 현대까지 한국 사회에서는 유사한 평가문화가 이어지고 있다고 해도 과언이 아니다. 적어도 형식적 시험인가, 도제적 평가인가의 관점에서 볼 때 한국에서는 후자의 관점이 역사적으로 득세하지 못했다. 조선시대에도 유형원같은 학자는 단회적인 필기시험에 의해 현덕자를 가려내기 어렵다고 보았고 중국에서 과거제 이전에 활용되던 인재 천거방식의 도입을 주장했다. 학문의 초기단계부터의 양성과정을 강화하고 단계적으로 상급기관으로 천거해 올리는 제도의 전면실시를 기획한 바 있다. 유능하고 현명한 관리를 선발하는 것이 목적이면 그것에 합당한 평가의 타당도가 높은 선발방식을 취해야 하며, 양성과 선발이 연계되도록 하자는 이 주장은오늘날의 학교교육정상화 논의와 유사한 관점에 터해 있다고 할 수 있다. 물론 이러한 제도는 현실화되지 못했다. 과거시험은 지필시험이기는 했지만 체계적이고 오랜 학습을 전제로 한 서술형시험이고 적어도 당시의 표준적 지식이라는 한계 안에서는 상당한 논리력과 지식응용력을 가져야만 최고 수준의 답안을 쓸 수 있는 시험이기는 했다. 뛰어난 유력 문인관료들인 성삼문이나 이이 등이 문과 장원급제자들이었던 점을 고려하면 우수한 자들을 가려내는 시험의 기능을 했다고 볼 수도 있을 것이다.

서양 근대의 산물인 능력주의가 중국의 전통적 과거시험과 같은 이질적인 시험형태를 취함으로써 관료선발시험과 같은 능력주의 기제를 형성시킨 점은, 동아시아 근대화과정에서 과거시험이 폐지되는 것(한국에서 1894년, 중국에서 1905년)에 비추어 보면 상당히 역설적이다. 19세기 말까지 존속한 과거시험의 문화적 효과는 보다 심각하게 재고할 필요가 있을 것이다. 전통사회

에서 양천의 신분제약이 있었지만 과거시험은 양인 이상에게 개방된 시험이었고 그 시험의 결과에 의해서 관직이 서열적으로 부여되는 제도였기 때문에 근대 능력주의 원리와 기본적으로 통하는 부분이 있는 것이다.

식민지시기 고등문관시험이나 대학입학시험도 일본어를 모국어로 쓰는 자와 그렇지 않은 자들 간의 원천적인 차별은 있었지만, 개인의 능력으로 해소될 수도 있는 문제였다. 1924년 경성제국대학 1회 입학자인 유진오와 같은 인물이 일본인과의 경쟁에서도 수석을 할 수 있었던 것은 이 모든 장애를 극복한 결과였던 것이다. 오히려 1930년대 이후 지식위주 입학시험에 대한 비판으로 학생부, 추천서를 강화하는 전형으로 전환하면서 한국인 입학에 장애가 생긴 사례들이 많다(광주학생운동 이후 해당 학교 학생들 입학 거부 등). 이러한 전환은 당시 시험위주문화를 바꾸기 위한 노력이기는 했지만 일제의 군국주의적 통제논리와 맞물려 악용된 것이었다. 흥미로운 것은 식민지시기 경성제일고보(현 경기고)에서 교사들 간에 큰 논란거리가 된 사건 중 경성제대 입학자이지만 출석미달자인 학생에게 졸업장을 줘야 하는가 하는 문제가 있었다. 결국 논란 끝에, 불성실하기는 해도 제대입학을 한 이상 앞길을 막을 수 없으니 이 학생에게 졸업장을 부여하게 된다(이충우, 1980). 학교교육보다 상급학교진학이 우선시되는 관행이 여실히 드러나는 대목이고, 이러한 양상은 오늘날까지도 크게 바뀌고 있지 않다.

이러한 평가문화는 여러 줄 세우기를 위한 다양한 전형의 도입에도 불구하고 근본적으로 변화하지 못하고 있다. 이를 위해서는 노동시장구조, 학벌구조 등이 변화해야 하지만 무엇보다 평가의 객관성과 공정성을 대하는 시각이 문화적으로 변화해야 할 것이다. 적어도 교육 장면에서 가르치고 배우는 맥락이 중시되고, 개개인의 다양한 능력이 존중되기 위한 다양성 기준에 대한 신념이 형성되고 강화될 필요가 있을 것이다.

□ 대안적 평가문화의 형성

여러 나라에서 표준화시험은 대학입학선발, 학교책무성강화 등을 위해 활용되고 있으며 '빠른 언어'가 지니는 장점도 있다. 그러나 여러 비판적 논의에서 제기되는 것처럼 이러한 시험에 대한 의존도가 지나치게 높아서 교수학습과정을 지배하고 인간의 능력에 대한 홀리스틱한 접근을 대체하고자 한다면 문제가 클 것이다. 표준화시험은 빠르고 쉽게 정보를 제공하는 방식이기는 하지만 보다 깊이 있는 배움을 이끌어내지 못한다면 교육적 의미는 약화될 것이다.

첫째, 변별보다는 성장과 발달을 중심으로 한 평가를 지향해야 한다. 교수학습과 평가가 연계되면서도 분리되기 위해서는 교육에 기반한 평가철학이 정립될 필요가 있다. 이제까지 표준화시험의 과도한 지배를 보완, 대체하기 위해 교수학습과정의 맥락에 충실한 평가방법들이 제안되어 왔다. 수행평가, 포트폴리오식평가, 참평가 등은 표준화하고 비교하기보다는 개인의 성장과 발달을 중심으로 하여 접근하는 방식으로 주목되어 왔다. 또한 교사의 교실평가(classroom assessment)의 중요성도 부각되어 왔다. 교수학습이 실제 일어나는 장면에서 배움과 평가가 동시에 일어나도록 하는 접근이 교육적으로 왜 중요한가의 문제는 평가철학이기도 하지만 교육관의 문제이기도 하다. 교육을 도구적 관점에서 바라보지 않고 개인의 성장 발달관점에서 접근하는 교육적 고려가 필요하다. 이를 위해서는 다양한 교육외적 요소들이 교육활동

에 개입하지 않도록 교실평가의 자율성이 충분히 보장될 필요가 있다. 변별 압박, 공정성에 대한 압박 등으로 인해 교육적 평가보다는 문제없는 평가를 지향하도록 하는 것이 현행 평가의 현실적 문제점이다. 객관성, 공정성에 집착하는 것을 비판하면서도 교사 평가의 전문적 자율성에 대해 불신하고 개개인의 발달에 대한 주관적 측면을 허용하는 것에 대해 불안함을 가지고 있는 것이 현실이다. 이러한 불안감은 북유럽과 같이 단위학교 자율성이 높은 나라에서도 나타나지만 이 불안이 외부적 모니터링에 의해서가 아니라 교사들의 전문적 협력공동체 안에서 해결될 수 있도록 방향을 잡는 것이 중요하다.

둘째, 표준화시험 그 자체보다도 표준화시험의 특정 형식과 유형이 교육과정을 협소화시키고 교육내용을 제한한다는 점을 고려할 필요가 있다. 유럽에서 실시하는 졸업입학자격시험인 아비투어나 바깔로레아의 경우 서술형평가로 이루어져 있어 그 성격상 표준화시험임에도 불구하고 자료분석능력, 지식적용능력, 이해 및 탐구능력 등을 종합적으로 확인할 수 있도록 이루어져 있다. 평가를 통해 깊이 있는 교수내용과 학습방법을 유도할 수 있다면 표준화에 따른 논란 그 자체의 문제는 상대적으로 크지 않을 수도 있다.

앞서 살펴 본 바 있는 바, 최악의 시험은 '상대평가로, 선다형에 의해서, 제한된 시간에, 반복적으로, 1학년부터 실시되는' 시험이라고 보는 알피 콘의 주장은 반드시 교육적으로 고려할 필요가 있다. 표준화 그 자체가 때때로 불가피하다 하더라도, '최악의' 형식을 피하는 방법은 교육적으로 고려될 필요가 있는 것이다. 특정 형식과 유형에 치우치다 보면 시험의 형식으로 인해 제한된 능력만을 확인할 수밖에 없고 그러다 보면 시험의 형식으로 인해 잘못된 교수학습방식과 학습습관이 의도하지 않게 형성될 수 있는 것이다. 즉 하나의 정답을 찾는 선다형 문항에 익숙한 학생들은 주어진 문제에 대해 스스로 사고하기보다는 추측을 하게 될 가능성이 높고 그러한 방식으로 사고가 반복적으로 훈련되다보면 다양한 관점에서 맥락적인 사고를 하는데 제한을 받는다. 또한 문제는 선다형시험 그 자체보다도 선다형시험을 반복적으로 준비하면서 갖게 되는 사고 습성이 문제라고 할 수 있다. 실생활에서 문제해결력을 가진 주체가 되기 위해서는 학생들이 열린 질문들에 대해 스스로 사고할 수 있는 능력을 갖추는 것이 중요하다. 고부담 선다형 표준화시험 준비를 위한 반복훈련에 익숙한 교육문화에서 이러한 능력은 키워지기 어렵다. 평가문화가 형성하는 사고 습관은 고등사고를 훈련하기 위한 각종 교육과정과 충돌할 수 밖에 없다.

셋째 표준화시험이 고부담평가와 연계되면서 나타나는 교육적 폐해는 매우 직접적이다. 미국의 NCLB의 경우 학력부진학생비율을 학교의 책무성과 직접적으로 연결하다 보니 교육과정협소화양상이 문제시되어 왔다. 문제는 부유층 학교의 경우 이러한 현상이 심각하지 않은데 비해 저소득층 학교일수록 부진아비율을 낮추려다 보니 교육과정이 표준화시험에 맞춰지는 양상이 높아진다는 것이다. 표준화시험과 직접적으로 연계된 교과들의 비중이 높아지다 보니 학생의 전인교육 기회가 줄어들고, 결국 사회불평등이 심화된다는 지적을 받게 되는 것이다. 표준화시험에 의한 진단은 장기적 모니터링 등을 위해 행정적으로 필요할 수도 있지만, 이를 학교책무성과 직접적으로 연결시킬 경우 학습자개인의 필요에 대응하는 접근보다는 평균 수치를 높이기 위한 접근이 강화될 수밖에 없다. 교육과정을 협소화하지 않는 책무성 점검의 방식이 교육적으로 고민될 필요가 있다.

넷째 과학성, 객관성, 공정성 등을 표방하며 질 관리를 하고자 하는 평가문화(audit culture)가 형성됨으로써 얻는 것과 잃는 것이 무엇인지 엄밀히 따질 필요가 있다. 표준화시험의 역사에서 드러나듯이 표준화시험은 외부평가이며 교수학습의 맥락을 제한하거나 통제하는 효과를 지닌다. 표준화시험을 통해 학교의 내부를 단순화한 지표로 드러냄으로써 교수학습의 일반적 수준을 확인하고 학생들의 진보정도를 파악할 수 있다. 교사와 학교의 역할이 학생의 성취(평가결과)와 동일시되어야 하는가의 문제는 논란이 된다. 그러나 이러한 평가결과를 통해 학생들의 일반적이고 전체적 수준을 제한된 수준에서나마 파악하는 것이 가능하다.

문제는 이 결과를 어떻게 해석하고 교수학습이나 선발, 정책제도에 어떻게 반영할 것인가의 문제이다. 평가결과가 무엇을 측정한 것인가에 대해 섬세하게 접근할 필요가 있다는 것이고, 액면 점수만으로 결과를 이해했을 때 어떤 오류를 범할 수 있는지 종합적인 고려가 필요하다. 평가 결과를 통해 교수학습의 질을 어느 수준까지 설명할 수 있고 그 개선방안에 대해 무엇을 말할 수 있는지 질문해봐야 한다. 평가결과에 영향을 미친 다각적 요인들에 대한 엄밀하고 공정한 분석 없이 학교 간 차이 등이 단순 비교되는 오류들에 대해 문제제기를 하지 않는 것은 주어진 점수의 객관성과 과학성에 대한 일종의 신화적 믿음이 전제되기 때문이다. 대학입학상의 다양한 전형요소들이 변별을 위해 결국 표준화시험(수능)의 점수로 귀결되는 것, 표준화시험의 결과가 그대로 고부담 정책으로 귀결되는 것 등은 평가를 대하는 문화를 반영하는 것이다.

표준화시험의 결과가 전체를 보여줄 수 없는데도 전체를 본 것으로 간주한다면 그것은 객관적이지도, 공정하지도, 과학적이지도 않다. 총체성이 드러나야 전체에 대한 판단이 가능하게 되는 것이다. 이를 위해서는 '빠른 언어(quick language)'를 넘어서는 종합적 접근이 필요하고, 빠른 언어가 아니라 복합적이고 맥락적인 질적 특성을 포함하는 '치밀한 언어(thick language)'를 사용할 수 있어야 하는 것이다.

부록1: 미국에서의 표준화시험 – 대학입시와 학력평가

미국에서의 표준화시험은 크게 대학입시와 학력평가의 두 차원에서 살펴볼 수 있다. 대학입시에서는 20세기 초부터 활용되어온 SAT와 ACT 등의 시험이 있으며, 일반 학력평가는 NCLB 법에 의해 시행되는 전국 단위 시험과 관련이 된다. 특히 후자는 학교의 책무성을 점검하는 평가의 성격을 지니고 있어 교육문화에 크게 영향을 주고 있다.

1. 대학입시에서의 표준화시험

가. 대학입시의 기본 체제[24)]

미국의 대학 입학제도는 대학마다 전형 방법과 전형 요소가 다르다. 학교의 규모, 공·사립 등 설립 유형, 사회적 책무성 고려의 정도 등에 따라 다양하다. 그러나 이러한 다양성에도 불구하고, 미국 대학 입학제도는 선별성(selectivity)의 정도를 기준으로 크게 다음 세 유형으로 분류가 가능하다.

- 개방형 입학제도(open admission)
- 선별형 입학제도(selective admission)
- 경쟁형 입학제도(competitive/ highly selective admission)

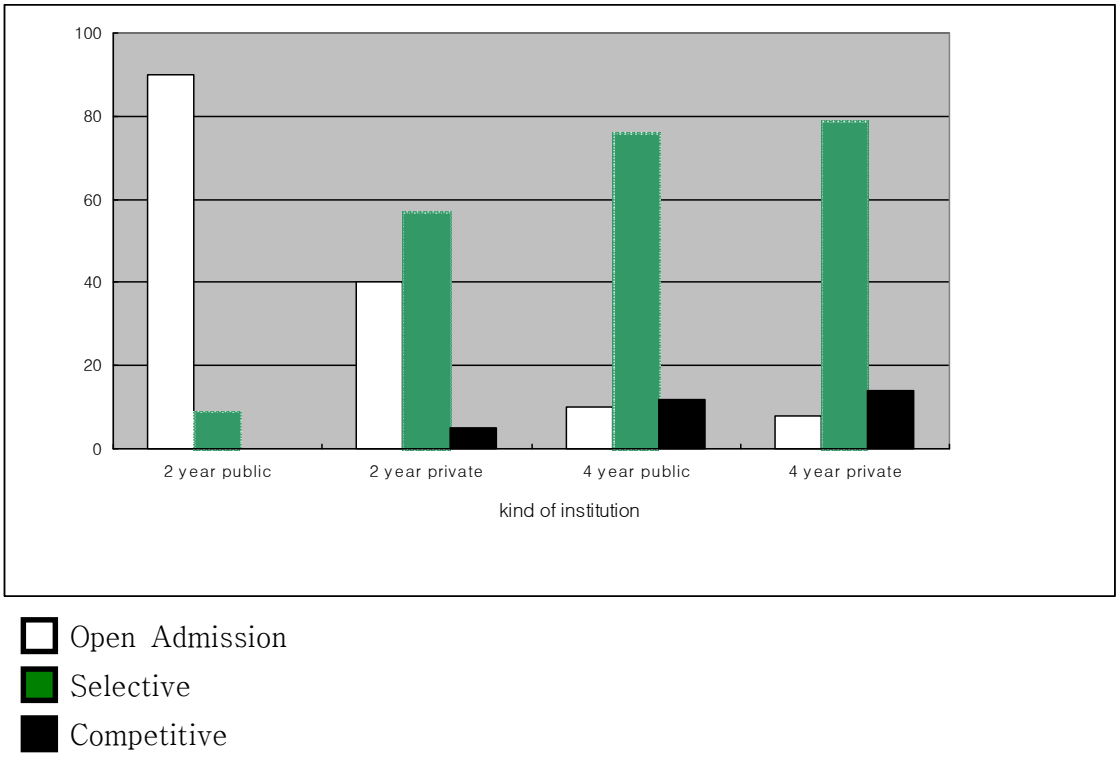
개방형 입학제도는 대학이 공적 책무를 강하게 인식하고 교육기회의 분배나 복지적 차원에서 접근할 때 사용하는 방법으로 대개 실질적인 경쟁률이 없다. 선별형은 학생을 선발하기는 하나 그 정도가 심하지 않은 주립 대학 등에서 널리 찾아볼 수 있는 방법이다. 경쟁형은 입학 경쟁이 매우 치열한 까닭에 다양한 요인을 고려하여 선발하는 방법으로 아이비리그 소속의 사립대학과 일부 주립 대학에서 채택하고 있다. 각각의 특징에 관한 보다 자세한 설명은 아래와 같은 유형 분류를 참고해볼만하다(김미란, 2010: 7-8 참조).

① 입학하고자 하는 모든 지원자들은 전통적인 학문적 자격에 대한 검토 없이 입학이 허가됨
② 고등학교 졸업이나 이와 동등한 학력을 갖춘 모든 지원자는 입학이 허가됨
③ 특정한 수준의 학업성취 수준 이상이거나, 특정 자격을 갖추거나, 고등학교 졸업 이상의 학력을 갖춘 지원자들 중에 다수가 입학이 허가됨
④ 특정한 수준의 학업성취 수준 이상이거나, 특정 자격을 갖추거나, 고등학교 졸업 이상의 학력을 갖춘 지원자들 중에 단지 소수만이 입학이 허가됨

이 분류를 기준으로 보면, ①과 ②에 해당하는 경우가 개방형 입학제도이다. ③을 채택하고 있는 경우가 선별형 입학제도이고, ④를 경쟁행 입학제도로 분류할 수 있다. <그림 1>은 브레란드(Breland)가 1995년에 조사 발표한 입학제도 유형별 분포인데, 2년제 공립대학을 제외한 대부분의 대학이 선별형 입학제도를 채택하고 있음을 알 수 있다. 이를 다시 등록자를 기준으로 분류해보면, 개방형 입시를 통해 대학에 진학하는 학생은 전체의 38%, 선별형 입시는 52%

24) 이 부분은 필자가 참여한 연구인 이윤미, 김성천, 김용일, 반상진, 『중등교육정상화를 위한 서울시립대 입학제도 개선 방안 연구』, 한국교육연구네트워크, 2012년 11월 보고서에 수록된 내용(33-42쪽)에 기초하고 있음.

이다. 반면, 경쟁형 입시를 통해 진학하는 학생은 전체의 10% 수준으로 나타난다(김미란, 2010: 8 참조).



<그림1> 대학 종류별 학생선발방법/정책[25] (이윤미 외, 2012: 34)

미국의 경우 입학조건에 따른 차이가 발생하는 것은 사실이지만, 우리 식의 대학 서열화와는 상당히 다르다. 다른 무엇보다 점수 분포에 따라 학교나 학과를 순서지우지 않고 일정한 조건에 부합하는지 아닌지를 기준으로 학생을 선발하는 계층적 평준화 정책을 취하고 있기 때문이다. 게다가 경쟁적 수준의 선별성을 취하는 공립대학에 입학하는 학생들 중 아이비리그에 진학하는 학생들보다 학교성적이나 수능점수가 높은 학생들이 많다. 공립대학이 등록금이 훨씬 저렴하면서도 사립대학에 못지않은 양질의 교육을 제공하고 있어 공립대학을 선택하는 우수 학생 자원이 많기 때문이다. 또한 학사관리가 엄격하여 졸업생 수준에서는 실력의 평준화가 나타나기도 한다(김영석, 2004: 143).

전형 방법과 요소도 다양하다. 먼저 전형방법은 크게 숫자를 이용하는 공식에 의한 전형과 총체적(holistic) 전형으로 나눌 수 있다. 공식에 의한 전형에는 고등학교 GPA나 순위 그리고 표준화된 시험성적(SAT, ACT 등)이 포함되며, 총체적 전형에는 지원서, 에세이, 추천서와 다른 정보를 포함하는 지원자의 전체 파일에 대한 검토가 이루어진다(김미란, 2010: 8). 공식에 의한 전형은 커뮤니티 칼리지나 주립대학들과 같이 지역의 모든 학생들에게 교육기회를 제공해야 하는 공적 책무를 갖는 대학들이 주로 사용하며, 앞에서 개방형 및 선택형 입시를 실시하는 학교

25) Breland et al (1995), 9-10/ 김영석(2004), 142에서 재인용

들이 여기에 속한다. 반면, 총체적 전형은 경쟁형 입시를 채택하고 있는 대학들이 주로 실시하는 방법으로 지원자의 경험, 재능, 관점, 배경 등 다양한 기준을 통해 학생을 선발한다. 양자의 특성을 정리하면, 아래의 <표 1>과 같다.

<표 1> 전형방법에 따른 주요 특징 비교

구 분	공식에 의한 전형	총체적 전형
주요 설립 유형	주립대학 > 사립대학	사립대학 > 주립대학
주요 입시 유형	개방형 또는 선별형 입시 – 주어진 조건을 충족시키는 지원자 대부분이 합격 – 경쟁률이 높지 않음	경쟁형 입시 – 주어진 조건을 충족시키는 지원자 중 소수만 합격 – 경쟁률이 높음
구성 원리	최소 기준 설정: 최소 기준 이상의 요건을 갖춘 집단	대학의 필요에 따라 구체적으로 설정: 대학자원의 활용과 대학에 대한 기여도를 바탕으로 대학이 구체적으로 설정
입시의 투명성	투명	불투명
사회적 책무성 (대학입시의 목적)	강함 (고등교육 기회분배 > 대학발전)	약함 (대학발전 > 고등교육 기회분배)

자료: 김미란(2010), 9쪽의 자료를 수정·보완함. 이윤미 외(2012: 35).

전형요소는 고교 내신 성적(GPA), 대학수학능력시험(SAT), 지원에세이, 교사 추천서, 학생의 자질 및 소양 등에 관한 학생 신상정보들로 이루어진다. 여기에 고교 교육과정의 난이도, 고교에서 이수한 대학예비과정(AP, Advanced Placement) 성적, 지원대학에 대한 관심도, 학년석차, 총 학점 평균 등도 중요한 참고사항이다(김미란, 2010: 10).

이밖에 면접(interview)을 실시하는 경우도 있다. 모든 대학이 면접을 필수로 요구하지는 않으며, 일반적으로 생각하고 있는 것과 달리 면접의 결과는 당락에 거의 영향을 미치지 않는다. 한편, 지원자가 재학하는 고등학교에서는 학교 현황(school profile)을 지원 대학에 제출해야 하는데, 이 자료에는 고등학교의 학생구성, 개설 중인 교육과정을 파악할 수 있는 정보나 통계 등이 포함된다. 그러나 표준화된 서식이 있는 것은 아니며 학교에 따라 다른 양식으로 보고한다.

다양한 전형 요소 가운데 가장 중요한 요소는 고등학교 GPA 또는 석차이다. 이는 4년제 국·공립대학과 사립대학을 대상으로 1979~2000에 걸친 네 차례의 조사에서 확인된 사실이다. 그 다음으로는 SAT, ACT와 같은 표준화 시험성적, 고등학교 교육과정 이수 내역, 대학수준 교과 이수 내역(이수 여부, 등급) 등이 주요 고려 대상이다. 이 밖에도 주(州) 거주 여부, 소수자 여부, 국적, 장애인 여부, 각 고등학교 출신자 비율, 성별, 재정적 지원 필요 등이 고려되고 있다(김미란, 2010: 10-11 참조).

입학사정 시 어떤 요인들이 중요시 여겨지는가에 있어서 학교의 크기에 큰 차이가 보이지 않고 있는 반면, 공립대학과 사립대학 간에는 중요한 차이를 보이고 있다. 공립대학이 계량화할 수 있는 객관적 요인에 따른 입학 결정에 의존하는 반면, 사립대학은 개개인의 면면이나 특성을 살피는데 더 치중한다고 볼 수 있다. 또 경쟁률이 높은 대학일수록 고교시절의 과외활동,

SATⅡ와 같은 과목별 시험, 추천서 등의 반영비율이 높다(김영석, 2004: 136).

1990년대 이후 미국에서는 개방형 입학제도가 퇴조하고, 학생 변별이 강화되는 추세를 보인다. 그 원동력은 대학입학 예비자원의 증가다. 1980년대 초반 베이비붐의 여파로 인한 학령아동의 증가와 공교육 투자에 따른 고등학교 졸업률의 증가로 인해 평등을 강조하는 추세에서 변별력을 중시하는 쪽으로의 변화가 이루어진 것이다. 또 다른 요인으로는 미국 사회의 보수화와 경쟁을 강조하는 풍조의 만연을 들 수 있다. 이런 풍조를 이끌어 가는 집단은 일부 명문 사립대인데, 사립대의 공격적 학생유치 전략으로 우수 학생을 빼앗기게 된 일부 주립대학들도 선별성 경쟁에 뛰어든 것이다(Breland et. al., 1999; 김영석 외, 2004: 141 에서 재인용).

선별성이 강조되는 현실에서 유명 사립대학 이상의 명성을 유지하면서도 공립대학의 이념에 충실한 사례 또한 눈여겨볼 필요가 있다. 캘리포니아의 대학시스템이 그것인데, 차별적 평준화 정책이라 명명할 수 있는 입학정책을 통해 공립대학이 추구하는 평등주의(egalitarianism)의 이상과 능력주의(meritocracy)를 동시에 추구하고 있다(이윤미 외, 2012: 37).

나. 주요 표준화시험 (SAT와 ACT)[26)]

미국대학들이 SAT나 ACT와 같은 표준화시험을 입학전형에 반영하는 이유는 연방체제하에서 학교들의 재정역량, 교육과정, 평가 등의 편차가 크기 때문이다. 즉 지역별 차이 및 학교 유형(사립학교, 원격교육, 홈스쿨링) 등에 차이가 많아 다른 나라들과 같은 단일한 대학입학시험체제를 유지하기 어렵기 때문이다. SAT나 ACT는 중등학교 학생부기록을 보완하기 위한 것이며, 지역적 데이터(교육과정, 학점, 석차)를 전국적 수준으로 상대화하기 위한 것이다. 대부분의 대학이 SAT나 ACT 중 하나를 선호하지 않고 둘 다 채택하는 경향 있다. 이러한 표준화시험들은 대학입시에서 대부분 전형요소의 하나로만 간주된다.

▫ SAT

SAT는 미국의 대학에서 가장 광범위하게 활용되는 표준화시험의 하나이다. SAT는 국가에서 운영하는 것이 아니라 사설 비영리 민간 기구인 College Board가 운영하며, Educational Testing Service (ETS)가 위탁 제작하여 시행되는 시험이다. 그 역사는 1926년으로 거슬러 올라간다. SAT의 원래 이름은 Scholastic Aptitude Test로 IQ 검사의 원리를 적용하는데서 시작되었으나 그 후 Scholastic Assessment Test로 개칭되고 SAT Reasoning Test 등을 거쳐 지금은 특별한 의미를 지칭하지 않는 SAT라는 이름을 유지하고 있다. 이 시험에서 평가하고자 하는 것은 대학 공부에 대한 준비도이다.

현행 SAT는 2005년에 개정된 것으로 3시간 45분 동안 실시되며 52.50불의 응시료를 낸다. SAT 점수는 800점 만점의 3개 영역(수학, 비판적 독해, 작문)으로 구성되며 총점은 600-2400점으로 분포된다. 현행 SAT는 학교 교육과정을 반영하지 않으며, 논리나 추상적 사고를 재고 있

26) http://en.wikipedia.org/wiki/SAT 및 http://en.wikipedia.org/wiki/ACT_(test) 참조하여 정리함.

지 않다. SAT는 재개정되어 2016년에 새롭게 시행될 것이라고 발표된 바 있다. 작문영역에 포함된 에세이쓰기가 선택으로 바뀌고 1600점 만점의 시험으로 되돌아갈 예정이다.

SAT는 고등학교 2-4학년 학생들이 응시하는 시험이다. SAT는 대학에서 필요로 하는 문해, 작문능력을 측정하는 시험으로 대학에서의 학업성공을 예측하려는 것이다. 이는 문제를 분석하고 해결하는 능력을 재고자 하는 것으로 대학에서 필요한 능력이라고 보고 있다. 시험은 변별력을 확보하기 위해 제한된 시간 동안 실시되기 때문에 우수하지만 시험에서의 순발력이 낮은 학생들이 평균 점수밖에 받지 못하게 한다고 지적되기도 한다. College Board에서는 고등학교 학점(GPA)과 SAT 점수를 함께 보는 것이 대학에서의 성공을 보다 잘 예측할 수 있다고 주장한다.

▫ SAT 시험의 역사

SAT는 브릭험(Carl Brigham)에 의해 개발되었는데, 사회경제적 배경이 다른 피검사자 간의 검사편향(test bias)을 제거하기 위해 고안되었다고 알려져 있다. College Board에 의해 실시된 최초의 시험은 1901년 6월 17일 973명의 학생을 대상으로 미국의 67개소와 유럽의 2개소에서 실시되었다. 당시 시험을 본 학생들은 대부분 뉴욕, 뉴저지, 펜실베니아주에서 왔고 사립학교, 아카데미 졸업자들이었다. 이들 중 60%는 컬럼비아대학에 지원했다. 시험과목은 영어, 불러, 독일어, 라틴어, 그리스어, 역사, 수학, 화학, 물리 등이었다. 당시의 시험은 선다형이 아니었고 에세이시험으로 탁월, 우수, 의심, 취약, 매우 취약으로 분류되었다(excellent, good, doubtful, poor, very poor)

최초의 SAT시험은 1926년 6월 23일에 실시되었다. 이 시험은 Scholastic Aptitude Test로 알려졌다. 앞서 언급했듯이 프린스턴대학의 심리학자 브릭험이 위원장인 위원회에 의해 개발된 이 시험은 개념정의, 수학, 분류, 인공언어, 반의어, 수계열, 유추, 논리추론, 문단독해 등으로 구성되었다. 300여개소의 시험장에서 8000여명이 응시했다. 성비는 60%가 남자였다. 응시자의 25% 가량이 예일대학과 스미스대학에 지원했다. 문항수는 315개였고 90분 동안 실시되었다.

1928년에는 언어영역이 7개로 줄었고 시험시간도 2시간으로 줄었다. 1929년에는 영역이 6개로 다소 줄어들었다. 이는 응시자들에게 시간제한을 완화시켰다. 언어영역에 초점을 두는 대신 수학은 완전히 없어졌다.

1930년에 SAT시험은 처음으로 언어와 수학 영역으로 구분되었고 이는 2004년까지 지속되었다. 1930년 언어 영역은 이전에 비해 협소한 범위를 다루었는데, 반의어, 중의개념(문장완성문항과 유사), 문단독해 등이 그것이다. 1936년 유추가 다시 추가되었다.

1936년과 1946년 사이에 80-115분간 250개의 언어영역문제를 (1/3이상이 반의어) 해결하도록 했다. 1930년 수학시험이 시작되었는데, 100개의 자유응답문항을 80분동안 풀도록 했으며 문제해결속도가 관건이었다.

1936년과 1941년 사이에는 1928-1929년 시험처럼 수학시험이 완전히 없어졌다. 1942년에 수학영역이 다시 추가되었을 때는 선다형시험으로 변경되었다.

1941년 이후 모든 SAT시험은 서로 다른 시험점수간의 비교 가능성을 높이기 위해, 해마다

응시자 수준에 의한 편차를 줄이기 위해 전 회 시험과 동등화되었다(equated). 1941년에는 언어영역, 1942년에는 수학영역이 동등화되었다. 이로써 1995년까지 응시자들은 1941년과 1942년 시험과 비교되었다.

1946년 시험에서는 언어영역의 문단독해가 읽기독해로 대체되었고, 이중개념 문항들은 문장완성형으로 대체되었다. 1946년에서 1957년까지 107에서 170개의 언어문제에 대해 시험시간이 90-100분으로 변경됐다.

1958년부터 시간제한은 더욱 안정적으로 되어 17년동안 1975년까지 90문항을 75분간 해결하도록 되었다. 1959년에는 수학영역에서 데이터충분성이 도입되었고 이후 1974년에는 양적 비교로 대체되었다. 1974년에는 언어, 수학 영역이 75분에서 60분으로 줄었고 이에 대응하여 시험구성이 바뀌었다(1960-70년대 생략. 점수 하락 추세).

1994년 초 SAT시험에서는 큰 변화들이 나타났다. 언어영역에서 반의어가 없어졌는데 이는 단순암기를 덜 유용하게 하기 위한 것이었다. 반면 제시문에 기반한 읽기자료가 30%에서 50%까지 증가했다. 이 제시문들은 대학에서 사용되는 전형적인 독해자료에 기반한 것이었다. 이는 College Board의 위원회가 발표한 1990년 보고서에 기반한 것으로 분석적 독해가 강조되었다. (에세이시험도 필수화할 계획이었으나 소수자들 및 채점 비용 등으로 인한 반대에 부딪쳐 시행되지 못했다.) 수학영역에서는 수학교사협회의 제안을 수용하여 계산기를 사용할 수 있게 했다. 1935년 이후 처음으로 선다형이 아닌 수학시험이 시작되었고 수학영역의 시험범위가 확대되었다.

2005년에는 언어, 수학에서 일부 문제유형이 폐지되고 에세이가 포함된 작문영역이 추가되었다. 수학영역은 고교 3년간의 수학을 범위로 삼도록 확대되었고, 언어영역은 비판적 읽기 영역으로 개칭되었다.

최근의 변화를 보면, 2012년부터 응시자들이 등록시 최근 사진을 제출하도록 되었고, 2014년 3월 5일에는 SAT를 고등학교 교육내용과 연결시키겠다고 발표했다. 새로운 시험은 2016년 봄부터 반영될 예정이다. 1600점 만점 체제로 다시 돌아가며, 에세이를 선택화하고, 틀린 답안에 대한 감점을 없앤다. 또한 College Board는 값비싼 준비프로그램들에 맞서기 위해 Khan Academy와 함께 무료온라인 예시문제와 비디오를 제공하기로 했다.

▫ ACT

ACT는 American College Testing의 준말이다. 고등학교 학업성취를 확인하는 표준화시험으로 SAT처럼 대학입시에서 활용되고 있다. ACT가 처음 시행된 것은 1959년 11월이었고 개발자는 린퀴스트(Everett Frankling Lindquist)였다. 린퀴스트는 SAT시험의 유명한 반대자였다. ACT는 영어, 수학, 사회과학, 자연과학의 4영역을 중심으로 실시되어오다가 1989년 사회과학을 독해영역으로 전환하고 사회과학은 하위영역으로 조정되었고 자연과학도 문제해결력을 강조하는 과학추론시험으로 개명되었다.

2005년 2월 ACT에 선택과목으로 작문시험이 부가되었는데 이는 같은 해 SAT시험에 작문이 부가된 것과 시기를 같이 한다. 2015년 봄부터 ACT는 컴퓨터기반시험으로 전환될 예정이다.

이때 주관식 문항(Constructed Response Questions)도 추가될 예정이지만 전체적인 시험내용, 점수구성, 선다형형식 등은 변경되지 않는다. 컴퓨터기반으로 준비가 안 된 시험장에서는 여전히 필기시험의 형식으로 실시될 예정이다.

ACT 응시자는 계속 증가해왔는데 2011년에는 처음으로 SAT 응시자를 넘어섰다. 모든 대학에서 ACT점수를 전형에서 채택하지만 상대적 반영비율(학교석차, GPA, 교과외활동 등)은 편차가 있다. 점수구성은 각 영역별 1-36점이며, 총점은 각 영역 점수의 평균점수로 계산한다.

SAT와 마찬가지로 ACT는 대학학업준비도를 측정하고자 한다. ACT의 각 영역점수는 대학에서 요구하는 영어, 수학, 사회과학, 인문학, 생물학 등의 학습과 대응되도록 하고 있다.

ACT는 개발과정에서 미국의 중등학교(중학교, 고등학교)의 교수목표를 반영하며 7-12학년 교과서를 검토한다. 또한 교육자들을 대상으로 고교졸업후 학습의 성공에 필요한 기술에 대한 설문조사를 실시한다. ACT는 대학신입생 대학에서의 학점(GPA) 예측을 위한 타당도제고 관련 연구결과보고서를 발간하기도 한다.

ACT는 대학진학학생이 아닌 경우에도 학교평가의 일환이나 졸업자격의 대용으로 실시되기도 한다. 콜로라도나 일리노이는 ACT 응시를 2001년 이후 의무화하고 있다. 미시간도 2007년 이후 ACT를 필수화하고 있으며, 켄터키는 테네시의 경우 고교 졸업 직전 학년(3/4)에게 전체적으로 응시하도록 하고 있으며 와이오밍도 ACT 혹은 ACT WorkKeys에 응시하도록 하고 있다.

2. 책무성기제로서의 표준화시험 - NCLB를 중심으로[27)]

The No Child Left Behind(NCLB)는 연방차원의 저소득층 지원 프로그램인 TITLE 1을 포함한 초중등교육법(ESEA)를 개정한 것으로 표준기반교육개혁을 지원하기 위한 것이다. 높은 표준을 설정하고 측정가능한 목표를 정하는 것이 교육에서의 결과를 개선할 것이라는 기대에서 시행되고 있다.

이 법안은 학생의 기본 기능(basic skill)에 대한 평가를 기본으로 한다. 연방의 학교지원금을 받기 위해서는 각 주가 학교들에 대해 정해진 학년에 평가를 실시해야 한다. 법안 자체가 국가수준의 성취 표준을 제시하고 있지는 않다. 각 주가 자체적으로 표준을 설정해야 한다. NCLB는 연례평가, 연례학업개선, 보고서, 교사자격, 재원변화 등을 통해 공교육에서 연방의 역할을 확대시켰다.

NCLB는 미국 의회(107차)에서 양당의 지지를 받으며 통과되어 2002년 1월 8일부터 시행되었다. 2001년 1월 23일 조지 W. 부시에 의해 제안되었고, 미국 하원에서 2001년 5월 23일, 상원에서 6월 14일에 통과되었다.

□ 법안의 내용

NCLB는 연방의 지원을 받는 모든 공립학교가 매해 모든 학생에게 주(State) 단위 표준화시험

27) http://en.wikipedia.org/wiki/No_Child_Left_Behind_Act

을 실시하도록 하고 있다. 저소득층학생에 대한 특별 연방지원(Title 1)을 지원받는 학교들은 시험점수에서 해마다 적정 개선정도를 보고하도록 되어 있다(Adequate Yearly Progress). 학교의 성취 점수가 낮을 경우 학교교육을 개선하기 위한 절차를 취하도록 되어 있다.

연속 2년 시행에서 AYP를 충족하지 못한 학교는 개선요망 학교로 공식 분류되어 학교에서 문제가 되는 교과에 대해 2년간 개선프로그램을 실시해야 한다. 학생들은 같은 학군내의 다른 학교에 전학을 할 수 있다.

3년간 AYP가 충족되지 못하면 학교는 부진학생에게 개인지도를 제공해야 하며 보완교육을 실시해야 한다. 4년간 미충족된 경우 학교는 교정대상으로 분류되며, 교원의 교체, 새로운 교육과정의 도입, 수업시간 증가 등의 조치를 취해야 한다. 5년간 미충족인 경우 학교 전체에 대한 재구조화계획 수립이 요구된다. 6년간 실패할 경우 학교재구조화 계획은 실행에 옮겨져야 한다. 공통적 조치에는 학교폐쇄, 차터스쿨로의 전환, 사설업체에 위탁운영, 주교육청의 직접 운영 등이 포함된다.

각 주에서 AYP를 정할 때는 다음의 사항을 필수적으로 반영해야 한다.

1. 각 주는 AYP에 대한 측정가능한 목표를 정해야 한다. 이는 모든 학생 및 특별 집단(경제취약층, 장애학생, 영어취약층)에 대해 설정되어야 한다.
2. 이 목표는 12년(2013-14 학년도 말)에 모든 학생을 능숙(proficient) 수준에 도달하도록 해야 한다.
3. AYP는 기본적으로 각 주의 평가에 기반해야 하며 추가적 지표를 반드시 하나 포함해야 한다.
4. AYP목표는 학교수준에서 평가되어야 한다. AYP목표를 2년간 충족하지 못하는 경우 개선대상으로 분류된다.
5. 학교 AYP는 학생별로 위의 범주에 근거하여 보고됨으로써 각 집단의 학생이 AYP 목표에 달성했는지를 확인할 수 있다.
6. 모든 그룹에서 적어도 95%의 학생이 주별 평가에 참여해야 한다.
7. 각 주는 AYP보고를 위해 3년간의 자료를 축적해야 한다.

이 법안은 모든 학생들에게 '질 높은' 교사를 제공할 것을 요구한다. 각 주는 질높은 교사가 무엇을 의미하는가에 대한 독자적 표준을 설정해야 한다. 유사하게 학생들에게도 '높고 도전적인 표준'을 설정하도록 요구하고 있다. 각 주별로 '높고 도전적인 표준'을 정할 수 있으며, 교육과정은 주 내부의 모든 학생에게 동일하게 적용되어야 한다.

□ 교육과정과 표준에 대한 효과

지방정부가 교육에 실패함에 따라 연방정부가 개입해서 질 낮은 교육에 대해 교정해야 한다는 주장들이 있다. NCLB의 결과로 많은 주들에서 NCLB의 목표에 도달했다는 보고들이 있다(위스컨신과 워싱턴 DC가 98%로 50개 주중 1위).

교육의 질 제고는 다음과 같은 접근에 기초하고 있다: 1) 학교들로 하여금 질 제고를 하도록 함으로써 도달, 2) 학교들이 교실, 부모참여프로그램, 비진학학생의 직업개발활동 등에 대한 과

학적 연구를 행하도록 함으로써 교수의 질을 높이도록 함, 3) Early Reading First를 시행하여 조기 문해를 도움, 4) 읽기, 언어기능, 수학, 과학 등을 핵심교과로 정함.

NCLB의 초점은 읽기, 쓰기, 수학 등 경제 발전과 관련된 기능에 치우쳐있다. 2000년대말 재정위기와 결합하여 많은 학교들이 NCLB의 책무성 표준과 관련없는 많은 교과들에서 수업을 폐지하고 재정을 삭감했다. 2007년 이후 약 71%의 학교가 역사, 예술, 언어, 음악 등의 수업시간을 줄이고 수학, 영어 시간을 늘렸다. 일부 학교에서는 수업시간은 그대로 두되 기본 교과에서 성취가 낮은 학생들은 예술, 체육 등의 선택과목 대신 읽기, 수학 보충학습을 하도록 보내졌다. 기본 교과에 대한 지나친 강조는 학생의 경험을 협소화하고 과학에 대한 호기심과 흥미를 떨어뜨린다는 비판을 받아왔다.

NCLB는 2014년까지 읽기와 수학에서 모든 학생이 동일한 표준에 도달해야 한다고 요구한다. 비판자들은 100% 도달은 어렵다고 말한다. 실제로도 NCLB에서는 95%를 요구하고 있다. 5%에 대해서는 보고를 하지 않아도 되며, 1%까지는 대안적 평가를 실시할 수 있도록 하고 있다. 일부 주에서는 포트폴리오식 평가를 포함한 대안적 평가를 소수에 대해 적용하기도 한다. 장애학생들에게 통합적 평가가 타당한지, 동일한 평가 적용이 차별적인지에 대해서는 찬반 논란이 있다.

참고문헌

김미란 (2010). 『외국의 대입제도』. 한국교육개발원 position paper. OR 2010-01-9

김영석 (2004). 「미국의 대입제도: 공립대학의 차별적 평준화정책을 중심으로」. 경상대학교 사회과학연구원 편, 『대학서열체제연구: 진단과 대안』, 경상대학교.

이양락외 (2009). 『일본대학입시센터시험 문항분석』. 한국교육과정평가원, 연구자료 ORM 2009-46

이윤미, 김성천, 김용일, 반상진 (2012). 『중등교육정상화를 위한 서울시립대 입학제도 개선 방안 연구』, 한국교육연구네트워크

이충우 (1980). 『경성제국대학』, 다락원

정영근 (2010). 「독일학교의 내신평가」. 『일제고사 그리고 대학 진학, 우리교육 평가제도를 평가한다』. 제5회 원탁토론 학술심포지엄(서울경기강원교육청 위탁 전문과정 연수자료), 2010년 10월 23일, 91-100

한국교육과정평가원 (2001). 『일본, 중국, 대만 대학입시제도연구』. 한국교육과정평가원, 연구보고 RRE 2001-12

일본 문부과학성 국립교육정책연구소 (2014). 『平成２６年度 全国学力・学習状況調査 調査結果のポイント』. 文部科学省 国立教育政策研究所, 2014.8.

竹内洋 (1995). 『日本のメリトクラシー』. 東京大學出版會

Atia, Samir El (2008). From Napoleon to Sarkozy: two hundred years of the Baccalauréat exam. Language Assessment Quarterly, 5(2), 142-153

Bily, Cynthia A. (2011). Standardized testing, Farmington Hills, MI: Greenhaven Press

Bower, Joe & Thomas, P. L. (2013). De-testing and de-grading schools: authentic alternatives to accountability and standardization, New York: Peter Lang

Cremin, Lawrence A. (1964). The transformation of the school: Progressivism in American education, 1876 - 1957, New York: Vintage Books

Dolezaler, Holly (2009). Standardized testing in schools, Edina, MN: Abdo publishing company

EURYDICE (2009). National testing of pupils in Europe: objectives, organisation and use of results, Education, Audiovisual and cultural executive agency, European Commission

Farley, Todd (2009). Making the grades: my misadventures in the standardized testing industry, San Fransico: Berrett- Koehler

Gardner, H. (1992). Assessment in context: the alternative to standardized testing. in B. R. Gifford et al. (eds.) Changing assessments. New York: Kluwer Academic Publishers

Garrison, Mark J. (2009). A measure of failure: the political origins of standardized testing, Buffalo: SUNY

Gorlewski, Julie & Brad, Porfilio (2013). Left behind in the race to the top: realities of school reform, Charlotte, NC: Information Age Publishing Inc.

Gould, Stephen Jay (1996). The mismeasure of man, New York: WW Norton & Company

Guba, E. G. & Lincoln, Y. S. (1989). Fourth generation evaluation, New York: SAGE publications

Reese, William J. (2013). Testing wars in the public schools: a forgotten history, Cambridge: Harvard University Press

Huber, S. G. & Gördel, B. (2006). Quality assurance in the German school system. European Educational Research Journal, 5(3&4), 196-210

Karabel, J. (2005). The chosen: the hidden history of admission and exclusion at Harvard, Yale and Princeton, Boston: Mariner Books

Lemann, Nicholas (2000). The big test: the secret history of the American meritocracy, New York: Farrar, Straus and Giroux

Sacks, Peter (2000). Standardized minds: the high price of America's testing culture and what we can do to change it, New York: Da Capo Press

Sedlacek, William E.(2004). Beyond the big test: non-cognitive assessment in higher education, San Francisco: Jossey-Bass

Shepard, Lorie A. (2000). The role of assessment in a learning culture. Educational Researcher. 29(7), 4-14

Spring, Joel (2014). Political agendas for education: from race to the top to saving the planet, New York: Routledge, 2014 (5th edition)

Haney, Walt (1978). Making sense of the competency testing movement, National Consortium on testing, staff circular no.2 (Research report)

Harris, Douglas N.(2012). Value-added measures in education, Cambridge: Harvard University press

Harris, P., Smith, B.M., Harris, J. (2011). The myths of standardized tests: why they don't tell you what you think they do, New York: Rowman & Littlefield Publishers, Inc.

Heafford, Michael (1995). The early history of the abitur as an administrative device, German History, 13(3), 285-304

Hoffmann, Banesh (1962). The tyranny of testing, New York: Dover Publications, Inc.

Kohn, Alfie (2000). The case against standardized testing: raising scores, ruining the schools, Portsmouth, NH: Heinemann

Koretz, Daniel (2008). Measuring up: what educational testing really tells us. Harvard University Press

Lysne, A. (2006). Assessment theory and practice of students' outcomes in the Nordic countries. Scandinavian Journal of Educational Research, 50(3), 327-359

Lundahl, C. & Waldow, F. (2009). Standardisation and 'quick languages': the shape-shifting of standardised measurement of pupil achievement in Sweden and Germany. Comparative Education, 45(3), 365-385

Martin, W. A. P. (1870). Competitive examinations in China. The North American Review, 111(228), 62-77

McGuinn, Patrick J. (2006). No Child Left Behind and the transformation of federal education policy 1965-2005, Lawrence: University of Kansas

Mertens, Donna M. & Wilson, Amy T. (2012). Program evaluation theory and practice, New York: Guilford Press

Phelps, Richard P. (ed.)(2006). Defending standardized testing, New York: Psychology Press

Teng, S. (1943). Chinese influence on the Western examination system (I. Introduction). Harvard Journal of Asiatic Studies, 7(4), 267-312

Wolf, D., Bixby, J., Glenn, J. III, Gardner, H. (1991). To use their minds well: investigating new forms of student assessment. Review of Research in Education, 17, 31-74

Wolf, Richard M. (1990). Evaluation in education: foundations of competency assessment and program review, New York: Praeger

Whitford, B. L. & Jones, K. (2000). Kentucky lesson: how high stakes school accountability undermines a performance-based curriculum vision. In B,L. Whitford & K. Jones (eds.), Accountability, assessment and teacher commitment: lessons from Kentucky's reform efforts. Albany, NY: State University of New York Press.

Young, M. (2008). The rise of the meritocracy, London: Transaction Publishers

Zenderland, Leila (2001). Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing, Cambridge: Cambridge University Press

Zeng, K. (1999). Dragon gate: competitive examinations and their consequences, London: Cassell

Zessoules, R. & Gardner, H. (1991). Authentic assessment: beyond the buzzword and into the classroom. in V. Perrone (ed.). Expanding student assessment. Association for supervision and curriculum development.