DOI 10.17939/hushss2016..53.002

교육의 효율성과 타당성의 대립: 한국과 서양의 교육문화 비교

류성창*

<국문요약>

한국의 학생들은 거의 모든 학업의 동기가 시험과 그 결과의 활용에 초점을 맞추고 있는 반면 교육학의 이론가들과 정책 입안가들은 교육적으로 옳은 방향에 초점을 맞추고 있어, 효율성을 중시하는 현실과 교육의 타당함을 중시하는 이론의 괴리가 인지된다. 본고는 미국을 비롯한 해외의 교육평가 사례들을 검토하고 전반적인 세계적 흐름을 고찰하며, 한국의 교육평가 개선 사례들을 대조적으로 분석하여, 교육에 있어서 효율성과 타당성이 결국 상호 대립적인 개념이 아니라 현실에 있어서 동시에 추구될 수 있음을 논증한다. 특히 효율성을 중시하는 한국의 교육현실에서 교육적 타당성이 보다 강조되어야 한다.

[주제어] *효율성, 타당성, 변별력, 교육평가, 역류효과, 역행설계, 수행평가*

* 이름: 류성창 (E-mail : ryoo@kookmin.ac.kr)
학위취득대학: Harvard University 교육학박사
현직: 국민대학교 교육학과 조교수

논문접수일 : 2016년 10월 03일, 논문수정일 : 2016년 11월 14일, 게재확정일 : 2016년 11월 20일

www.kci.go.kr

Efficiency versus Validity in Education: International and Domestic Trends

Ryoo, Josh Sung-chang

<Abstract>

In most cases, Korean students find their actual motivation for learning in obtaining good results in various evaluations, whereas theorists and scholars in their academic works tend to value educationally desirable directions. Thus the concepts of efficiency and validity in educational practices seem to be in a contradictory, indeed irreconcilable relationship. This article examines several international and domestic cases in which educational evaluation has been revised in pursuit of validity, and discusses important implications for the betterment of Korean educational culture. In particular it requests that participants more highly value validity in practical realms.

[Keywords] *efficiency, validity, discriminatory power, washback effect, backward design, performance assessment.*

www.kci.go.kr

I. 서론

한국의 교육 현실에서 교사와 학생들은 시험 성적을 잘 받기 위한, 즉 좋은 결과를 얻기에 최적화된 학습을 추구하는가, 아니면 마땅히 배워야 할 내용을 마땅히 수행해야 하는 학습의 방법으로 배우고 있는가?

최소의 비용으로 최대의 효과를 얻는 효율성과 마땅히 배워야 하는 것을 가르치고 배우는가의 타당성의 문제는 한국교육의 현실과 한계점을 지적함에 있어서 늘 경쟁하는 두 핵심 개념으로 인식되어 왔다. 그리고 한국의 교육문화는 효율성을 추구하는 현장과 타당성을 강조하는 학계가 서로 대립하는 구도에 익숙해져 있다. 심지어 타당성을 강조하는 이론가들의 주장은 마치 현실을 모르고 탁상공론만을 일삼는 사람들인 양 생각하는 경우가 있기도 하고, 아무리 학계나 정부에서 교육적으로 옳은 방향을 가리켜도 현실은 최소 비용과 최대 효과를 얻는데에만 요지부동으로 관심을 쏟는 경우도 많다.

교육에 있어서 효율성과 타당성의 문제는 이와 같이 화해할 수 없고 동시에 반영할 수 없는 서로 모순된 개념인가? 아무리 이론과 정책이 타당성을 주장하여도 효율성과 결과만을 추구하는 한국교육의 문화는 이대로 괜찮은 것인가? 효율성과 타당성의 이분법적인 이해는 과연 옳은 것인가? 효율성과 타당성을 중요시하는 교육관계자들이 서로 대화할 수 없는 상황은 극복할 수 없는 것인가?

본고는 이와 같은 문제의식 하에 효율성과 타당성이 동시에 추구될 수 있는 해외의 사례들을 검토하여 어떻게 한국의 교육문화가 발전할 수 있는지를 고민하고자 한다.

본고에서는 미국의 SAT 시험과 TOEFL 시험의 변천, 다국적 프로젝트인 ATC21S 그리고 한국의 교육평가 변혁의 과정과 실제 수능문항 등에 대한 국내외의 관련 흐름을 비교교육학적으로 문헌분석을 통하여 고찰하고, 궁극적으로 선진적인 교육문화의 형성을 위해서는 효율성과 타당성이 복합적으로 추진되는 교육의 현실이 자리 잡아야 함을 결론적으로 주장하려 한다. 이에 따라 본고의 연구문제는 다음과 같다:

1. 교육평가 문항에 있어서 효율성과 타당성은 대립적인 개념이므로 취사선택할 수밖에 없는 개념인가, 아니면 동시에 추구될 수 있는 개념인가?

2. 교육평가의 효율성과 타당성 관련 국내·외의 동향은 어떠하며 한국교육에 주는 이론적 함의점은 무엇인가?

II. 국내 · 외 교육평가 개선의 동향

2014년 3월 미국의 대입 시험에 해당하는 SAT(Scholastic Aptitude Test)를 주관하는 College Board사는 SAT의 문항 구성과 성격을 수정하겠다는 발표를 하였다. College Board사의 회장인 데이비드 콜맨(David Coleman)은 2,400점 만점에서 1,600점 만점으로, 그리고 에세이(논술) 시험 점수를 따로 산출하는 방식의 새로운 점수를 기반으로 하여 2016년부터 적용되는 SAT 개선안을 발표하였다(CNN 2014. 3.5). 이러한 SAT의 개정발표에는 콜맨(Coleman)회장에 따르면 크게 세 가지의 이유가 있다. 첫째, 기존의 SAT 문항들이

www.kci.go.kr

고등학교 학생들의 학교에서 배우는 내용과 거리가 있다는 점, 둘째, 고등학교의 학습내용과 거리가 있는 기존의 시험 내용이 대학에서 수학할 수 있는 능력과도 거리가 있으므로 개선이 필요하다는 점, 그리고 마지막으로 시험 문제를 풀기 위해 필요한 '문제풀이 기술(test-taking skills)'들이 학생들에게 부당한 부담을 준다는 점이 그 이유이다(Ibid.). 대학의 수준이 나뉘어 있어 어느 정도의 학벌 문화가 형성되어 있는 미국과 같은 사회에서는, 한국과 같이 심각한 수준은 아니지만, 대입 시험을 잘 치르고자 하는 욕심 있는 학생들은 입시시험을 대비하여 대입준비용 교재를 공부하는 경우가 적지 않다.

그리고 이러한 소위 문제집들이 설명하는 문제풀이 기술들은 한국 학생들이 대입 시험을 치르기 위해서 배우는 기술들과 크게 다르지 않다. 예를 들어, 가장 정답에 가깝지 않은 것들을 하나씩 지워가는 기술인 '제거법(the elimination method)'은 이들 교재들이 강조하는 대표적인 문제풀이 기술에 해당한다. 새로이 개정될 SAT에서는 이와 같이 문제풀이 테크닉에 집중하는 학생들의 시험 준비 행태를 바꿀 수 있도록, 증거기반의 독해와 작문을 요구하는, 보다 고차원적인 사고가 가능한 문항들로 대입 시험을 구성한다는 계획이다.

콜맨(Coleman) 회장이 언급한 세 가지 문제점은 사실 하나의 문제로 귀결된다고도 볼 수 있다. SAT는 시험의 본래 목적에 맞게 '미국의 대학에서 공부할 수 있는 능력'을 측정할 수 있는 시험이어야 하며, 따라서 본 시험을 준비하는 학생들은 고등학교에서 배우는 것들과 대학 공부를 준비하기 위해서 배워야 하는 것들을 학습함으로써 본 시험에도 자연스럽게 준비가 되어야 하는데, 실제로 SAT가 운영되는 것을 보면 학생들은 시험문제 풀이 기술에 더 많이 집중하고 있고 마땅히 배워야 할 것들에는 신경을 덜 쓰게 되더라는 것이다.

SAT가 이와 같은 개선 노력을 보이는 것은 비단 2014년 한 번의 일이 아니다. 2005년 College Board는 원래 1,600점 만점이었던 SAT 시험을 에세이 점수 800점을 포함하는 2,400점으로 개정하고, 그리고 독해 문항들이 기존에 지나치게 어려운 단어를 물어보던 것을 보다 대학 공부에서 활용이 많이 되는 단어 위주로 전환되는 내용을 골자로 하는 개정안을 발표한 바 있기도 하다.

이와 같이 SAT 시험이 끊임없이 개선 방안을 발표하는 이면에는 사실 사설 기관에 의해서 운영되는 SAT와 ACT(American College Testing)라는 두 가지의 대입 표준화 시험(standardized tests)의 경쟁구도를 무시할 수 없는데, 수험생들에 의해서 선택되어야 수익을 더 많이 올릴 수 있는 시험 시장에 있어서 보다 매력적인 시험을 구성하려는 동기가 많이 작용하고 있기도 하다. 에세이 시험이 없고, 본래 SAT보다 실생활 단어 위주의 문항으로 구성되어 있으며, 과학 교과가 포함되어 있는 ACT는 1959년에 처음 실시되어, 급기야는 1900년대 초부터 실시되어 온 SAT의 경쟁자로 자리잡는데 성공하게 된다.

전통적으로는 중서부(the Mid-west) 지역의 학생들이 ACT를 많이 치고 동부와 서부의 학생들은 SAT 시험을 더 많이 선택하는 것으로 알려져 있던 상황에서, 보

www.kci.go.kr

다 고등학교 교육과정에 맞춰져 있고 실제 문제해결력에 더 초점이 맞추어져 있다는 평을 얻기 시작한 ACT가 동서부의 수험생들에게도 인기를 얻기 시작하면서 (The New York Times 2007.11.4), SAT를 주관하는 College Board는 회장과 운영진이 교체되고, 수년이 멀다하며 개선안을 발표하는 등 대입 시험 시장에서 경쟁력을 높이기 위해 많은 노력을 하고 있다.

여기에서 특이할 만한 사항은, 이러한 대입 시험의 경쟁구도가 있다 하더라도, 보다 올바른 교육과 대입 문화를 만들어 가기 위한 미국 교육계의 다양한 노력이 이러한 변화에 큰 영향을 미치고 있다.

예를 들면, 미국 대입상담교사들의 전국연합체인'대입진학상담연합회(National Association for College Admission Counselling)'는 '2013년 대입 현황 보고서(2013 State of College Admission Report)'에서 미국의 대입 표준화 시험을 강력히 비판한 바 있는데, 실제 대입 사정의 근거는 학생들의 내신 성적과 고등학교 수업 수강의 열의[1]가 표준화 시험 점수나 고등학교의 평판(고교 등급)보다 더 큰 영향을 미쳤다는 분석결과를 인용하며, 대입 표준화 시험들이 보다 정확한 대입 수학 능력을 측정해야 함을 역설한 바 있다(NACAC 2013).

즉, SAT나 ACT와 같은 미국의 대입 시험들이 그 문항의 성격이나 내용을 개선하고 전체 점수의 산출 방법을 끊임없이 고민하고 있는 이유는 일차적으로는 경쟁사와의 시장경쟁에서 우위를 점하고자 하는 동기가 있겠으나, 그러한 경쟁구도 하에서도 궁극적으로 교육적으로 옳은 방향으로 선도적인 개선을 시도하는 등 무언가 보다 올바른 종류의 경쟁이 이루어지고 있는 것을 그 특징으로 볼 수 있다.

이러한 표준화 시험의 개선이 비단 시장경쟁에서의 영업 우위만을 위해서 이루어지는 것은 아니라고 볼 수 또 하나의 근거는, 이와 비슷한 종류의 시험 개선 노력이 시장경쟁이 약한 맥락에서도 이루어져 왔던 TOEFL(Test of English as a Foreign Language)의 개선 과정에서 발견할 수 있다. 시험을 주관하는 미국의 또 하나의 사설 기관인 ETS(Educational Testing Service)는 영어가 모국어가 아닌 학생을 대상으로 미국 대학에서 공부할 수 있는 영어 실력을 측정하기 위한 시험인 TOEFL을 1964년부터 시행해 왔다. TOEFL은 읽기, 듣기, 말하기, 쓰기의 네 가지 영역으로 구성되어 있으며, 수험자가 문항에 반응하는 정도에 따라 다음 문항이 결정되고 말하기의 녹음파일과 쓰기의 기록이 인터넷을 통하여 ETS로 바로 직접 전송되는 인터넷 기반 시험, 즉 IBT(Internet-Based Test)로 치러지고 있다.

본 시험은 잘 알려진 대로 과거 종이 기반 시험인 PBT(Paper-Based Test)였던 것이 2000년대 초반을 기점으로 컴퓨터 기반 시험인 CBT(Computer-based Test)로 변모하였으며, 인터넷 기반 시험 관리를 강화한 현재의 IBT로 발전하였다. 주목할 만한 사항은 예전의 TOEFL은 단어의

1) 미국의 고등학교 학생들은 한국의 대학생들이 학기 초에 수강신청을 하듯이 자신이 수강하고자 하는 수업의 계획표를 짤 수 있고, 따라서 최소한의 수업만 이수하는 학생과 학업 열의를 가지고 심화과목이나 다양한 과목을 수강하는 학생의 차이가 성적표를 통해서 드러난다.

www.kci.go.kr

의미만을 묻는 문항들이나 틀린 문법을 찾는 영역이 있었고, 쓰기 영역이나 말하기 영역은 상상할 수도 없었으며, 심지어 아주 오래 전의 TOEFL에는 듣기 영역조차 없었다는 사실이다. 그러나 단어의 단순 암기 여부를 묻던 문항에서 현재는 단어의 뜻을 독해 지문 상의 맥락에서 유추할 수 있도록 하는 문항으로, 없었던 쓰기나 말하기 영역이 추가된 방식으로, 그리고 읽기와 듣기 등 복수의 영역이 융합된(integrated) 형태의 복합문항 등으로 많은 변경이 이루어진 것이다.

가장 주목할 만한 변화는 말하기 영역에 있는데, 2000년대 전후에 TOEFL을 준비했던 사람이라면, 6가지나 되는 문항들에 대해서 말해야 하며 독해와 듣기가 안되면 말하기도 할 수 없는 융합형 말하기 문항들이 포함되어 있는 현재의 TOEFL을 접하면 적지 않게 놀라게 될 것이다. 외국인 학생의 미국 대학입학 문제를 다루기 위해서 1960년대 초에 구성된 국가 협의체인 '외국어로서의 영어교육위원회(National Council on the Testing of English as a Foreign Language)'에서 제안하여 ETS의 실행으로 2016년 현재까지 TOEFL은 시행되어왔다.

TOEFL은 영국에서 개발된 같은 목적의 시험인 IELTS(International English Language Testing System) 점수를 인정하는 극소수의 미국 대학을 제외하면, 현재까지도 거의 대다수의 미국 대학에서 외국인 유학생을 대상으로 요구하는 영어시험으로 활용되고 있다. 따라서 미국 내에서 크게 경쟁할 만한 대체시험이 없었던 점을 감안하면, TOEFL의 개선 과정에 있어서는 SAT와 ACT의 시장경쟁에서의 승리와 같은 동기는 거의 없다고 볼 수 있다. 즉, TOEFL의 개선은 시장경쟁과 같은 교육의 의미와 무관한 사업적 동기보다는, ETS(2011, 2)에서 밝히고 있는 바와 같이, 영어가 모국어가 아닌 학생이 개발해야 하는 언어기술을 올바르게 측정하고 있느냐의 문제, 즉 타당성(validity)의 문제가 보다 근본적으로 작용했다고 보는 것이 옳을 것이다.

고등학교교육이나 영어교육보다 다소 폭이 넓은 맥락에서 이와 비슷한 평가 개선이 시도된 최근의 사례가 있다. 21세기에 들어가면서 앞으로의 경제 인력은 정보의 축적이나 저장에서 벗어나, 다양한 전문성을 가진 타인과의 협력이나 업무진행에 있어 발견되는 갈등을 해소하고 목표로 한 바를 이루어 낼 수 있는 대인관계능력 및 의사소통능력 등이 필요하게 되는데 반해서, 기존의 학교교육은 여전히 지식 축적 중심의 교육이 이루어지고 있어, 21세기가 필요로 하는 경제 인력을 제대로 길러낼 수 없다는 자각이 수년전부터 제기되어 왔다.

이를 개선하기 위해 인텔, 마이크로소프트, 시스코와 같은 국제 대기업들이 주도권을 가지고, 평가를 개선하여 학교교육의 현실을 바꾸어보려는 시도를 한 사례가 있다. 국제 대기업들의 연구자금을 바탕으로 미국, 호주, 캐나다 등의 교육학자들이 참여한 거대 교육평가도구 개발 프로젝트인 ATC21S(Assessment and Teaching for 21st Century Skills)이 바로 그것이다(Griffin, McGaw & Care 2011).

호주 멜버른 대학(University of Melbourne)에 연구센터를 두고 2주기 연구가 진행 중인 있는 본 프로젝트는 미국, 핀란드, 싱가포르, 호주, 네덜란드 및 코스타리

www.kci.go.kr

카에서 시범 운영 중에 있기도 하다(ATC21S 2001).[2)]

위의 사례에서 볼 수 있는 바와 같은 평가시험의 개선을 통한 학습의 정상화는 한국의 경우에도 물론 오랫동안 시도되었다고 볼 수 있다. 이런 시도는 1955년 서울시 고등학교장들의 대입 관련 좌담회에서 고등학교 내신 성적을 대입 전형의 일부로 사용할 것을 논의하였던 것에서부터 출발한다. 다음으로 실제로 내신이 대입 전형의 일부로 공식 도입되었던 1980년 7·30 교육개혁(정광희 외 2010)과 이전의 대입예비고사를 개선하고자 1982년부터 1993년까지 실시된 학력고사제도가 이어졌다.

그러나 암기력과 문제풀이 기술로만 문제를 풀 수 있었던 학력고사를 극복하여 고등수학능력을 측정하기 위하여 1994년부터 대학수학능력시험이 실시되었다. 다음으로 대입 전형 요소의 다양화와 단순 학업능력 뿐 아니라 학생들의 소질과 적성을 대입에 반영하기 위하여 2008학년부터 도입된 입학사정관제 등은 한국의 교육계가 평가의 개혁을 통하여 학습의 변화를 꾀해온 발자취로 볼 수 있다. 대입이라는 한국 교육체제의 가장 큰 시험 관문을 개선하여 초·중·고 학교현장의 교육을 정상화시키고, 학생들이 대입준비 과정을 통해 대학학업이 실질적으로 필요로 하는 능력을 성장시키는 것이 가능하게 하기 위한 일련의 방편들이었다고 볼 수 있는 것이다(류성창 외 2012, 20-25).

특히 최근 주목할 만한 흐름은, 시·도 교육청별로 다르겠으나, 내신 평가의 30~40%에 해당하는 부분을 학생의 실제 학업 행위 과정과 결과를 질적으로(qualitatively) 관찰하여 평가하는 수행평가(Performance Assessment; 허경철 외 1999, 20)가 도입되어 정착단계에 이르고 있으며, 수행평가를 그 취지에 맞게 실행하는 것 또한 중요하다는 인식이 확산되고 있기도 하다(서지영 외 2008; Lee, Ryoo & Lee 2014).

이와 같이 평가가 그 이전에 일어나는 실제 학습에 영향을 미치는 현상을 '역류효과(The Washback Effect)'라고 한다(Hughes 1989).[3)] 역류효과는 마치 강물이 상류에서 하류로 흘러야 하는데 바닷물이 밀려들어 하류에서 상류로 거꾸로 흘러가는 현상을 빗대어 표현한 것으로, 교육현실에 있어서 평가의 내용과 방법에 따라 그 평가를 준비하는 학습활동의 성격이 부정적이거나 긍정적으로 영향을 받는 현상을 의미한다(Taylor 2005, 154). 본래 교육의 목표, 내용, 학습활동 그리고 평가 방법이 순차적으로 구성되고 실현되어야 한다고 보았던 틸러나 타바(Tyler·Taba 1962)와 같은 초기 교육과정학자의 생각과는 반대로, 교육의

2) ATC21S 프로젝트의 연구책임자인 Griffin 교수를 필자가 2014년에 면담한 바에 따르면, 본 평가방법은 피험자의 대인관계능력이나 의사소통능력, 협동력 등을 측정하기 위해서 내용이 아닌 참여 학생들의 온라인 대화 반응의 빈도와 속도를 중심으로 컴퓨터 기반 평가를 개발하고 있어, 대화와 협력의 내용을 파악하지 못한다는 한계를 갖고 있다. 장기적으로는 언어학, 뇌과학, 컴퓨터공학 등과의 공동연구를 통해서 대화와 협력의 내용을 파악할 수 있는 컴퓨터 기반 평가방법이 개발될 수도 있을 것이다.

3) 언어학이나 언어교육학에서는 'washback'을, 그리고 그 밖의 영역에서는 'backwash'라는 표현을 더 자주 쓴다고 하나(Taylor 2005, 154), 한글에서는 구분되지 않으므로 어느 영어 표현을 쓰느냐는 크게 중요하지 않다.

www.kci.go.kr

목표 설정 이후에 곧바로 평가 방법과 내용을 구상하고, 그 이후 교육의 내용과 학습활동을 구성할 수 있다고 본 역행설계모형(The Backward Model; Wiggins & McTighe 2004)의 기본가정과 그 의미를 같이 한다고 볼 수 있다. 위에서 언급되었던 SAT과 TOEFL의 변화 그리고 한국의 대입제도의 변천이 사실 역행설계모형을 바탕으로 치밀하게 짜여진 교육과정 계획이나 관련 이론에 근거하여 의도적으로 실시된 것이라고 보기는 어렵겠으나, 암묵적이건 명시적이건 간에, 평가를 개편해야 실제로 이루어지는 학습의 내용과 활동의 종류가 바뀔 수 있다는 논리에 바탕을 두고 이러한 변화가 추진되어 왔다는 점은 부인할 수 없다(Bailey 1999).

위에서 거론된 사례들 이외에도, 시험성적을 바탕으로 학업성취도가 낮은 학생이나 학교에 대한 지원을 도모하였던 미국의 No Child Left Behind(NCLB) 정책이나, 이전 정부가 추진했던 NCLB 정책을 계승 및 수정하여 문항의 개선과 참여하는 주(State)들의 자율성을 제고하고자 하였던 미국의 Race to the Top Assessment 정책 역시 평가의 개선을 통한 교육현실의 발전을 도모하였다는 점에 있어서 역류효과를 기본 논리로 삼은 정책사례들이다(류성창 외 2012, 41-46). 이러한 국내외적인 흐름은 결국 '무엇이 교육적으로 옳은 것인가'를 탐구하고 그것을 현실에 옮겨 놓기 위한 부단한 노력 혹은 몸부림이라고 해석된다. 특히 그 해법을 평가 우선의 전략에서 찾아, 평가를 바꿔 현실을 개선해 보려는, 인간의 행동과 심리에 대한 보다 현실적인 인식에 기반을 하고 있다. 그렇다면 과연 교육적으로 옳은 것은 무엇인가? 그리고 만약 교육적으로 옳은 것이 있다면, 그것을 평가에 담아내어 실제 학습에 까지 이어지게 할 수 있는 방법은 무엇인가?

Ⅲ. 변별력·타당성·효율성의 논리: 수능문항의 예

'교육적으로 옳은 것'을 탐색하기에 앞서 아래 <표 1>의 문항을 검토한다. 아래의 문항은 2010년 11월에 시행된 대입수학능력평가시험의 사회탐구 한국근현대사 영역에 실제로 출제되었던 문항이며, 다른 문항보다 높은 3점이 배점된 문항이다. 문항 밑에 제시된 설명과 정답을 보지 않고 우선 아래 문항을 자세히 한 번 검토해 볼 필요가 있다.

<표 1> 2011년 대학수학능력시험사회탐구 한국현대사 문항예시(출처: 한국교육방송공사 2016)

6. 다음 가상 일기가 쓰인 당시의 모습으로 가장 적절한 것은? [3점]

> 오늘 서대문과 청량리 사이에 처음 설치된 전차의 개통식이 있다고 해서 명동 성당 앞에서 친구를 만나 함께 구경을 갔다. 식장에는 고관대작, 각국의 공사 그리고 민간 유지 등 초청받은 사람들뿐만 아니라 소문을 듣고 온 지방 사람들까지 많은 사람들이 운집하여 매우 혼잡스러웠다. 마침내 화려하게 장식된 전차가 출발하자 함성이 절로 터져 나왔다.

①토지측량업무를 하고 있는 양지아문관리
②고종의 환궁을 요구하고 있는 독립협회회원
③원산학사의 개교 소식을 듣고 기뻐하는 상인
④원각사에서 신극은세계공연을 보고 있는 관객
⑤대한매일신보에 실린 독사신론을 읽고 있는 독자

www.kci.go.kr

위 문항의 해설과 정답은 다음과 같다. 우선 지문에 나타난 행사는 서대문-청량리 구간 전차 개통식으로 1899년의 일이었다. 그리고 아래의 5개의 선택항의 년도는 ① 1899~1903년, ② 1896~97년, ③ 1883년, ④는 1909년 그리고 ⑤는 1908년의 상황이다. 따라서 지문과 기간이 겹치는 ①번이 정답이다. 문항이 출제된 시기의 교육과정인 2009 개정 교육과정의 총론에 따르면 고등학교의 전체적인 교육목표는 성숙한 자기개발능력, 비판적이고 창의적인 능력, 풍부한 문화적 능력 그리고 국가와 국제사회에 대한 안목 등 네 가지의 주요 능력 사항을 핵심적인 교육과정상 목표로 규정하고 있다(교육부 2014, 34-35). 또한 고등학교 한국사 교과의 2009 개정 교육과정 상 교육목표를 보면, 역사적 사건에 대한 세계사적 맥락에서의 이해, 한국사의 정체성 유지과정에 대한 이해, 역사 자료를 분석·비판·종합하여 역사적으로 탐구하고 상상하며 판단하기, 그리고 사회 발전에 역사적 태도로 임하기 등 총 네 가지로 구성되어 있다(국가교육과정정보센터 2015).

이렇게 구체적으로 명시된 교육과정 총론과 교과별 교육과정 상에 제시된 총 여덟 가지의 교육목표가 <표 1>의 대입 시험 문제에 어떻게 반영되어 있는지 심각하게 점검해 볼 필요가 있다. 본 문항은, 최대한 그 의미를 합리화해 보더라도, 총 6가지의 역사적 사건들의 구체적인 년도를 파악하고 기억해내는 것 이외의 다른 능력을 요구하지 않는다. 서로 간의 논리적, 인과적, 유추적이거나 시대적인 연관성이 전혀 없이 6가지의 서로 상관없는 사건들을 오로지 그 발생년도에 대한 기억에 의존해서 풀 수밖에 없는 이와 같은 문항을 통해서 어떻게 한국의 국가교육과정에서 목표로 하고 있는 한국사 교과교육의 본래 취지를 살려낼 수 있을 것인가? 한국의 학생들은 이와 같은 문항이 대입 시험에 출제될 것을 아는 상황에서 과연 한국의 국가교육과정에서 목표로 하는 핵심 능력 혹은 핵심 역량을 제대로 길러낼 수 있을 것인가? 동시에 이러한 상황에서 교사와 학부모들 역시 국가에서 목표로 하는 '제대로 된 교육'에 관심을 가질 수 있는가?

이러한 문항을 왜 출제하게 되는가? 아마도 한국의 교육문화를 경험하였거나 잘 아는 사람이라면 '변별력'이라는 정체불명의 단어를 쉽게 떠올릴 수 있을 것이다. '변별력'이란 무엇인가? 변별력 혹은 변별도(Discrimination)는 문항통계처리에 있어서 고려되는 항목으로 주로 심리측정이나 심리검사의 결과에 있어서 피검자의 결과가 전체적으로 골고루 퍼지고 서로 상이하게 분포되어야 서로 간의 구분이 가능해 지기 때문에 통계처리 상 중요하게 고려되는 요소이다(김성훈 외 2010, 358).

더 나아가서 변별력은 그 개념 상 "집단 의존적(group dependent)이어서, 동일한 검사를 능력이 우수한 집단에 시행하면 정답률이 매우 높아지며 반대로 능력이 낮은 집단에 시행하면 정답률은 매우 낮아진다(김성훈 외 2010, 358)." 즉 변별력이라는 것은 그 본질에 있어서 '능력' 즉 마땅히 길러야 하는 내용과 능력이 이미 고정되어 있는 것이고 변별력의 높고 낮음은 결국 시험을 치르는 사람들의 특성에 따라 좌우되는 것이라는 점이다. 따라서 여기에서 변별력과 관련하여 고려해

www.kci.go.kr

야 할 핵심적인 사안은, 교육과정과는 상관이 없더라도, 변별력만 높일 수 있는 위의 수능 문항과 같은 문항을 굳이 만들어 내어 실제 적용하는 것이 마땅한 것인지, 해당 시험이나 검사가 측정해야 하는 핵심적인 내용과 능력을 벗어나지 않는 선에서도 변별력이 높은 문항을 구성할 수 있는지 등에 관한 것이다.

문항의 변별력과 관련하여 주목해 볼 필요가 있는 두 가지의 세부개념이 있다면 그것은 타당성과 효율성이다. 타당성 혹은 타당도(validity)란 간단히 정의하면 시험이나 검사가 "측정하고자 하는 것을 얼마나 충실히 측정하였는가"를 의미한다(성태제 2002, 341). 측정결과나 연구결과의 검증 항목으로 캠벨과 스탠리(Campbell and Stanley 1966)에 의해 처음 논의된 타당도의 개념은 그 파악하는 방법에 따라 전문가가 검토한 의견에 따라 파악하는 내용타당도, 구인(construct: 하위 요소)들 간에 통계적으로 일관적인 정도(상관계수)에 따라 파악하는 구인타당도, 측정 결과와 향후 실제 수행도 사이의 관계를 분석하여 파악하는 예측타당도, 기존에 널리 인정된 공식 시험이나 검사와의 유사성에 근거하여 파악하는 공인타당도, 그리고 사회 전반에 미치는 영향에 따라 파악하는 결과타당도 등이 있다(Ibid.).

통계적인 산출결과에 의존하는 구인타당도나,[4] 기존의 검사도구와의 비교를 중시하는 공인타당도를 제외하고는, 타당도의 개념은 기본적으로 해당 평가도구나 검사도구가 얼마나 측정내용의 본질을 정확히 반영하고 있는가에 관여하고 있다.[5] 결국, 위 수능 문항과 같은 문항을 위해서 공부하는 한, 교육과정에서 목표로 하는 여덟 가지의 핵심 능력 중 그 어느 것 하나도 길러질 수 없다는 점에 동의할 수 있다면, 위 문항의 (내용)타당도는 매우 낮게 파악될 수밖에 없다.

변별력과 관련된 또 한 가지 개념의 축은 효율성(Efficiency)인데, 시험이나 검사를 치른 사람들 사이의 결과치를 서로 다르게 분산하여 일렬이 되거나 그룹별로 서로 다르게 분류될 수 있는 효율의 정도를 의미한다. 위 수능 문항이 변별력을 위해서 제작되어 시험에 포함되었다면, 문항의 내용타당도는 잃더라도 오히려 한국근대사와 관련하여 교육과정 상 타당한 내용을 마땅히 공부하였을 학생들은 모를 법한 지엽적인 사건의 년도에 관한 문제를 출제하여, 아무 의미 없이 어렵기만 한 문항을 통해서 변별의 효율을 높인 것으로 알 수 있다.

그러나 변별력에 있어서 타당도와 효율성이 모두 중요하다는 것은 상식적으로 생각해 봐도 당연한 논리이다. 예를 들어, 사람의 몸무게를 측정하는 저울이 있다고 가정하면, 사람의 몸무게 수치를 낮은 단위까지 안정적으로 변별해 내는 기능적 효율성도 중요하겠지만, 특정 그룹의 사람

4) 실제 능력의 변화가 측정 결과의 변화와 같은 수준으로 일어난다면 그 타당도를 인정할 수 있다는 새로운 개념의 타당도(유동타당도; Borsboom, Mellenbergh & van Heerden 2004) 역시, 같은 그룹의 문항들을 통계적으로 묶는 구인타당도와 비슷한 방식으로 통계적인 기법을 통하여 타당도의 정도를 파악한다는 점에서, 내용과 관련이 적은 방식으로 타당도를 산출하는 방식으로 볼 수 있다.

5) 실제로 타당도의 개념이 통계적으로 처리되는 타당도와 내용, 이론, 윤리적으로 마땅한 지에 대한 타당도로 나눠 보기도 한다(Messick 1980)

www.kci.go.kr

들의 몸무게가 잘 구분이 되지 않는다고 해서 본래 측정하려고 했던 몸무게가 아닌 허벅지 두께, 체지방 비율, 혹은 콜레스테롤 수치 등을 대신 측정할 수는 없는 노릇이다. 이러한 상황에서 타당도를 쉽게 포기하고 변별 효율성만으로도 문항을 구성하고 있는 한국의 시험 문화는 매우 비상식적인 논리에 근거하고 있다고 해도 과언이 아니다.

교육평가에 있어서 변별력이라는 요소는 중요하게 고려되어야 하는 항목임에는 틀림없겠으나, 한국교육의 현장에서는 수능문제출제를 담당하는 교육전문가를 비롯해서 대다수의 사람들이 그 근본적인 의미를 깨닫지 못하고, 오로지 학생들(대입지원자들)을 효율적으로 가려내야 하는 현실적 필요성과 비슷한 것으로만 오해해서 현실에 적용하고 있다.

한국교육의 현장에서는 변별력과 관련하여 단순히 문항의 기능적인 효율성만을 고려하여 마치 타당도는 낮아도, 즉 해당 교과의 핵심 학습 내용이 아닌 지엽적이고 불필요한 내용이라고 하더라도, 변별력만 높일 수 있다면 중간 중간에 포함시켜도 괜찮은 것처럼 암묵적으로 잘못된 합의가 이루어져 있다. 그러나 교육평가의 핵심 개념 중 하나인 '문항의 변별력'이 적용될 때에는 반드시 그것은 어디까지나 '교육적'으로 타당한 문항이어야 하고 '기능적'으로 효율적이어야 한다는 두 가지 조건이 충족되어야 마땅하다.

Ⅳ. 결론: 한국교육의 문화

2014년 영국의 런던에서는 매우 흥미로운 프로젝트가 발표되었다. 행복한 지도(Happy Map)라 불리는 이 프로젝트는 기존의 인터넷 기반 경로 찾기 프로그램들이 목적지까지의 최단 시간이나 최단 경비 등 효율성에 기초한 조건으로만 검색이 가능했던 것을, 오래 걸리고 돌아가더라도 많은 사람들이 행복한 추억이 있는 장소로 기록된 곳들, 아름답다고 표시한 곳들, 다양한 냄새와 색깔이 배어 있는 곳들을 거쳐서 갈 수 있는 그야 말로 거쳐가면서 행복해 질 수 있는 경로를 검색해서 표시해 주는 소프트웨어와 스마트폰 앱을 개발한 것이다(Armitage 2014). 2016년 현재는 런던대학의 연구자들이 런던을 중심으로 행복한 지도를 완성하고 있으며, 점차 구글맵(Google Map)을 기반으로 한 Happy Map을 확대하기 위해 행복한 정보수집을 주 목적으로 하는 'Mappiness 프로젝트'도 가동하고 있는 것으로 알려지고 있다(Ibid.).[6] 지금까지는 거리, 시간, 연비 등 효율성 관련 지표들에 기초한 도구들이 개발되었다면, 앞으로는 효율성이 떨어져도 보다 인간다운, 보다 행복한, 보다 의미 있고 즐거운 것들을 추구할 수 있는 '인간적으로 타당한(humanly valid)' 방향으로 각종 도구들이 개발될 징조를 보이기 시작했다.

한국의 교육은 이제 효율성에서 타당성으로 그 관심을 옮겨야 할 중요한 시기에 놓여 있다.[7] 한국의 시험 문제는 이제

6) 명연사들이 강연하는 TED에도 관련 연설이 올려져 있으며, 아래의 URL 주소에 가면 시청할 수 있다.
https://www.youtube.com/watch?v=AJg9SXIcPiM

7) 한국에는 교육 분야뿐 아니라 다른 분야에도 전반적으로 효율성을 타당성보다 우선시하고 중요시하는 문화가 넓게 자리 잡고 있다. 예를 들어, 주거유형 선택에 있어서도 효율성이 중요한 요인으로 파악된다(정재영

www.kci.go.kr

얼마나 학생들을 줄을 잘 세울 수 있는지의 문제에서 얼마나 마땅히 교육적으로 옳은 것을 충실하게 평가할 수 있는지에 초점을 맞추어야 할 시점이다. 한국교육계의 혜안과 노력으로 사실 앞서 언급한 여러 종류의 교육개혁이 오랜 시간 동안 추진되어 왔으며, 사실상 거의 모든 교육개혁의 노력은 원칙적으로 타당성을 목표로 하여 실행되었다고 해도 과언이 아니다.

다만, 그러한 정책적인 노력들이 현장의 교수학습의 변화를 이끌어 내기에 역부족이었던 이유가 앞서 검토한 것과 같이 '문항의 문제'를 등한시하였기 때문은 아닌가를 고려해 보아야 한다. 큰 틀에서 상대평가를 절대평가로 바꾸고, 입학사정관제도를 도입해서 다양한 근거자료를 반영한다 해도, 실제로 학생들이 풀어내야 하는 구체적인 문항들이 여전히 교육적 타당성보다는 기능적 효율성에 기반을 둔 문항들로 구성되고 있는 한, 실제 학교교실에서 일어나는 교수학습의 현실은 크게 바뀌기 어려울 것이다. 결국 문항이 바뀌지 않으면 한국의 교육은 바뀌지 않을 것이다.

이러한 문제점도 현장의 교사와 교육전문가들을 너무 잘 알고는 있으나 현실에 옮기는 것이 어려운 상황이라는 점을 지적할 필요가 있다. 예를 들어, 최근에 도입된 수행평가의 경우 교육적으로는 옳다는 그 중요성과 필요성은 대체적으로 공감하면서도 실제 적용에 있어서는 표면적인 공정성과 객관성, 그리고 학부모의 공감도가 부족하다는 이유로 인해 정상적으로 실행되고 있지 않는 현실적인 한계점이 있다(Lee et. al. 2014; 김성훈 외 2010, 257). 이러한 한계 상황에서 한국교육은 여전히 타당성에 집중해야 할 것인가, 아니면 타당성을 희생한 효율성을 여전히 반영할 것인가?

이 문제를 고려함에 있어서 자동차를 비유로 한 번 생각해 볼 수 있다. 자동차의 경우 타당성과 효율성의 문제가 다소 상식적으로 정착된 사례라고 볼 수 있다. 예를 들어, 소위 말하는 비싼 자동차의 경우는 연비나 유지비, 수리비 등 높은 비용에 얻는 것이 적어 효율성은 낮다. 반면 각종 편의장치나 안전성 등에 있어서는 우위를 보이기 때문에 효율성이 낮아도 타당성이 높은 '좋은 자동차'로 분류된다. 반대로, 경차의 경우, 편리함이나 안전성에 있어서는 한계가 있지만, 연비, 유지비, 수리비 등 효율성은 월등하다. 즉, 소위 비싼 자동차는 효율성보다 자동차의 본래 목적을 충실하게 달성하는 그 타당성이 높아 좋은 자동차로 인정받고 그 가격도 높게 형성되어 있는 반면, 그 반대의 경우는 타당성보다 효율성이 높게 제작되었지만 가치와 가격은 낮게 설정된 것을 알 수 있다.

이와 비슷하게도, TOEFL시험의 말하기나 쓰기의 채점 과정이 간혹 주관적일 수도 있고 일관적이지 않을 수도 있겠으나, TOEFL은 말하기와 쓰기 등 소위 질적 평가(qualitative assessment)를 50%까지 끌어 올리는 방식으로 진화하였고, 이

·윤태권 2008, 56). 2016년 현재 모든 주택유형 중 30%정도만이 단독주택이며 나머지 70%는 아파트나 여러 가구가 한 건물에 모여 사는 형태로 파악되고 있다(최막중·임영진 2001, 71). 자연환경에 둘러싸인 단독주택을 이상적인 주거형태로 여기는 서양과 비교하면 효율성 중심의 주거형태 역시 한국문화의 특징을 잘 보여주는 사례라 하겠다.

www.kci.go.kr

러한 과정에서 타당성이 효율성보다 더 중요시된 것을 알 수 있다. 즉 이와 같이 개선되어야 실제로 TOEFL을 대비하는 학생들이 마땅히 공부해야 하는 내용, 즉 말하기와 쓰기를 연습하게 될 것이며, 평가와 학습에 있어서 그 보다 더 중요하게 고려해야 할 요소는 없다는 인식이 자리 잡은 것이다. 마찬가지로 한국교육이 선진 교육으로 도약하기 위해서는 궁극적으로 효율성보다 타당성을 더 중요시하는 교육문화가 자리 잡아야 할 필요가 있다.

여기에 바로 '교육적으로 옳은 것'을 파악하는 일이 한국교육을 구상함에 있어서 중요한 과제로 인식된다. 만약 효율성보다 타당성을 중요시하는 교육이 좋은 교육이라면, 과연 교육적 타당성, 즉 교육적으로 마땅히 가르치고 배워야 하는 것은 도대체 무엇인가? 과연 교육적으로 옳은 것은 무엇인가? 교육이 정치발전을 위해 도움이 되면 그것은 교육적으로 옳은 것인가? 혹은 경제발전에 도움이 되거나 사회계층의 재구조화에 기여할 수 있으면 교육적으로 옳은 것인가? 아니면 또 다른 차원의 '그 자체로서 교육적으로 옳은' 방향이나 목표, 혹은 본질이 과연 있는 것일까? 교육의 이론이 '교육받은 인간'의 개념을 가장 근본적인 논점의 출발점으로 삼는 것 역시 교육의 효과, 결과, 효율을 추구하기보다는 교육적으로 옳은 방향을 탐구하기 위한 것이다. 이에 본고는 한국교육의 정책과 현실의 방향 탐구에 있어서 교육적 타당성을 보다 핵심적으로 추구하여 반영할 것을 요청한다.

www.kci.go.kr

참고문헌

교육부. 2014. 2009 『개정 교육과정의 부분 개정에 따른 고등학교 교육과정 해설 총론 증보편』

국가교육과정정보센터. 2015. “우리나라 교육과정-영역/교과목별-사회-한국사-2009개정시기-고등학교.” http://ncic.kice.re.kr/nation.kri-.org4.inventoryList.do, 2015/12/01

김성훈·김신영·김재철·반재천·백순근·서민원. 2010. 『예비교사를 위한 교육평가』. 서울: 학지사.

류성창·장혜승 서예원·강태훈 정제영. 2012. 『역량 중심 대입제도 개편 가능성 탐색 연구』. 한국교육개발원. OR2012-09.

서지영·김소영·홍수진·신명선. 2008. 『학교교육 내실화를 위한 수행평가 개선 연구(Ⅰ)』. 한국교육과정평가원. RRE2008-1.

성태제. 2002. 『현대교육평가』. 서울: 학지사.

정광희·박병영·조덕주·한혜정·이관희. 2010. 『고교-대학 연계를 위한 대입전형 연구(Ⅷ): 고교-대학 연계 실천 모델 및 운영 방안 개발』. 한국교육개발원. RR2010-21-1.

정재영·윤태권. 2008. “아파트 수요자의 선호요소에 대한 AHP 분석에 관한 연구.” 『한국건축시공학회 논문집』 8(3), pp. 51-58.

최막중·임영진. 2001. “가구특성에 따른 주거입지 및 주택유형 수요에 관한 실증분석.” 『국토계획』 36(6), pp. 69-81.

허경철·백순근·박경미·최미숙·양길석·김광주. 1999. 『수행평가 정책 시행 실태 분석과 개선 대책 연구』. 한국교육과정평가원. CRE1999-2.

한국교육방송공사. 2016. “지난수능/학평보기.” http://www.ebsi.co.kr/ebs/xip/xi-pa/retrieveSCVLastExamList.ebs, 2016/01/01

www.kci.go.kr

References

Armitage, Jacob. 2014. “Happy Map: Mapping Happiness.” Goldsmiths, University of London. http://jacobarmitage.co.uk/wp-content/uploads/2014/03/Happy-Map-Mapping-Happiness-By-Jacob-Armitage-March-2014.pdf. Accessed 2015/12/1.

ATC21S. 2015. “Assessment and Teaching of 21st Century Skills.” http://www.atc21s.org. Accessed 2015/9/1.

Bailey, Kathleen M. 1999. *Washback in Language Testing*. ETS Research Memorandum RM-99-04, TOEFL-MS-15. Princeton, NJ: Educational Testing Service.

Campbell, Daniel T. and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Design for Research.* Chicago: Rand McNally.

CNN. 2014.04.03. “Major Changes Coming to 2016 SAT Test: Here’s What, How and Why.” http://edition.cnn.com/2014/03/05/living/sat-test-changes-schools. Accessed 2015/12/1.

ETS. 2011. *TOEFL iBT Research: TOEFL Program History.* Series I, Volume 6. NJ: Educational Testing Service.

Griffin, Patrick and Esther Care, eds. 2011. *Assessment and Teaching of 21st Century Skills: Methods and Approach.* New York: Springer.

Hughes, Arthur. 1989. *Testing for Language Teachers.* Cambridge: Cambridge University Press.

Lee, Ju-Ho, Sung Chang Ryoo, and Sam-Ho Lee. 2014. “From Multiple Choices to Performance Assessment: Theory, Practice and Strategy.” *KDI School Working Paper Series.* Seoul, Korea: KDI School.

Mesick, Samuel. 1980. “Test Validity and the Ethics of Assessment.” *American Psychologist* 35 (11): 1012-1027.

NACAC. 2013. “State of College Admission Report.” http://www.nacacnet.org/research/PublicationsResources/Marketplace/research/Pages/StateofCollegeAdmission.aspx. Accessed 2015/12/1.

Slatella, Michelle. 2007. “ACT vs. SAT.” *New York Times*, November 7. http://www.nytimes.com/2007/11/04/education/edlife/guidance.html.

Taba, Hilda. 1962. *Curriculum Development: Theory and Practice.* New York: Harcourt Brace and World.

Taylor, Lynda. 2005. “Washback and Impact.” *ELT Journal* 59 (2): 154-155.

Wiggins, Grant and Jay McTighe. 1998. *Understanding by Design.* Alexandria, VA: Association for Supervision and Curriculum Development.

www.kci.go.kr

국문참고문헌의 영어표기

(Korean References in Roman Alphabet)

Choi Mak-Jung and Lim Young-Jin. 2001. "Empirical Analyses of the Relationships Between Household Characteristics and Preference of Residential Location and Housing Types." *The Journal of Korea Planners Association* 36 (6): 69-81.

Chung Jae-Young and Yoon Tae-Kwon. 2008. "A Study on the Preference Analysis of Apartment Purchaser Using the AHP Method." *Journal of the Korea Institute of Building Construction* 8 (3): 51-58.

Department of Education. 2014. *2009 National High School Curriculum Guidebook, Based on Partial Revision.*

Educational Broadcasting System. 2016. "Viewing Past Korean CSATs." http://www.ebsi.co.kr/ebs/xip/xipa/retrieveSCVLastExamList.ebs. Accessed January 1, 2016.

Hur Kyung-Cheol, Back Soon-geun, Park Kyung-Mi, Choi Mi-Suk, Yang Kil-Suk, and Kim Kwang-ju. 1999. "Analysis on the Implementation and Improvement of Performance Assessment Policy." Korea Institute of Curriculum and Evaluation. CRE 1999-2.

Jung Kwang-Hee, Park Byung-Young, Cho Duck-Joo, Han Hye-Jung, and Lee Kwan-Hee. 2010. *Development of Action Models for the Connection between High School and College Education and the Development of Operation Strategy.* Korean Educational Development Institute. RR2010-21-1.

Kim Sung-Hoon, Kim Shin-Young, Kim Jae-Chul, Ban Jae-Chun, Back Soon-Geun, and Seo Min-Won. 2010. *Educational Assessment for Pre-service Teachers*. Seoul: Hakjisa.

National Curriculum Information Center. 2015. "National Curriculum-Areas/Subjects-Social Sciences-History-2009 Revision-High School." http://ncic.kice.re.kr/nation.kri.org4.inventoryList.do. Accessed December 1, 2015.

Ryoo Josh Sung-Chang, Chang Maria Hye-Seung, Seo Ye-Won, Kang Tae-Hoon, and Chung Jae-Young. 2012. *Competency-based College Admissions Policy Reformation.* Korean Educational Development Institute. OR2012-09.

Seo Ji-Young, Kim So-Young, Hong Su-Jin, and Shin Myoung-Sun. 2008. *Research on the Improvement of Performance Assessment for More Effective School Education in Korea (I).* Korea Institute for Curriculum and Evaluation. RRE 2008-1.

Sung Tae-Je. 2002. *Contemporary Educational Assessment.* Seoul: Hakjisa.

www.kci.go.kr