

행동과학에서 측정평가의 과거와 현재: 한국사에서의 고찰*

이 순 목	김 아 영	백 승 근 [†]	강 민 경
성균관대학교	이화여자대학교	아주대학교	성균관대학교

본 연구에서는 한국문화에서 인간행동의 측정 및 평가에 관한 흐름을 검토하고 학문적 정체성과 발전의 현주소를 밝히고자 하였다. 한국사에서 측정평가의 문화적 기반을 찾기 위해 역사문헌을 검토하고 해방 후 연구를 포괄하고자 여러 학술DB로부터 현대의 측정평가 연구에 대한 기초적인 통계를 산출하였다. 고조선 시대의 수리적 사고로부터 삼국시대와 통일신라의 판단적 평가인 교육기관 추천, 관직에의 천거, 그리고 객관적 평가로서의 통일신라의 독서삼품과, 고려시대의 과거제도 도입, 그리고 조선시대의 과거제도 확대까지 시대의 변천에 따라 주관적 평가에서 객관적 평가로 무게중심이 이동하였다. 서구학문의 유입은 측정 및 검사이론, 통계학의 수입으로 기존의 행동평가를 정교화시켰고, 실험과 조사방법의 도입은 유학의 질적 방법론에 현대적 연구방법들을 더하였다. 현대적 측정평가는 해방 이후 도입되어 측정 및 검사이론, 통계, 도구(개발) 및 평가, 그리고 연구방법으로 분화 발전하였다. 도입기(대한민국 수립 후 15년간)-구축기(60년대부터 20여 년간)-발전기(80년대~2000년)를 거쳐 21세기에는 14년간 측정평가의 학술활동이 비약적으로 발전하였다(2014년 말까지 총 3,539건). 1960년대부터는 학술지 중심의 연구 활동이 빠르게 성장하였고, 영역별로는 도구/평가 영역의 발전이 두드러졌으며 검사이론이 그 뒤를 따른다. 산업 측면에서 심리검사 영역은 공식적으로 집계되지 않고 있지만 최소 30개 이상의 회사가 설립·운영되고 있으며, 여론 및 사회조사 회사는 411개 기업에 8,184명이 종사(2013년 현재)하는 규모로 성장하였다. 마지막으로 질적 측면을 충분히 감안하는 판단적 평가의 중요성을 재강조하고, 양적 평가와 질적 평가의 균형 발전을 위한 노력에 대하여 논의하였다.

주요어: 측정 및 평가, 판단적 평가, 객관적 평가, 유교적 가치, 역사

* 이 글의 초고에 유익한 지적을 주신 심사위원께 감사드린다. 이 글은 2015년 7월 14일 중국 북경에서 있었던 국제심리측정학회(IMPS)에서의 심포지엄 “Psychometrics in Asia”에서 발표된 “Past and Present of Psychometric Work in Korea”를 수정 보완 한 것임을 밝힌다.

† 교신저자: 백승근, 아주대학교, 주소

Tel: 031-219-1930, E-mail: 100tomak@hanmail.net

본 연구는 한국문화에서 인간행동(내적/외적 행동)에 대한 측정 및 평가에 관한 흐름을 검토하고 일제 강점기 동안 단절을 경험한 학문의 역사에서 과거와 현재를 이어봄으로써 한국문화에서의 오래된 관점과 방법을 통해 측정 및 평가 분야의 학문적 정체성을 명확히 하고자 진행하였다. 또한 새로운 기술의 발전과 도입에도 불구하고 측정과 평가에 있어 동일한 주제의 문제점이 반복되고 있음을 확인하고 개선을 위한 미래 연구방향을 제시하고자 하였다.

한국인은 유교 또는 유학을 인간 심성(心性)의 학문으로서(이승환, 1998) 기원전부터 접해왔다. 유학을 수입하여 교육과 통치의 원리로 삼고자 했던 삼국시대 이래로 공식/비공식으로 이루어진 교육기관 입학용 평가, 관리 선발용 평가, 관료의 역량 평가에 따른 관등의 부여, 무술 및 병략의 평가에 따른 군 지휘권 부여는 국가 역량강화의 목적 하에 교육, 행정, 군사 방면에서 유교적 가치를 기반으로 한 행동평가의 오랜 제도와 관행이 있어왔음을 말해준다. 또한 고대 국가 체제를 정비하기 위한 일환으로 역사서의 편찬, 율령의 제정은 당대 최고의 학자들로 하여금 유교적 가치를 기반으로 문자에 의한 기록, 경전의 해석은 물론 논리비평적 접근을 통해 자연스럽게 심성에 대한 진지한 숙고를 하게 하였다. 진지한 숙고의 도구로서 기록, 해석, 논리비평적 접근은 모두가 오늘날의 질적 연구방법에 속하는 것들이다. 삼국시대 이래로 해석/논리비평의 방식이 꾸준히 발전되어 왔음은 유교적 문화 속에서 이미 오래전부터 심성의 연구에 질적 방법이 깊이 있게 사용되어 왔음을 의미한다. 이처럼 한국사에서 행동평가 및 연구방법은 근대의 서구 문물이 수입되기 훨씬

전부터 고대의 국가적 목적에서 일찍이 시작했음을 알 수 있다.

한국 역사에서 문화로 자리 잡은 행동과 사유에 대한 유교적 습성(조궁호, 1998)은 인간행동의 측정평가에 커다란 밑그림이 되었다. 고대국가의 성립 이래로 실시된 행동평가와 질적 연구방법의 전통은 심성에 대한 평가 및 연구의 영역으로 계속 발전되어 왔으며 현대에 와서 행동과학의 수입과 더불어 측정/검사 이론과 통계의 영역이 추가되면서 확대되었다고 볼 수 있다. 즉, 오래된 심성평가 및 연구의 문제를 현대적 행동과학과 방법에 의해 보다 기술적으로 접근하게 된 것이다. 따라서 현대 한국에서 발생하는 측정평가의 개발 및 문제에 대하여 20세기의 학문적 논의와 함께 한국문화에서의 맥락에 기반한 이해를 더 할 필요가 있다. 그러기 위해서는 과거 2천 4백여 년에 걸쳐 교육, 행정, 군사 분야에서 이루어진 평가 및 연구방법에 대한 고찰이 있어야 한다.

오늘날 행동과학에서 측정평가의 영역은 크게 네 가지 분야, 즉 측정/검사 이론, 통계, 도구(개발) 및 평가, 연구방법으로 나누어 볼 수 있다. 앞의 두 영역은 60년대 이후 서구 학문을 습득한 개척자들이 국내에 씨앗을 뿌리면서 시작되었다. 세 번째 영역인 도구개발과 평가는 사회조사, 개인 또는 기관평가의 양적 발전으로 나타났다. 그런데 도구/평가와 연구방법의 두 영역은 한국의 오랜 역사에서 일찍이 출발한 상태에서 서구 학문의 수입으로 더욱 폭넓게 발달하였다. 현대의 측정과 평가에 관한 문제를 논의함에 있어서도 서구에서 유입된 이론과 기법에 더해 우리 역사 속에서 형성된 과거의 학술적 고뇌와 문화적 바탕을 고려한다면 타당성과 효용성에 있어 보다 큰

가치를 가진 논의가 될 수 있을 것이다. 통합적이고 주체적인 논의를 위해서는 한국 역사 속의 측정평가를 살펴봄으로써 학문적 현주소의 기원에 대한 자부심을 가지고 그에 더하여 미래지향적 연구를 생각할 필요가 있다.

에 의해 해석되고 구성되는 것임을 감안할 때 본 연구에서 수용할 수 있는 연구 방법으로 채택되었다.

결 과

방 법

유교적 가치의 영향

본 연구의 성격상 경험 자료의 분석에 따른 실증 연구의 성격을 갖지는 않는다. 과거에 대해서는 역사 서적에서 인용하면서 해석을 하였고, 현재에 대해서는 해방 후의 학술활동의 통계를 구하였다. 따라서 어느 정도의 통계는 사용되지만, 상당부분 해석적인 방식을 사용하고, 후반부에는 논리비평적 방식을 활용하였다.¹⁾ 현대 한국의 행동과학에서 측정평가 연구에 대한 통계를 구하기 위하여 한국심리학회 DB, 한국교육학술정보원(KERIS), 한국학술정보원(KISS), 국내 생산되는 모든 학술논문이 소장되는 국회도서관 DB, 그리고 국내서적이 가장 많이 소장되는 국립중앙도서관 DB, 더하여 서적에 대한 분류가 잘 되어 있는 성균관대학교 도서관 DB를 섭렵하여 문헌을 검색하고 정리하였다. 이후의 기술에서, 한국사에서의 심리측정 부분을 탐색하고 정리하는 관점에서 사용된 해석 및 논리적 비평이 주관적 판단으로 비추어질 수 있으나 역사는 필자

‘유교적 가치’는 한국사에서 국가의 교육체제와 행정체제 운용, 통치원리 확립에 심대한 영향을 미쳤고, 그 영향은 현대에도 지속되어 교육체제, 행정기관, 군대뿐 아니라 기업경영에도 중요한 맥락을 제공한다. 큰 맥락에서 우리 민족의 두드러지는 특징으로 서구의 개인주의적 가치관과 가장 크게 대비되는 집단주의 문화는 유교적 사상과 크게 관련이 있다(조공호, 2007). 예를 들어, 우리의 가치관, 행동, 및 사고방식이라는 심리적 개념들의 윤곽을 그려내고 내용을 채우는데 가장 큰 몫을 차지하고 있는 유교적 영향들 가운데는 수재교육제도, 근면/성실/청렴의 인성을 강조하는 유교적 가치관, 신용과 신의에 기반을 둔 공동체로서의 사회, 지도자들에게 기대되는 모범적 도덕행동이 포함된다(이광세, 1998). 즉, 유학사상의 인성론과 유교적 가치는 오늘날 우수한 인재의 선발/교육을 목표로 하는 교육체제, 행정관료의 선발 기준으로서의 가치관, 기업의 사회적 책임, 그리고 민간과 군대사회를 가리지 않고 지도자에게 기대되는 모범적 행동 등을 이해하는데 중요한 기반이 된다(김충렬, 1982).

이렇게 한국사회에 전반적 영향을 미친 유교적 가치와 원리는 고조선 시대부터 시작하여 2천 4백년 이상을 거치면서 한국인의 심리

1) 질적연구방법으로서 해석과 논리비평은 각각 구분되는 방법론이자 활동이며, 본 연구에서는 관찰의 대상이 되는 현상이 그것의 맥락에서 갖는 의미를 이해하고 밝히는 과정을 해석으로, 현상이 전제하고 있는 원칙에 따라 적절성과 타당성을 논하는 것을 논리비평으로 구분하여 다루었다.

적 요소와 활동에 큰 부분을 차지하게 되었고 측정평가도 예외는 아니다. 현대 심리학에서 다루는 측정평가의 4개 하위영역 중 먼저 도구/평가와 연구방법에 관한 사례들을 한국사에서 살펴볼 수 있다. 역사를 고찰하는 맥락에서 볼 때 도구는 다양하지 않았으므로 도구/평가는 “평가” 또는 “행동평가”로 부르기로 한다. 연구방법은 초기에는 주로 질적인 방법이었으나 수학의 발전과 더불어 양적 방법도 사용되었다.

이를 상세히 살펴보면 우선 평가 영역에는 수리적 작업이 포함되는데, 이미 고조선 시대(BC 4세기 이전-108 BC) 그리고 비슷한 시기에 출발하여 고구려 장수왕 때(494 AD)까지 존속한 부여에서 백성을 다스리기 위한 법조문을 보면 양적 개념과 행동에 대한 평가를 볼 수가 있다(예: 고조선의 8조의 법금). 개인의 능력과 행동을 평가하는 판단적 평가나 추천제도는 고대국가 이전 연맹왕국의 모습인 고구려나 부여에서 귀족들이 왕을 선출하는 것에서 볼 수 있으며, 더 한층 구체적으로는 고대국가 체제를 갖춘 삼국시대부터 볼 수가 있다(이기백, 1998; 한영우, 2015). 그 시대에 관리가 되기 위해서는 국가 차원의 교육기관에 입학하기 위한 자격이 있어야 하고, 교육 후 국가 기관의 직책을 맡기 위해서는 넓은 의미의 행동(오늘날 행동과학에서 말하는 내적 행동, 외적 행동)에 대한 평가를 거치는 것이 요구되었다. 이러한 평가의 목적은 크게 두 가지를 판단하기 위함이었다. 즉, 당시의 통치원리 및 사회규범의 핵심이 되는 유학의 학습과 인성, 그리고 주변국과의 전쟁가능성에 대비한 무인으로서의 준비도라고 할 수 있으며, 삼국사기에는 신라의 화랑도를 통해 우수한 자를 조정에 천거한 것에 대해 “어진 보

필자와 충신은 이로부터 나왔고, 훌륭한 장수와 용감한 병졸은 이로부터 생겼다.”고 표현하였다. 즉, 이러한 평가의 두 영역은 결국 유교적 원리에 입각한 개인의 능력과 수행을 포함한 행동전반의 평가를 의미한다(노중국, 2014).

오늘날 국내에서 평가는 학교, 사기업, 공조직, 군인을 양성하는 군조직, 교육기관 및 많은 검사회사의 중요 업무가 되어 있다. 각종 학교 입학시험 및 학업평가는 물론 군조직에서의 장병선발 및 정신건강에 대한 평가, 조직에서 고용관련의 평가, 임상장면의 검사 및 진단은 모두가 오랜 한국역사 속에서 문화의 한 부분이 된 행동평가의 방법론이 서구에서 유입된 측정/검사 이론 및 행동과학 통계의 발달과 융합되어 오늘날 폭넓게 적용되고 있다.

연구방법 역시 한국 역사에서 전통적인 방법인 해석적 방식과 논리비평적 방식이 서구에서 유입된 많은 양적/질적 방법론과 조화되어 양적으로는 실험방법과 조사방법을 주축으로 하고, 질적으로는 철학적 방법론을 비롯하여 사례연구, 문화기술적 방법, 근거이론 접근 등 다양한 방법이 도입되었다. 오늘날 질적 연구의 방법론에서 한 쪽이 양적 연구에 가까운 근거이론(grounded theory)이라면(Corbin & Strauss, 2008), 다른 한 쪽은 철학에 가까운 현상학적 질적 연구(Giorgi, 1985), 그리고 철학적 방식인 해석/논리비평(Russell, 1959/1990)이 있으며, 이 중 해석/논리 비평은 한국사의 여러 장면에서 등장하는 전통적인 질적 연구방법에 해당한다.

유교의 역할과 행동평가

유교의 유입에 대한 정확한 연대는 알 수

없으나 중국과 국경을 접하고 있던 고조선(B.C. 4세기 이전~B.C. 108년) 시대 또는 그 이전부터 문화교류를 통해서 그 사상이 한국 문화에 유입되기 시작한 것으로 보고 있다(한국민족문화대백과, '한국상고 및 삼국시대와 유교'). 한국사에서 청동기의 시작은 B.C. 10세기에서 B.C. 4세기까지로 추정하고 있는데 이 기간 중에 여러 성읍국가가 탄생하였고 그 중에서 가장 선진사회는 고조선이었다. 대체로 B.C. 4세기경에는 중국에까지 고조선의 존재가 알려졌다(이기백, 1998). 양적인 개념(quantitative concept)과 행동평가는 이미 고조선 시대의 8조의 법금(法禁)에서 발견된다. 예로서 고조선의 8조의 법조문 가운데 “남의 물건을 훔친 자는 데려다 노비로 삼는다. 단 스스로 속죄하려는 자는 1인당 50만전을 낸다”(한영우, 2015, p.90)고 할만큼 양적으로 행동을 평가하는 개념이 발달하였다.

유교적 가치가 한국문화에서 평가의 필요를 가져오게 되는 것은, 같은 씨족의 구성원들 속에서 지배계층이 생겨나는 성읍국가(walled-town state)가 성장하고, 그 연맹체인 연맹왕국(confederated kingdom) 체제를 넘어 강력한 중앙집권 국가로 발전하는 과정에서 당시 한국의 통치체제였던 유교적 가치의 유지 및 강화를 위해 인적 속성평가의 필요성이 발달하였다. 고조선이 성읍국가로 출발하여 연맹왕국을 거쳐 후반기인 위만조선에 이르는 과정을 살펴보면 이를 확인할 수 있다²⁾. 위만조선이 주변의 유이민 집단을 흡수하는 정복국가로 발전함에 따라 인구증가, 행정관리의 존재, 신

분의 분화, 전문직 종사자의 필요, 전쟁, 징세, 징병 등의 국가 업무가 수행되었다(최몽룡, 2002). 물론 현재까지 발견된 사료는 8조 법금이 유일하나 마치 두 번의 세계대전을 치르는 과정에서 양질의 병력을 선발하고 양성하기 위해 심리검사가 발달한 것과 같이, 한국 고대사에서 통치체제가 성장하는 과정에서 필요한 전문적인 인력을 수급하기 위해 개인의 능력과 수행에 대한 평가가 중요했음을 유추할 수 있다.

그러나 삼국시대, 즉 고구려(37 B.C.-668 A.D.), 백제(18 B.C.-660 A.D.), 그리고 신라(57 B.C.-668 A.D.)에 왕을 중심으로 한 중앙집권의 강화, 통치원리와, 사회적 규범, 그리고 교육체제를 제공할 필요에서 유학이 중요시되었고 그에 따른 인적 자원의 선발, 관등의 분류, 승진 등의 결정을 위한 업무평가, 군사지도자의 선발 등을 위해 측정평가의 중요성이 대두되었다. 당시의 관행으로 보아 시험제도가 도입되기까지는 특별한 도구 없이 추천(천거)제도를 통한 판단적 평가가 주된 부분이었다. 당시에 삼국은 상당한 수준의 문화적 교류가 있었으므로, 국가의 교육기관에서 훈련시킬 학생을 선발하고자 후보자의 사회적 위치나 혈통을 평가하는 방식이 일반적이었고, 그 기관에서 훈련받은 학생 가운데 정부 관리나 군사 지도자로 선발하기 위해서 평가를 하는 전통이 확립되었다. 예로서, 고구려(372, 태학), 백제(4C 후반, 오경박사제도), 및 신라(682, 국학)는 다투어 유교적 교육제도를 정비하였다. 고구려의 국립대학인 태학에서는 당시 중국의 수학 교과서인 구장산술 및 주비산경을 교과서로 사용했는데, 고구려 고분에 대한 사료 발굴에 비추어 볼 때 수학의 수입과 사용은 태학을 세우기 훨씬 이전부터로 추정된다(이

2) 고조선의 후반부를 차지하는 위만조선의 창업자 위만은 과거 연나라가 침공하여 연에 복속된 고조선 지역의 토착 인물로서 조선계로 파악된다(최몽룡, 2002).

장주, 2014). 구장산술(九章算術) 과 같은 초기 수학 교재들이 현실의 문제를 해결하기 위해 필요한 수학적 절차들을 안내하고 있음을 감안하면 시대의 한계는 있으나 역사의 한 시기에 수학이 발달한 수준에 걸맞은 통계적 발상이 행동평가의 관행을 지원해 주었을 것이라 짐작할 수 있다. Michell(1990, p.9 참조)은 심리 측정과 양적 개념이 온전히 현대 심리학자들(예: Campbell, Helmholtz, Hölder, Russell)의 노력으로 이룬 성취라고 생각하는 “20세기적 오만(20th century arrogance)”에 경종을 울리고 그 기원을 고대 그리스의 수학자들과 철학자들의 공헌, 즉 피타고라스(582?-500? B.C.)와 플라톤(427-347 B.C.)의 계량적 사고와 그 뒤의 아리스토텔레스(384-322 B.C.)의 양-질 구분에 돌리고 있다. 당연히 우리의 측정평가의 역사도 유학 및 수학 발전의 역사와 궤를 같이하는 면이 있을 것이나 그것은 이 글의 범위를 넘는다.

국립교육기관의 설립은 새로운 관료체제에 인재를 공급하게 하였고, 중앙정부에서는 관료의 역량을 평가하여 그 수준에 따라 관등의 분류(고구려는 14관등제, 백제는 16관등제, 신라는 17관등제)가 정비되었으며, 아울러 군대에서도 왕을 충사령관으로 하는 전국적인 군대편성이 이루어졌다(이기백, 1998). 국가 수준의 교육체제에 입학하는 자격, 거기서 훈련받은 자에 대한 평가를 통한 관료로의 선발, 왕의 군대에서 지휘관으로 일할 무장의 선발은 모두가 유교의 원리에 기반하여 통치체계를 갖추고자 하는 삼국시대에 필요한 행동평가의 논리와 근거를 제공하게 되었다.

이러한 교육, 행정, 군사 측면에서 선발/훈련/업무수행에 대한 행동평가의 전통은 삼국을 통일한 통일신라, 그 후의 고려, 그리고 마

지막 왕조인 조선에 이르러 시험제도가 도입 발전되고 점수가 산출되면서 점점 더 객관적 방법을 도입하는 추세를 보였다. 통일신라 때는 과거제도의 전사(前史)인 독서삼품과(전덕제, 2010), 고려에서는 후주인 쌍기의 건의로 실시하게 된 과거제도(문관에게만 실시)가 있었고, 조선에서는 과거가 문관과 무관 모두에게 실시되면서 양반 관료체제를 만들어내었다(이남희, 2008). 행동평가에서 주관적 판단의 부분을 축소하고 객관적 결과에 중점을 두는 경향의 지속은 한국 역사에서 유교적 통치와 행정 및 교육을 이끌어갈 인재선발과 평가의 발전사적 특징이 되었다. 이러한 유교적 영향을 배경으로 한국사에서 측정평가의 과거와 현재를 조명하였다.

행동평가의 역사: 추천과 시험

그림 1에서와 같이 한국사에서 평가의 역사는 판단적 평가(추천)와 객관적 평가(시험점수 중심) 간에 무게 중심이 지속적으로 객관적 방향으로 움직여 왔다고 볼 수 있다. 그에 따라 판단적 평가를 상당부분 사용하고 양적인 개념이 활용되는 초기에 해당하는 삼국시대로부터 현대의 측정이론과 연구방법론이 사용되기 전인 가장 최근의 조선왕조에 이르기까지 살펴보면 다음과 같다.

삼국시대

삼국시대의 판단적 평가방법은 추천(referral)으로 국립교육기관에 입학 대상자를 추천하는 것과 관직에 추천(당시 용어로 천거)하는 방법이 사용되었다. 국립교육기관(고구려의 태학, 신라의 국학)이나 사학교육기관(고구려의 경당, 신라의 초기 화랑도)에 입학하기 위해서는

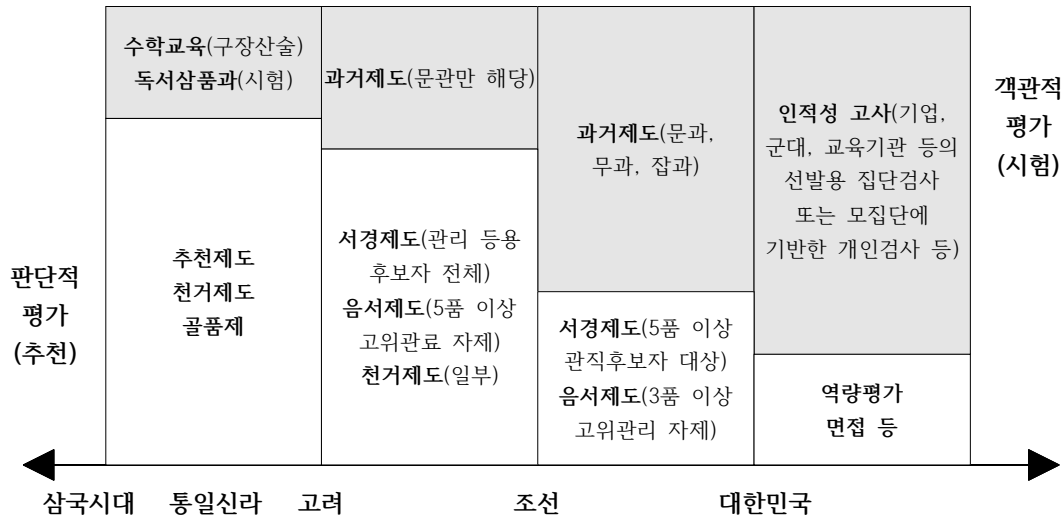


그림 1. 행동평가에서 두 경향의 비중 변화

일정한 사회계층, 성별의 제한(남성 위주)이 입학 추천하는 기준이 되었다. 시험제도가 있었다는 사료는 없으나 삼국사기에 따르면 고구려 유리왕 37년(18 BC)에 토지에 대해서 이미 경(頃)이라는 단위가 있어 농지의 측량이 있었고(김용운, 김용국, 2009, pp 77-106 참조), 국립교육기관인 태학에서는 신분예 상관없이 구장산술을 이용하여 수학을 가르쳤다. 구장산술은 총 9개 장에 걸쳐 각 18~46개의 문제를 이용해 수학적 계산법을 안내하는 책으로, 삼국사기에는 태학에서 9년간의 학업기간 동안 재질이 노둔하여 인재가 될 가능성이 없는 자는 퇴학시켰다고 기재되어 있어 학업과정에서 역량이나 성취도를 측정하기 위한 평가가 있었고 그 결과를 바탕으로 의사결정이 이루어졌음을 알 수 있다.

천거는 왕이 관직에 임명할 사람을 신하들로부터 추천받는 제도로 정기적인 요청이 있었던 것은 아니나 문관과 무관을 가리지 않고 역량과 가문이 좋은 자들을 추천받아 관직에

등용하였다(정구선, 1999). 천거제가 하나의 관리선발체제로 자리를 잡은 것은 한 대의 찰거제察舉制(선거제選舉制)를 시작으로 보는데, 삼국사기에 등장하기까지 전래된 방법과 시기에 대해서는 기록이 남아 있지 않다. 삼국사기에는 당시의 관리선발체도로 궁술(弓術), 기사(騎射), 전공(戰功), 음서(蔭叙) 등도 함께 활용된 것으로 기록이 남아있다. 음서는 천거와 함께 해설적 논평을 담은 추천제도의 하나로 음서가 고위직 관료의 집안과 배경을 중심으로 하는 추천방식인 반면, 천거는 모든 관등에게 평등하게 1인, 또는 최대 5인까지도 후보자를 추천받는 방식으로 진행하였다(정구선, 1998). 삼국시대에는 역량과 가문이 행동평가의 주된 기준이었다.

삼국사기에 따르면 천거의 사례들이 제시되어 있고, 우리가 잘 아는 고구려의 재상 을파소나 신라의 무장 관창은 모두 고관들이 추천하여 직위에 오른 예이다(정구선, 1999). 중요한 것은 이 당시에는 시험이나 검사가 있어

그 결과로 양적 평가를 한 것이 아니라 판단적 평가였지만, 오늘날 타당도를 논변(argument)으로 보는 관점(AERA, APA, NCME, 1999, 2014; Cronbach, 1988; Kane, 2006)에서 보면 추천받는 후보자가 나중에 교육, 행정, 군사 장면이라는 현실 맥락에서 어떻게 수행할 것인지 예상되는 결과(consequence)를 가늠할 수 있는 많은 정보를 수집하여 질적인 해석을 한 것이다.

삼국시대의 판단적 평가를 요즈음의 논변적 타당화 관점에서 볼 때 명시적인 타당도 논변이 제시되었는지의 자료는 없으나, 당시 가문이나 공동체의 명망(reputation)과 번영을 담보로 하는 추천이었기에 후보자에 대한 정보를 놓고 매우 신중하게 해석적 논변(예: Kane, 2006에서 이야기하는 많은 정보의 질적 해석, 예상되는 결과에 기반한 의사결정)이 있었을 것임을 알 수 있다. 예로서 고구려나 신라의 초기에는 귀족들의 회의에서 능력과 수행이 뛰어난 사람을 왕으로 선출하였고(이기백, 1998), 왕권이 강화된 후에는 귀족사회에서 문벌중심으로(예: 신라의 골품제도) 추천이 이루어졌다. 이것은 그림 1에서 맨 좌측의 상황이었다. 국가의 규모가 작을수록 상대적으로 적은 수의 관리들만으로도 국가 운영이 가능했기 때문에, 왕권이 안정된 직후의 초기 삼국은 문벌중심의 제한된 인재등용만으로도 충분히 숙고를 거쳐 필요한 인재를 선발하는 절차를 운영할 수 있었다. 그러나 이것은 일부 귀족세력만이 추천권을 가진다는 문제를 낳게 되었고 삼국의 인구가 썩이고 국가가 팽창하는 시기인 통일신라로 넘어가면서 개선의 필요성이 대두되었다.

통일신라

추천제도로서의 천거는 통일신라(688-935)에서도 계속되었으나, 이제는 문벌이 독점하는 인재공급을 벗어나 통일된 국토의 다양한 계층에서 인재를 구할 필요가 제기되었다. 그에 따라 과거제도의 전사인 독서삼품과라는 시험제도(원성왕 4년, 788)가 실시되었고 이것이 한국사에서 시험제도의 효시라고 할 수 있다. 이 제도는 이전 시대에 귀족들의 추천과 영향력으로 관료가 되는 관행을 차단하기 위해서 실시된 것이지만 이미 강하게 유지되던 골품제(bone-rank system)를 극복하지는 못하였다(한 국민족문화대백과사전, '독서삼품과'). 즉, 통일 이전의 성골/진골 구분이 통합된 진골이라는 신분과 남성에만 국한된 추천이 행동평가의 주축을 이루었다. 유교적 학술과 사회규범에 대한 학습의 기회가 많았던 신분계층과 남성에 대한 사회적 선택이 천거라는 추천제도를 강력하게 유지하도록 했던 것이다. 그림1에서 볼 때 통일신라는 아직은 판단적 평가가 주가 되는 체제였다.

고려

고려 때(918-1392)부터는 왕이 지방 호족의 세력을 누르고 중앙집권을 강화하기 위한 목적에서 객관적인 시험제도로서 과거를 본격적으로 실시하게 되었다(광종, 958). 그러나 삼국 시대 및 통일신라시대의 추천제도인 천거가 일부 남아있었고(고려사, 권75, 지29, 선거3, 천거), 천거적 요소가 포함되어 귀족 자손들에게만 해당되는 음서제도가 본격적으로 도입되어 왕족이나 공신을 포함하여 5품관 이상 고위관료들의 자식들만이 이 제도에 의해 관직에 추천을 받을 수 있었다(고려사, 권75, 지29, 선거3, 음서).

한편 이때의 과거에서 시험 내용은 문관에만 해당되는 제술과(문예시험), 명경과(경전시험), 잡과(기술관 시험)가 있었다. 한 가지 특기할 것은 시험결과로 “성적”에 따라 합격이 결정되지만 막상 관리로 임용되는 단계에 가서는 대신들에 의한 “판단” 즉, 동의를 받는 서경제도가 있었다. 과거시험 점수만으로 관리임용을 보장받는 것은 아니고, 좀 더 현실 맥락에 비추어 집단에 의한 평가가 이루어지는 부분이다. 음서와는 다른 또 하나의 판단적 평가체제로 볼 수 있다. 고려 때는 관리등용 후보자 “전체”가 서경제도 적용의 대상이었고 객관적인 시험결과를 집단의 판단에 의해서 뒤집을 수 있는 공식적인 제도였으며 이는 조선시대까지 이어졌다. 그림1에서 고려는 비록 음서제도와 천거제도가 시행되고 있었지만 판단적 평가(음서, 서경)와 객관적 평가(과거)가 비교적 균형 잡힌 평가의 시대로 볼 수 있다.

조선

마지막 왕조시대인 조선왕조(1392-1910)에서 과거제도는 더욱 확대되어 문관은 물론 무관 선발에도 적용이 되었다. 조선시대의 과거제도는 고려 때까지 강력한 역할을 한 음서제도의 위력을 누르고 공식적인 선발체계로서 위상이 높아졌고, 그 결과로 조선은 “문무의 관직을 차지할 수 있는 사회적 신분층”(이기백, 1998, p.196)으로서의 양반들이 이끌어가는 사회가 되었다. 그에 따라 추천제도인 음서제도는 3품관 이상만이 추천권을 갖도록 제한되어 고려 때와 달리 보조적인 제도로 그 역할이 제한되었다. 특히 신분보다 능력을 중시하는 조선사회에서는 음서가 천시되어 이 추천제도에 의해 요직에 진출하기는 거의 불가능했고,

결국 과거제에 의한 객관적 평가가 우세한 사회가 되었다.

조선시대 과거의 내용을 보면 문과, 무과, 잡과가 있어서 문/무과는 양반들을 대상으로 하였고 잡과는 서얼, 중인, 일반 상민을 대상으로 하였다(통역, 음악, 의료, 법률 분야). 대체로 정기적으로 실시되었으나 나라에 경사가 있거나 왕의 결정이 있으면 특별한 실시가 가능하였다. 한편 장애인도 과거에 급제하고 정신지체자도 종1품에 올랐으며, 시각장애인에게 현송(거문고를 타며 시를 읊는 관직)과 기우제를 관장하게 하고, 척추장애인으로 좌의정과 우의정을 지낸 허조, 간질장애인으로 우의정을 지낸 권균, 지체장애인으로 좌의정을 지낸 심희수 등 능력이 있다면 장애인에게도 차별이 없음을 보여주는 적극적인 정책을 시행하였다(김동욱, 1996).

고려 때와 같이 서경제도가 있어 5품 이상의 관직 후보자들에 대하여 실시되었으므로 그 범위가 축소되었고, 과거시험 점수가 평가에서 더 큰 역할을 하였다. 그림 1에서 현대 한국 이전의 조선시대의 위치는 음서제도가 약화되고 서경의 범위도 축소되어 판단적 평가의 비중이 약화되고, 객관적 평가가 강화된 것을 보여준다. 즉, 과거시험에서의 점수가 평가의 중심이 되고, 현실 맥락이 반영되는 질적 해석은 상대적으로 위축되었다. 이 시험제도는 당시 양반가의 젊은이들에게 거의 유일하게 출세의 길이 되었으나, 조선 후기에 많은 부정부패의 여파로 인하여 본래의 의미를 잃었고(승정원일기, 고종 10년 5월 7일) 고종 31년(1894) 일본의 강압으로 실시한 갑오경장 이후 폐지되었다(한민족독립운동사, 1권 국권수호운동1, 「동학혁명운동과 갑오개혁」).

초기 위원회 방식 평가의 특징

현대에 와서 위원회 방식의 평가(평가센터; assessment center)는 2차 대전 때부터 출발하여 현재의 사기업과 공조직에 걸쳐 광범위하게 사용되고 있는 집단 평가방식이다. 2000년대부터는 국내에서도 사기업은 물론 고위공무원단 평가에서도 이 방식이 널리 사용되고 있다. 일반적으로 평가센터 방식의 집단평가는 평가자들이 다른 자료를 보지 않고 피평가자의 행동만을 평가대상으로 삼는다. 하지만 이 방식을 미국에서 2차 대전 시에 최초로 도입한 전략업무국(Office of Strategic Services)은 평가의 모든 단계에서 모든 종류의 정보를 함께 참조하면서 전반적인 평가를 실시하였다(Thornton, 1992). 특이하게도 한국 역사에서 고려와 조선 시대의 서경제도 역시 시험에 합격한 사람들을 왕이 관리로 임명하기 전 대신들이 시험점수를 포함한 모든 정보를 참조하여 가부를 제시하였다. 이는 집단평가체로서 오늘날의 평가센터 보다 훨씬 앞서 실시된 한국 고유의 집단평가로, 당시 주된 평가체제로 정착되어 가던 과거가 객관적 평가인데 반하여 서경제도는 중신들이 피평가자에 대한 많은 정보를 질적으로 해석하여 임용이 정당한지의 논변을 제공하는 중요한 판단적 평가체계였다. 이렇듯 초기의 위원회 방식 평가인 미국 전략업무국의 그것과 우리의 서경제도는 모두 피평가자의 가용한 정보를 전부 활용한다는 것과 전반적인 질적평가를 한다는 측면에서 유사점을 갖고 있다.

판단적 평가와 객관적 평가의 비교

그림 1을 보면 우리문화에서의 측정과 평가는 판단적 평가 중심에서 객관적 평가 중심으로 변화해 왔다. 삼국시대와 통일신라시대에

는 천거라는 추천제도가 주가 되었으나, 고려 때는 귀족 자제 일부에 대한 음서제도와 모든 관직 후보자에게 적용된 서경제도라는 판단적 평가가 있었고 과거제도라는 객관적 평가가 있어 균형 잡힌 시기였다. 조선시대에 와서 과거가 출세의 거의 유일한 길이 되면서 객관적 평가는 평가의 주요한 흐름으로 자리를 잡았다.

주관성과 불안정성을 띠기 쉬운 판단적 평가에 비해 객관적 평가는 다수에 의해 합의가 가능한 평가라는 점에서 공정성을 높일 것이라는 일반적 기대를 갖게 한다. 하지만 반드시 객관적 평가가 옳고 그름에 대한 보다 정확한 판단을 내리도록 하는 것에 대해서는 내용 타당도에 대한 전제가 필요하며 이는 어떤 방법을 써서 평가하는 가와는 필연성이 없다. 평가의 타당도 관점에서 살펴보면 과거제도의 원래 목적은 훌륭한 관리가 될 사람을 평가하는 것이었으나 문관들에 대한 과거시험의 주된 내용은 경전해석과 시문짓기였으므로 문학이나 철학자적 자질, 그 중에서도 인지적 능력 중심의 평가였으니 관리로서 필요한 투명함, 원만함, 공정한 태도, 협상능력 등과는 크게 관련이 없는 시험이었다. 오늘날 문학가나 철학자를 정치인이나 관료와 동일시하지는 않는다. 그러나 과거제도의 특성상 그러한 동일시, 즉 내용타당도가 결여되었던 시대가 있었다. 그러한 타당도가 부족한 과거제도가 빚어낸 비극 중의 하나는 송강 정철의 예시에서 찾을 수 있다. 가사 문학의 3대 거장이자 충신으로 이름이 높았던 그가 정여립 사건시 기축옥사(1589)의 위관으로서 향후 3년여에 걸쳐 임진왜란이 터지기까지 무려 1,000여명의 상대와 선비들을 처형하는 냉혹한 면모를 보였다. 서인에 속하는 그가 반대측의 동인측 인사를

개인적인 원한에 의해 무자비하게 숙청함으로써 멸족의 화를 당한 가문들의 원한을 샀고 이후 300년 간의 당쟁의 불씨를 당기는 역할을 하였다. 당시 그가 행한 숙청과 그로인해 파생된 결과들로부터 그가 정치인 또는 관리로서의 재목으로는 적합하지 않았다고 보는 것(조선왕조실록, 1594년)이 오늘날 일반적인 해석이다(역사저널 그날, KBS1, 2014). 즉, 문학적 소질을 위주로 역량을 평가한 당시의 과거시험이 타당도를 결여함으로써 얼마나 크고 잘못된 결정을 하게 만들 수 있는지 극단적인 사례를 보여주었다.

그에 반해 추천제도(천거, 음서)나 집단평가(서경제도)와 같은 판단적 평가는 오히려 보다 많은 정보를 고려하여 질적 해석을 통해 결정을 내린다는 점에서 고대국가처럼 규모가 작은 공동체에서는, 추천될 사람이 실제로 수행을 잘 할 것인가에 대하여 충분히 타당한 의사결정을 내리게 할 수 있는 방식이다. 당시의 의사결정은 대부분 국가차원의 직위나 인제에 대한 것이므로 가문이나 공동체의 운명에 직결되어, 피추천자가 현실적으로 적절한 수행을 보일 것을 담보하지 않을 수 없는 상황이다. 예로서, 고려 때는 천거된 자가 역량이 없으면 추천한 자를 벌주라는 왕의 명이 있었다(고려사, 권75, 지 29, 선거3, 천거). 이것은 최근 타당도의 논의가 논변적 접근으로 가고 있음을 참조할 때 현실맥락과의 관련성을 강조하는 왕의 정책이었다. 즉, 단순히 점수에 대한 해석을 넘어, 그 해석의 결과 또는 사용이 목적인 결과를 가져오는지의 증거를 제시하는 논변적 접근(argument-based approach) (Kane, 2006)을 적용하던 시대였다.

당시에 현대적 의미의 타당도 논의가 있었다고 보기는 어려우나, 가용한 모든 정보를

활용하여 추천 후 예상되는 결과에 대해 질적인 해석과 예측을 한 것으로 볼 수 있다. 이것은 현실맥락에 충분히 연결된 논변을 통해 개별적 의사결정을 한 것으로서, 시험이라는 객관적으로 보이는 평가체제에서는 확보하기 힘든 질적 의사결정을 할 수 있는 장점이 있다. 즉, 타당도의 논변적 관점에서 볼 때, “질적 평가는 수행에 영향을 미치는 요인들 간의 상호작용에 주목함으로써, 그러한 상호작용이 현실 규범이 되는 환경에서 보다 풍부한 해석을 제공할 수 있다”(Kane, 2006, p.47). 즉, 학교 장면에서도 다양한 요인들 간의 상호작용을 원소로 하는 역량(예: 리더십)을 평가하거나, 학교 장면을 벗어나 실생활 장면으로 갈수록 질적 평가가 매우 중요해진다. 따라서 교육기관에의 추천, 관직에의 천거, 음서, 서경과 같은 판단적 평가체계가 시험제도보다 덜 타당하다고 하기 어려운 측면이 있고, 왕조시대의 맥락에 비추어 볼 때 질적 해석이 기여한 부분에 대해서도 충분히 인식해야 할 필요가 있다. 반면 타당도의 논변적 관점에서 보면 시험제도의 운영이 점수지상주의로 치닫게 될 때, 평가의 결과가 현실에서 어떻게 나타날지가 고려되지 않아 타당도 없는 평가로 전략할 위험을 내포하게 된다. 조선시대 과거가 유일한 입신양명의 도구가 되었던 시기와 90년대 본고사 및 수능시험에 의해서만 대학진학이 가능했을 때 낮은 시험점수가 사회적 진출을 막는 사회적 문제로 대두되었던 현상들은 이러한 문제점을 명확히 보여준다. 최근 몇 년 간에 대학사회에서 입학사정관 제도는 그러한 결함을 보완하기 위한 대안으로서 대두된 것이다.

연구방법의 역사

한국사의 초기시대로 갈수록 학문은 유학으로 집중이 되고 연구방법론 역시 해석적이거나 논리비평적이 된다. 이러한 질적 방법은 국내에 현대심리학이 유입된 이후에도 맥을 이어왔으며(예: 김성태, 1989; 임능빈, 1983; 조궁호, 1990, 1997, 1998; 한덕웅, 1994), 최근 매년 8월 심리학회 때마다 질적 연구마당이 지난 수년간을 이어져 오면서 질적 연구의 분위기를 높여가고 있다. 한국 역사에서 질적 연구방식이 큰 비중을 차지하긴 하였으나 이미 언급된 수학의 발전 및 시험제도는 양적 연구의 큰 흐름을 도입했다고 볼 수 있다.

국내의 연구를 질적인 방법과 양적인 방법에 의한 것으로 구분할 때 질적 연구방법에는 초기의 경전 해석적 연구와 논리 비평적 연구가 있고 현대적인 방법인 근거이론, 현상학적 접근, 문화 기술적 접근, 사례연구, 내러티브 연구가 있다. 그리고 양적 연구방법에는 크게 실험과 조사의 두 분류가 있는데 조사의 경우 현대에 와서 융성한 경향이 있지만 조선시대에도 이미 실시된 기록이 있다.

한국 역사에서 조선시대 세종입금 때(1430) 전국의 17만 2,648명에게 여론조사가 실시되었는데, 조사관을 각 고을에 파견하여 새로이 세금 걷는 방법에 대하여 백성들의 가부 의견을 청취하였다(조선왕조실록, 1454). 이것은 현대화된 방식의 과학적 접근은 아니었으나 대규모의 설문조사로 볼 수 있고, 서구 역사에서 최초의 여론조사로 기록되는 1824년 미국 대통령 선거의 경우보다 400년이나 앞선 것이었다. 또한 선조 때 세자인 광해군이 분조의 군졸들에게 권점(점찍어 찬/반 실시)을 통하여 주둔지의 결정에 활용한 것은 일종의 공개투

표를 실시한 것이었다(조선왕조실록, 1616). 조사는 해방 후 정치적 의견과 사회 여론을 파악하려는 필요에 의해 평가도구로서 널리 사용되기 시작하였으며 최근에는 마케팅을 위한 여론조사 분야가 크게 발전하고 있다.

서구 심리학의 유입기

조선왕조 말기에 서구의 학문이 유입되면서 행동과학 강의가 시작되었고(1908, 이화학당), 그리고서 38년 후 한국심리학회(1946, 당시 명칭은 조선심리학회)가 출발하였으며 1953년에는 행동과학의 또 하나의 축으로서 한국교육학회(1953)가 창립되었다. 이제 100년이 넘는 국내 행동과학의 역사에서 현대적인 행동평가의 역사는 상대적으로 짧은 편이다. 비록 1908년에 최초의 심리학 강좌가 개설되었으나 일제강점기에 한국인이 남긴 심리학 분야의 업적은 희소하다(한국심리학회, 1996). 해방이 되고 미군정 교육고문이었다던 한국인에 의해서 1947년 처음으로 통계학이 심리학도들에게 가르쳐졌고 한국전쟁(1950-1953) 기간 중에야 행동과학자들은 심리검사를 개발하기 시작했다. 이후로 교육학자, 산업 및 조직 심리학자, 임상 및 상담 심리학자, 그리고 발달심리학자들이 많은 교육검사 및 심리검사를 개발하였다.

기존의 전통인 판단적 평가와 시험에 더하여 심리검사를 포함한 행동평가의 현대적 전통은 통계학과 함께 서구로부터 수입되어 측정평가의 현주소를 구성하는 부분이 되었다. 결과로 왕조시대로부터 물려받은 교육장면, 관리 채용선발 및 인사고과 장면에서의 행동평가를 측정 및 검사이론, 통계, 도구(개발) 및 평가로 분화 발전시켜 주었다. 또한 서구 학문의 체제가 유입되면서 연구의 방법론도 전

통적인 유학 연구의 질적 방법론(경전해석, 논리비평적 방법)에 추가하여 새로운 질적 방법론들이 도입되었다.

해방 후 측정평가의 양적 발전

행동과학에서 현대적 측정평가는 해방이후 도입되었다고 할 수 있다. 측정평가는 상당부분이 양적 영역으로 보이지만 삼국시대 또는 그 이전의 기원을 찾을 수가 있었고, 그림 1에서 보듯이 출발은 추천을 중심으로 한 판단적 평가에서 하였으나 시험제도의 도입과 더불어 객관적 평가의 길을 추구해 왔다. 그러나 조선 후기의 세도정치로 인하여 객관적 평가의 흐름이 흐려지고 외세에 의한 강점기를 거치는 동안 한국사의 여타 부분과 같이 침체할 수밖에 없었다. 비록 일제가 실시한 고등 문관시험과 같은 평가제도가 있었으나 우리 문화에서 그 실시의 주체를 담당하지 않았으므로 접근이 어려운 상태이다.

해방 이후 측정평가의 학술활동은 표 1에 통계자료가 제시되었고, 그것을 요약하기 위하여 그림 2, 3, 4가 제시되었다. 각 기간의 구분은 이미 교육학에서의 측정평가 분야를 정리한 강상진(2004)의 구분을 따른 것이다. 일제 강점의 어두운 그림자는 건국 후에도 상당기간 드리워져 대한민국 수립 후에도 15년 동안(도입기)은 측정평가가 거의 불모지와 다름없었다. 학회, 학술지, 학위논문이 거의 없던 시대이다. 60년대부터 20여 년간 기반이 구축(학문적 기반 구축기)되는 가운데 도구(개발)/평가가 약간의 강세를 보였다. 그 후 80년대 10년간 비약적인 증가를 보였다. 1960년부터 시작된 30여년의 군사정권이 막을 내리고 문민정부가 들어선 20세기의 마지막 10년대에

는 측정 및 검사이론/통계/평가/연구방법의 각 분야별 학회지와 저서를 통한 전문성 있는 연구와 학술활동(전문화 시대)이 기록되었다(표 1 참조). 21세기에 들어와서는 첫 14년간에 측정평가의 학술활동은 비약적인 발전을 하였고 표 1의 맨 아래 전반적 통계를 보면 총 3,595건의 연구 가운데 도구/평가가 2,467건, 측정/검사이론이 567건, 통계가 313건, 연구방법이 248건이다. 도구개발 및 평가 분야가 단연코 비약적인 연구실적을 보이고 있음은, 실용적 성격이 강한 영역임을 감안할 때 놀랄 일이 아니다. 그에 수반하여 검사판매 및 관련 컨설팅이 산업으로(33개 회사) 발전하고 있다(부록1). 즉, 측정평가의 산업화가 시작된 시기이다.

21세기에 와서 국내의 측정 및 검사이론과 도구/평가의 눈부신 발전에도 불구하고 검사 산업은 아직 통계청에서 한 분야로 인식하고 통계를 파악할만한 규모가 되지 않는다. 그래서 이 연구에서는 검사산업의 발전을 고무하는 차원에서 한국 심리학회 내 분과들(현재는 임상, 산업/조직)에 등록되어 있는 심리검사 기관들을 부록에 소개하였다. 그에 반해 마케팅 관련 기업활동에 기여하는 조사 회사들은 측정평가의 하위영역중 연구방법(“조사” 포함)에 관련된 실제 적용 영역으로서 이미 통계청에서 통계 파악의 대상이 되어 2013년 기준으로 411개 회사에서 8,184여 명이 종사하는 규모를 보이고 있다(통계청, 2013).

그림 2를 보면 가장 큰 양적 발전의 분야는 도구(개발) 및 평가의 영역이고, 다음은 측정/검사 이론, 통계, 연구방법론의 순서이다. 도구(개발) 및 평가가 다른 영역과는 현저히 다르게 큰 확장을 보이는 것은 그만큼 사회에서의 수요가 증가하고 서비스에 대한 시장이 형

표 1. 현대 한국에서 측정평가의 학술활동 요약

		학술지	지역서	논문집	학위논문	소계
1945 -1960	측정/검사 이론	0	0	0	3	3
	통계	0	4	0	0	4
	도구/평가	0	2	0	0	2
	연구방법	0	0	0	0	0
	소계	0	6	0	3	9
1961 -1980	측정/검사 이론	7	10	1	0	18
	통계	1	13	0	0	14
	도구/평가	28	13	3	0	44
	연구방법	5	1	0	0	6
	소계	41	37	4	0	82
1981 -1990	측정/검사 이론	25	9	0	4	38
	통계	3	24	1	0	29
	도구/평가	75	13	3	7	98
	연구방법	6	9	1	1	17
	소계	109	55	5	12	182
1991 -2000	측정/검사 이론	112	10	12	15	149
	통계	31	28	6	8	73
	도구/평가	299	52	79	8	438
	연구방법	19	22	2	0	43
	소계	461	112	99	31	703
2001 -2014	측정/검사 이론	263	25	14	57	359
	통계	74	59	13	48	194
	도구/평가	1,453	120	272	53	1,898
	연구방법	104	48	28	7	187
	소계	1,894	252	327	165	2,638
Total	측정/검사 이론	407	54	27	79	567
	통계	109	128	20	56	313
	도구/평가	1,842	200	357	68	2,467
	연구방법	129	80	31	8	248
	소계	2,487	462	435	211	3,595

출처. 학술지와 논문집: 한국심리학회 DB, 한국 교육학술정보원, 한국 학술정보원, 강상진(2004)

지역서: 국립중앙도서관+성균관대도서관 DB, 강상진(2004)

학위논문: 국회도서관 DB, 강상진(2004)

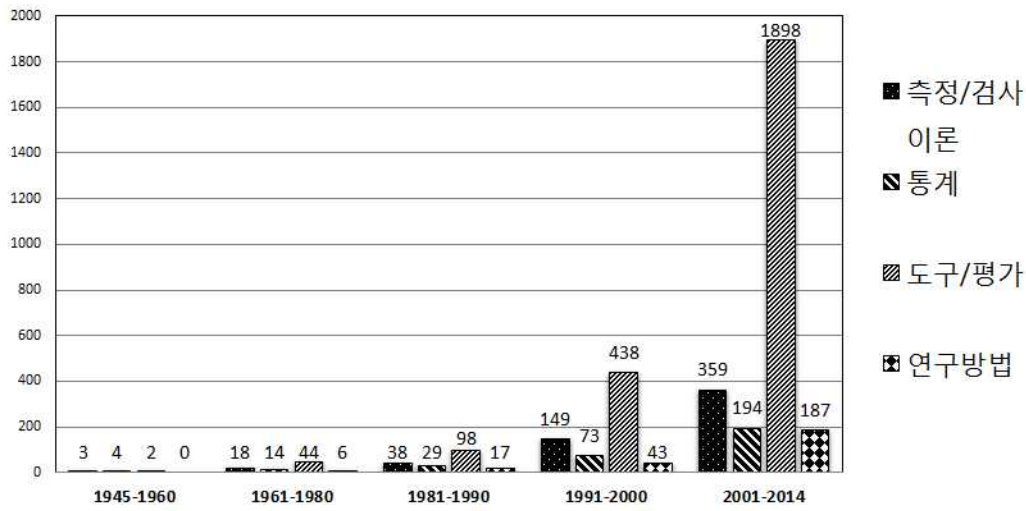


그림 2. 각 기간별 하위영역의 학술활동 비교

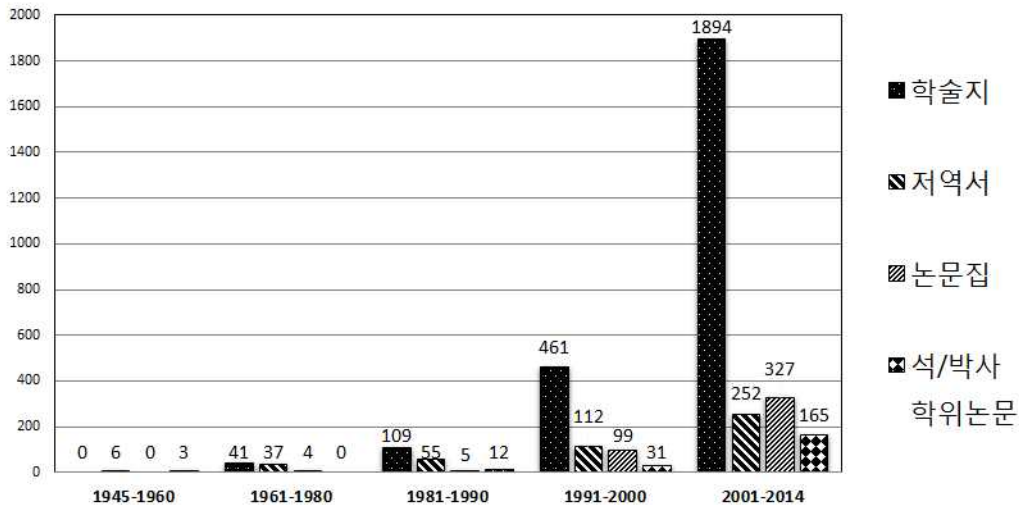


그림 3. 각 기간별 학술활동 창구의 성장 비교

성되고 있음을 의미한다. 실용영역인 도구/평가를 뒷받침하는 측정/검사 이론과 통계는 상대적으로 절대량은 적으나 최근에 전기 대비 2.5배 전후한 양적 성장을 보였다.

측정평가의 넓은 저변을 구성하는 연구방법 분야 역시 양적 규모는 적으나 최근에 전기

대비 4배 이상의 양적 성장을 하였다.

그림 3은 학술활동의 창구가 연도별로 성장하는 모습을 보여주고 있다. 창구로서의 출발은 학술지가 없던 시절에 저역서가 더 빨랐으나 1961년부터는 계속해서 학술지가 저역서를 앞질러서 학술지의 기능이 정착된 것을 볼 수

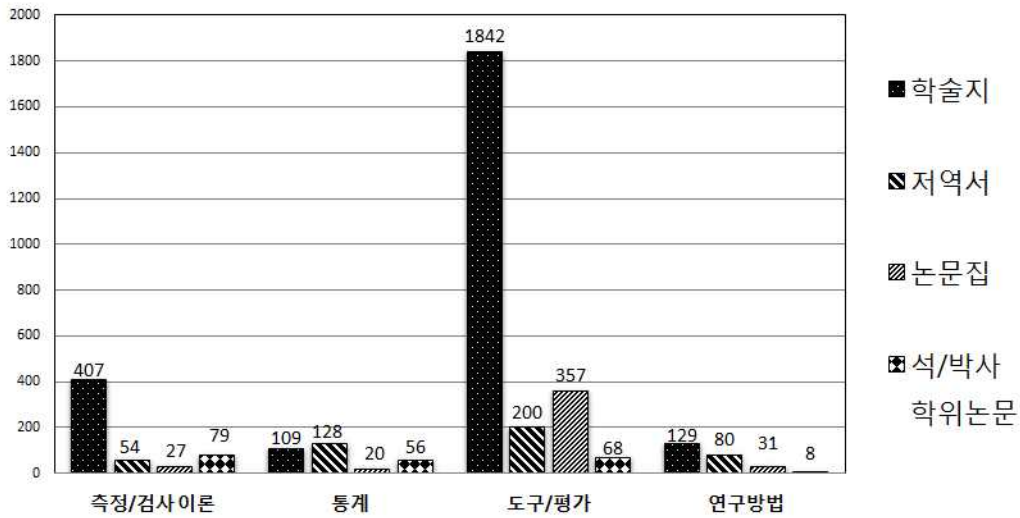


그림 4. 각 하위영역별 청구 선호 비교

있다. 학회활동도 활발해짐에 따라 학술논문집도 90년대부터는 현저하게 성장하여 지금은 학술지에 버금가는 청구로 활용되고 있음을 알 수 있다. 21세기에 들어서 첫 14년간에는 학술지가 현저하게(1,894건) 성장하였고 가장 열세인 학위논문도 최근에 전기 대비 5배 이상의 성장을 하였다. 그러나 14년간에 165건이므로 1년에 12명이 안되는 석박사를 배출하는 것이 측정평가의 열악한 학문기반이다. 그럼에도 연간 135건 이상의 학술논문이 생산되는 것은 연구자들이 얼마나 열정적인 학술활동을 하고 있는지를 짐작케 한다.

끝으로 그림 4를 보면 해방이후부터 현재까지 통계를 제외한 모든 하위영역에서 학술지를 통한 활동이 가장 높다. 그리고 도구/평가를 제외하면 나머지 세 분야 모두 학회발표 논문집이 거의 활용되지 않고 있음은 측정평가 분야의 학회활동이 아직은 충분히 활성화되어 있지 않음을 의미한다. 또한 석박사 학위논문 주제로서 다른 세 분야는 고르게 많은

데 비해서 연구방법을 선택하는 경우가 가장 적고 절대적인 숫자 역시 매우 적어서 앞으로는 이 분야의 주제 선택을 고무할 필요가 있다. 학위논문 중 측정/검사 이론에 대한 논문이 가장 많은 것은 측정 및 평가에 관한 이론적 논의에 대해 관심이 높은 국내 학술적 분위기를 반영한다.

논 의

앞서 한국사에서 각 시대별 선발제도의 양상을 비교 검토함으로써 객관적 평가와 해석적 평가의 비중변화와 그 맥락을 살펴보았다. 판단적 평가로부터 객관적 평가로의 전환에서 유교는 나라의 통치이념으로서 시대적 과제에 따라 변화와 통합을 이끌어 내는 국가의 신념체계를 제공하였다. 특히 국가 경영을 위한 인재를 선발함에 있어서 그러한 유교적 가치를 내면화하고 행동으로 표현하는 인적 속성

을 평가하고 관리로 등용함으로써 국력을 강화하는 정책으로 활용하기 위함이었다. 또한 수학의 전래와 발달, 초기 과거제도를 참고한 독서삼품과 등의 적용은 양적평가체계 수립과 적절한 방법론을 제공함으로써 일관된 신념체계가 사회적 요구에 따라 유연하게 변화하고 적응하도록 돕는 역할을 하였다. 이러한 맥락에서 각 시대로부터 주어진 사회적 요구와 그에 상응하는 적응적 변화가 시사하는 점은 여전히 검토할 가치가 있다.

오래된 문제와 새로운 기술

판단적 평가를 강조하는 타당도 논변

먼저 국내 행동과학의 과거와 현재를 연속성 있게 검토하면서 발견한 것 중 가장 큰 성과는 도구/평가의 하위영역에서 과거와 현재 간에 극명한 대비가 있다는 것이다. 그림 1에서 보듯이 과거로 갈수록 전반적으로 판단적 평가가 있다면, 현대에 올수록 시험제도의 도입과 더불어 객관적 평가가 큰 비중을 차지하고 있다. 이러한 변화가 올바른 방향으로의 발전을 뜻하는지, 즉 타당도를 높이는데 있어 충분히 긍정적인 방향으로 작용한 것인지에 대해 현대의 타당도/타당화 이론으로 검토해 볼 필요가 있다.

타당도를 “검사의 사용을 위해 제시되는 점수해석을 증거나 지식이 지지하는 정도”(AERA, APA, NCME, 1999, p.9; AERA, APA, NCME, 2014, .11)로 정의한다. 이에 따라 평가의 타당화는 평가에서 얻은 점수를 어떤 실제 목적에 사용하는 것이 정당화되는 정도라고 하는 논변적 관점이 최근의 추세이다(AERA, APA, & NCME, 1999, 2014; Cronbach, 1988, Kane, 2006). 따라서 논변적 관점의 타당화를

위해서는 “측정하고자 하는 것을 측정하는 정도”(Allen & Yen, 1979, p.95)를 넘어서 점수의 해석이 정당한지에 대한 평가를 해야 한다(Cronbach, 1988). 그러기 위해서 필요한 것이 해석적(interpretive) 논변과 타당도(validity) 논변이다. 해석적 논변은 평가의 사용 및 그를 위해 제안된 해석을 명세화하기 위해 수행의 관찰-채점-해석(예: 결론, 의사결정)을 제시하는 것이고, 타당도 논변은 해석적 논변이 응집력 있고(coherent), 추론이 합리적이며(reasonable), 가정이 적절하다(plausible)는 “주장”(claim)이다(Kane, 2006, p.23 참조).

미국의 검사표준서 5판이라고 할 수 있는 1989년 표준서에서는 타당화에 대하여 해석적 논변까지 언급이 되었으나, 그 후에 발간된 최근의 두 표준서에서는 그것을 넘어 타당도 논변까지 포함할 것을 분명히 하고 있다. 즉, 1999년 표준서에서는 타당화를 “검사점수의 해석 및 그것이 검사의 사용에 대하여 가지는 관련성을 지지하는 과학적인 타당도 논변의 개발”(AERA, APA, & NCME, 1999, p.9)로 제시하였고, 가장 최근의 표준서에서는 “검사점수의 해석 및 그것이 검사사용에 대하여 가지는 관련성에 대한 찬반의 논변을 구축하고 평가하는 과정”(AERA, APA, & NCME, 2014, p.11)이라고 한 것이 바로 해석적 논변을 넘어 타당도 논변이 있어야 타당화라고 정의함을 의미한다.

물론 해석적 논변이 충실하지 않으면 타당도 논변이 강화될 수가 없다. 따라서 해석적 논변의 영역으로서 종래의 특질 해석, 이론기반 해석, 질적 해석, 그리고 의사결정용 해석이 중요해진다(Kane, 2006, pp.30-56). 과거 한국사에서 행동평가의 초기에 큰 비중을 가졌던 판단적 평가는 정보의 질적 해석과 의사결

정용 해석의 논변이라는 점에서 해석적 논변이 이루어졌다고 볼 수 있다. 그러나 시험제도가 도입되면서 오늘날에 이르기까지 강화된 객관적 평가는 불필요한 주관이 배제되었다는 장점은 있으나 표준화된 검사의 특성상 1점, 0.5점의 점수차이로 당사자의 운명을 결정하는데서 질적 해석이 지나치게 배제되는 문제가 있다. 대상의 풍부한 질적 속성을 비교하는 과정에서 축약된 숫자 값 간의 작은 차이가 질적 차이를 전적으로 대체하여 해석되는 것인데, 이러한 문제는 현대 한국에서 측정평가의 모든 영역에서 양적 방법의 편향 및 신화적 믿음으로 인하여 해석적 논변을 위축시키는 역할을 했다. 즉, “타당한 평가”라는 오래된 문제를 공략하는데 새로운 접근(검사 이론, 통계, 양적 평가, 양적 연구방법)에 대한 편향은 평가자가 검사의 목적에 근거하여 관찰한 속성, 또는 행동에 대해 평가하고 의미를 부여하는 채점과정 일체를 대상으로 하는 해석적 논변을 위축시킨다. 그로부터 검사의 사용에 대한 찬반의 논변을 구축하고 평가하는 타당도 논변을 억제하는 문제를 야기한다.

평가에서 현실맥락에 기반한 판단 필요

한국사에서 평가 역사의 초기에 많이 사용되었던 판단적 평가가 타당한지 여부는 그로 인한 결정이 좋은/나쁜 결과를 가져오느냐에 달려있으므로 논변적 관점에서 논의할 수 있다. 한국사에서 연맹왕국 시대의 초기에는 각 성읍국가의 장들이 선거를 하여 왕을 선출하였다. 이때의 평가는 선출이라는 의사결정을 수반하였고, 그 선출이 타당한 것이었는지 아닌지는 왕이 된 후 그 사회에 미친 결과를 보고 판단하였다. 예로서 “부여에서 흉년이 들면 그 책임을 왕에게 돌려 혹은 갈자느니 혹은

죽이자느니 했다”(이기백, 1998, p.46)는 사실은 판단적 평가의 시대에 현실에서의 수행에 관련성을 가지는 타당도의 개념은 오히려 더 치열했다고 볼 수 있다. 판단적 평가 시 판단의 잘못은 가족을 넘어 공동체의 운명에도 직결되었으므로 현실맥락이 올바른 평가를 요구하는 큰 압력으로 작용하였다. 그 후 그림 1에서와 같이 판단적 평가의 비중이 감소되고 점수중심의 평가가 되면서 피평가자의 수행이라는 맥락에의 관련성이 감소되는 평가가 되었다.

해방 후 일제 강점기의 시험제도를 모방한 고등고시가 생기고 각 급 학교의 입시, 기업체 입사전형, 군 장교들의 승진심사는 수많은 이해당사자들의 눈과 귀를 사로잡는 초미의 관심사가 되었다. 그러나 조선 때만 해도 현실맥락을 고려한 판단적 평가가 있었고 또한 시험점수에만 의존하지 않고 관리로서의 수행이라는 맥락에 비추어, 점수에 대한 질적인 해석을 함으로써(서경제도) 시험제도의 타당도를 높이는 절차가 있었다. 그러나 대한민국 건국 후 서구에서 도입된 양적 평가 기술은 이해관계자들의 다양한 불만이나 비평을 회피하고자 단 1점 또는 그보다 작은 점수 차이라도 높은 점수의 사람이 낮은 점수의 사람보다 더 유리한 대우를 받게 하는 사회적 분위기를 형성하였다. 즉, 이것은 결정의 결과에 대한 타당도를 고려하지 않는 결정으로, 1점이나 그 이하의 점수 차이가 과연 실제 입학, 임용, 승진 후 현실 맥락에서 수행의 우위로 나타날 것인지에 대한 충분한 고려 없이 눈감고 숫자에 지나치게 의존하거나 권위를 부여하는 관행을 출현시켰다.

공학에서도 제품의 디자인이 서비스를 받을 인간들에게 정말로 기여할 것인지의 맥락을

반영하여 공학의 실천이 이루어지는데(예: 인간공학, Woodson, Tillman, & Tillman, 1992), 수많은 주변사람, 동료, 근무자, 장병, 국민 일반이라는 맥락에서 점수 일변도의 결정이 어떤 결과를 가져올 것인지에 대한 검토는 당연히 강조되어야 한다. 그러나 우리의 현실은 그러한 가능한 많은 정보를 검토하고 판단함으로써 타당도를 향상시킬 수 있는 여지를 상당부분 생략한 채 평가제도를 운영하고 있다. 그러한 제도에서 유리한 결정을 받고 학교에서, 기업에서, 정부에서, 법조계에서, 군에서 활동하는 인재들이 과연 온전히 바람직한 사회적 결과를 보일 것인지에 관해 논변할 수 있는 여지가 없다.

국내의 평가제도에서 현실 맥락에의 관련성을 가지는 평가라고 주장할 수 있는 절차가 누락된 상태에서 사회적 결과는 바람직하기 어렵다. 학교에서는 학생을 육성하는 목표가 단지 많이 외우고 숙달하여 높은 점수를 산출하는 학생이 좋은 학생의 표본으로 제시되어, 실제에서의 좋은 자녀, 좋은 이웃집 아이, 심신이 건강한 아이가 되기 어려운 교육이 이루어지고 있다. 건전한 예비 사회인을 생산하는 것을 목적으로 하는 대학의 경우 입시에서 졸업에 이르기까지 인지능력을 충분히 평가하는 것에 비해 다른 측면의 적격성을 판단할 수 있는 절차는 미미하다. 즉 타당도의 패러다임이 점수중심 방식에 갇혀 한계를 명확히 드러내고 있다.

최근 국민적 공분을 사고 있는 법조인 사이의 전관예우에 의한 폐단이나, 브로커 변호사들에 의한 서민 갈취적 법률행위(한인섭, 1999) 등은 그들이 대학 졸업 후 적합한 사법시험을 통과한 적격성을 갖추었는가에 대해 의문을 갖게 한다. 건국 후 다년간 시행된 사

법고시가 온전히 바람직한 법조인을 양성했다고 주장하기에는 많은 실망스러운 법조인들이 마스크를 통해 알려지고 있다. 과거의 사법고시가 인지능력 위주의 시험이었다면, 바뀌었다고 하는 법학 전문대학원에 대해서는 입학에서 변호사 시험에 이르기까지 불투명하다는 비평이 끊이지 않고 있다(이성진, 2015). 현재 법학 전문대학원 입학을 위한 1단계(LEET, GPA, 공인영어 성적, 자기소개서, 학업계획서) 및 2단계(1단계의 총점, 논술, 면접)의 절차 가운데 면접만이 변호사라는 현실 맥락에 비추어 후보자의 적격성 판단을 필요로 하고 나머지 절차는 모두 인지능력으로 판가름이 나는 부분이다. 지원자의 점수를 해석할 때는 국민을 위한 법률 서비스 제공자로서의 자질을 평가하는 폭넓은 증거기반의 해석이 필요한데 현재의 절차는 이를 담보하기 어렵게 설계되어 있다.

반면 기업체에서의 선발평가는 인지능력이라는 최대수행능력과 성격, 동기, 태도 등 전형적 수행능력을 모두 평가대상으로 한다. 단순히 시험점수만이 아닌 열린 면접이나 추천을 활용하고 있으며, 경우에 따라서는 인턴십과 같이 장기간 후보자를 관찰할 수 있는 체계를 활용하는 경우도 있다. 관찰할 수 있는 장면이 다양해지고 장기적이 될수록 회사업무의 맥락에 보다 관련성이 높은 결정을 내릴 수가 있으므로 보다 현실 맥락에 관련성이 있는 평가라고 할 수 있다. 그에 비해 사법고시, 행정고시, 외무고시를 비롯한 각종 공직시험에서는 인지능력 중심의 점수에 대한 과도한 의존을 견제하고 평가에 필요한 보다 많은 질적 정보를 확보하여 현실맥락에 연계하는 논변적 접근이 매우 부족하다.

물론 기업들 역시 평가에서 논변적 접근이

충분한지에 대해서는 여전히 따져볼 여지가 있다. 대기업의 관리자들이 중소기업이나 협력기업에 자행하는 “갑질”, 기업들이 지역사회 공익을 외면하는 사회적 책임의 부재는 기업인이 될 수 있는 자격의 평가에서 중요한 절차가 누락되었음을 시사한다. 이윤추구를 목적으로 하는 기업인이 사회공익을 추구해야 하는가에 대한 부분은 의문일 수 있으나 사회구성원으로서 호혜적 분위기를 구성하는 것이 직간접적이고 장기적인 이윤을 보장한다는 측면에서 기업 윤리적 맥락을 형성하는 과정은 매우 중요한 것이다. 그리고 이러한 윤리적 맥락이 반영되는 평가가 충분히 이루어지지 못해서 나타나는 부적합한 기업활동의 예시들을 우리는 일상적인 기사를 통해 자주 접하고 있다.

평가과정을 통과한 사람들이 이후 활동하거나 업무를 수행하게 될 현실 맥락에 대한 고려가 부족한 경우는 군대 사회에서도 흔히 기사화 된다. 평가의 허술함이 잠재적으로 문제가 있는 사병의 군입대를 허용하기도 하고, 장교로서의 자질이 부족하여 문제 장교나 문제 장군이 되어 갖은 추문을 만들어 내기도 한다.

결국 평가는 단순히 점수를 산출해서 결정을 내리는 것이 아니고 평가를 통과한 사람들이 참여할 생활, 일, 그리고 업무라는 맥락을 감안하여 모든 가능한 관련 정보를 가지고 그 점수의 해석 및 사용 결과를 논의하는 절차까지 포함하여야 한다. 평가결과가 사용되는 맥락에의 관련성이 담보되지 않는 평가는 충분히 타당하다고 말하기 어렵다.

미국의 경우 고용 맥락에서의 작은 점수차이가 공정한 선발인가의 논쟁에 대하여 이미 오래 전에 충분히 고민하고 여러 가지 유동적

인 방식들(Cascio, Ourtz, Zedeck, & Goldstein, 1991)을 강구한 바 있다. 이에 대한 참조는 우리 사회의 점수지상주의를 개선하는데 현실적 대안을 제시해 줄 수 있다. 또한 국내의 많은 시험제도 운영에서 이해당사자들로부터 있을 수 있는 강력한 비평에 대처하기 위해 점수 해석 및 사용이 현실맥락에 비추어 정당함을 주장하고 설득할 수 있는 준비를 해야 할 것이다.

경험적 평가와 판단적 평가의 균형 필요

해석적 논변에도 판단적 평가, 즉 정성적이고 질적인 평가가 많이 포함되어 있으나 타당도 논변은 훨씬 더 많은 질적 평가를 필요로 하는 추세를 보여준다. 실무적인 측면에서 질적 평가만으로 조사가 마무리 되는 경우도 있고(예: 산업 장면에서의 직무분석), 또 한편으로는 질적평가는 주로 검사점수를 이용하는 양적평가를 보완하고 평가의 질을 높이는 중요한 역할을 한다. 평가 대상의 속성에는 정량화에 보다 용이한 측면이 있는 반면 그렇지 못한 요소가 함께 담겨 있는 경우가 많고 이는 평가의 정확성과 용이성을 놓고 저울질 할 때 환원주의 대 전체론적인 접근(Verschuren, 2001), 다중 패러다임론 대 변증법적 입장(김미숙, 2006) 등 철학적 배경에 대한 논의까지 확장될 수 있다. 양적평가를 통해 정확성을 기하려는 노력은 점수화가 용이하고 편향이나 오염이 적도록 구체적인 수준으로 국소화시키는 것에 집중하는 경향을 강화시킬 수 있다. 하지만 질적평가는 정확성과 용이성에 대한 고려에서 반드시 평가 대상이 되는 속성을 단일 차원까지 좁혀 나가야 할 필요는 없다. 특히 양적평가 지표는 수량화하기 쉬운 좁고 단일한 차원의 속성들 중에서 의사결정에 중요

한 측면만을 추출해서 평가가 이루어지도록 구성되지만, 양적평가로 다루기에 연구가 충분하지 않거나 영역이 모호한 것들은 질적으로 평가한다. 이후 경험적 수량자료에 기반한 양적평가 지표와 병행하여 전체 평가결과를 의사결정에 반영하는 방식으로 활용할 수 있고, 중요한 의사결정에서는 이러한 병행체계를 갖고 있는 경우를 어렵지 않게 찾을 수 있다(예를 들어, 기업 입사전형이나 대학의 입학 전형에서 지원자의 역량을 평가하는 체계 등). 단지 양적인 평가지표를 갖춘 경우조차 원래는 대상 속성에 대한 이해와 그 수준에 대한 질적인 판단을 통해 점수화된(Kline, 2012)는 점을 상기한다면 판단적 평가를 보다 진지하게 할 필요가 있다. 그냥 편하게 하기 쉽게 하는 것이 판단적 평가가 아니고, 지식을 탐구하는 자세로 해석과 논리비평을 사용하여 치열하게 실시해야 할 것이다. 최근 국내에서 한국심리학회내에 심리측정 및 평가 분과가 창립되고 양적연구와 질적연구의 균형을 추구하는 노력은 이러한 당위적 필요에 부응하는 긍정적인 흐름이라고 판단할 수 있다.

참고문헌

- 강상진 (2004). 반세기 한국교육평가학의 역사와 전망. 이화여자대학교 한국문화연구원 (편), *교육학 연구 50년*, 321-378. 서울: 휴안.
- 고려사. 국사편찬위원회, 한국사데이터베이스: 고려시대: <http://db.history.go.kr/KOREA/>
- 김동욱 (1996). *국역 동패략송*. 경기: 아세아문화사.
- 김미숙 (2006). 양적 방법과 질적 방법의 통합에 대하여. *교육사회학연구*, 16(3), 43-64.
- 김성태 (1989). *경과 주의*. 서울: 고려대학교 출판부.
- 김용운, 김용국 (2009). *한국수학사(수학의 창을 통해 본 한국인의 사상과 문화)*. 경기: 살림MATH.
- 김충렬 (1982). *동양 인성론의 서설*. 한국동양철학회 편, *동양철학의 본체론과 인성론*. 169-184. 서울: 연세대학교 출판부.
- 노중국 (2014). 삼국시대의 인문 정신과 세계관. *동서인문학*, 48, 25-50.
- 승정원일기. 국사편찬위원회, 한국사데이터베이스: *사료고종시대사 6*: <http://db.history.go.kr/KOREA/>
- 역사저널 그날. (2014). '정철, 기축옥사 특검 되던 날', KBS1, 시사/교양, 28회.
- 이광세 (1998). "근대화, 근대성, 그리고 유교" 동양과 서양: 두 지평선의 융합. 서울: 길, 64-93.
- 이기백 (1998). *한국사 신론*. 서울: 일조각.
- 이남희 (2008). 과거제도, 그 빛과 그늘. *오늘의 동양사상*, 18, 117-136.
- 이성진 (2015). 로스쿨, 사법시험 등 법조인력 양성에 대한 현업 변호사들의 생각. *법률저널*, 2015. 7. 14.
- 이승환 (1998). 후기 근대 유학담론의 두 유형: 뚜웨이밍(杜維明)과 에임스(Roger Ames)를 중심으로 <<동아연구>>(서강대학교 동아연구소). 35, 363-416.
- 이장주 (2014). *우리 역사속 수학이야기*. 서울: 성균관대학교 출판부.
- 임능빈 (1983). *성리학의 적응심리학적 접근*. 미간행 박사학위 논문. 서울대학교.
- 전덕제 (2010). 신라의 독서삼품과: 한국 과거제도의 전사. *한국사 시민강좌*, 46, 1-18.

- 정구선 (1999). 高麗時代의 現官薦擧制. *동국사학*, 33, 19-43.
- 조공호 (1990). 맹자에 나타난 심리학적 함의 (I): 인성론을 중심으로. *한국심리학회지: 사회*, 5(1), 59-81.
- 조공호 (1997). 선진유학의 심리학적 함의. *한국심리학회편, <동양심리학의 모색> 한국심리학회 1997년도 추계 심포지움 자료집*. 서울: 한국심리학회. 41-406.
- 조공호 (1998). 유학심리학: 맹자·순자편. *나남 신서* 648. 서울: 나남출판.
- 조공호 (2007). 동아시아 집단주의와 유학 사상: 그 관련성의 심리학적 탐색. *한국심리학회지: 사회 및 성격*, 21(4), 11-54.
- 조선왕조실록. 6대 단종, 1454년, 국사편찬위원회: <http://sillok.history.go.kr/>
- 조선왕조실록. 14대 선조, 1594년. 선조실록 57권, 선조 27년 11월 12일, 국사편찬위원회: <http://sillok.history.go.kr/>
- 조선왕조실록. 15대 광해군, 1616년, 국사편찬위원회: <http://sillok.history.go.kr/>
- 최몽룡 (2002). *신편 한국사*. 서울: 국사편찬위원회. <http://db.history.go.kr>, 2015. 4. 10 인출.
- 통계청 (2013). *전국 사업체 조사 (국가 통계포털 KOSIS의 온라인 통계자료)*. 서울: 대한민국 통계청.
- 한국민족문화대백과사전. 한국학중앙연구원: <https://encykorea.aks.ac.kr/>
- 한국심리학회 (1996). *한국심리학회 50년사*. 서울: 교육과학사.
- 한덕웅 (1994). *퇴계심리학*. 서울: 성균관대학교 출판부.
- 한민족독립운동사, 1권 국권수호운동1, 「동학혁명운동과 갑오개혁」: <http://db.history.go.kr/>
- 한영우 (2015). *다시 찾는 우리 역사*. 서울: 경세원.
- 한인섭 (1999). 법조비리-문제와 대안. *법학(서울대학교)*. 39(1), 164188. 서울: 서울대 법학연구소.
- AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington, D. C.: American Educational Research Association.
- AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Belmont: Wadsworth, Inc.
- Cascio, W. F., Outtz, J., Zedeck, S., & Goldstein, I. L. (1991). Statistical implications of six methods of test score use in personnel selection. *Human Performance*, 4(4), 233-264.
- Corbin, J. M. & Strauss, A. (2008). *Basics of qualitative research: techniques and procedures for developing grounded theory*. 3rd Ed. Thousand Oaks, CA: Sage.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 3-17). Hillsdale, NJ: LEA.
- Giorgi, A. (1985). *Phenomenology and psychological research*. Pittsburgh, PA: Duquesne U. Press.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.) *Educational measurement* (4th ed. pp. 17-64). NY: Council on Education and Praeger Publishers.
- Kline, R. B. (2012). *Beyond significance testing: statistics reform in the behavioral sciences*. 2nd Ed. American Psychological Association,

- Washington, DC.
- Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: LEA.
- Russell, B. (1959). 서양의 지혜: 그림과 함께 보는 서양철학사. *Wisdom of the West: A Historical Survey of Western Philosophy in its Social and Political Setting* (이명숙, 광강제역, 1990). 서울: 서광사.
- Thornton, G. C. (1992). *Assessment centers in human resource management*. NY, NY: Addison-Wesley Publishing.
- Verschuren, P. J. M. (2001). Holism versus Reductionism in Modern Social Science Research. *Quality & Quantity*, 35, 389-405.
- Woodson, W. E., Tillman, B., & Tillman, P. (1992). *Human factors design handbook: Information and guidelines for the design of systems, facilities, equipment, and products for human use*. 2nd Ed. NY, NY: McGraw-Hill.
- 1차원고접수 : 2019. 02. 03.
2차원고접수 : 2019. 04. 08.
3차원고접수 : 2019. 06. 13.
최종게재결정 : 2019. 08. 20.

Past and Present of Measurement and Assessment in Behavioral Sciences: A review of Korean history

Soonmook Lee

Sungkyunkwan Univ.

Ahyoung Kim

Ewha Womans Univ.

SeungGuen Baek

Ajou Univ.

Min-Kyeong Kang

Sungkyunkwan Univ.

We traced the root of behavioral measurement and assessment in Korean history in an attempt to locate where we are in the context of history and modern development of the field. We discovered that the balance between judgmental assessment and objective assessment has continuously moved from “judgmental” to “objective”, which is consistent with the developmental trajectory of the field in the western world. Through searching books, articles, and conference proceedings, we could build statistics about the modern development of the field and academic performance of scholars. We reviewed evaluation practices for the referral, recommendation, and selection of government posts in Korean history and found evidence of judgmental and objective assessment throughout the Old Chosun, three Kingdoms, Unified Silla, Koryo, and Chosun dynasties. Through the dramatic development in the last decade of 20C we have demonstrated tremendous achievement in the subdomain of instruments and assessment, and the subdomain of theory of measurement and testing. As a result we see more than 400 companies doing business in the field of psychological/social survey. Also we are close to witnessing the advent of testing industry in the nation. Lastly we raise some questions or point out some problems in current assessment practices in an attempt to pursue the next milestone on the route to full blooming days of behavioral measurement and assessment in Korea.

Keywords: measurement, assessment, subjective judgement, objective evaluation, Confucian value

부록. 검사제작 및 관련 컨설팅을 업으로 하는 기관 및 검사

도서출판 신경심리: 노인용 인지검사, 아동용 Rey-Kim 기억검사, 성인용 Rey-Kim 기억검사, 성인용 Rey-Kim 기억검사-II, 아동용 Kims 전두엽-관리기능 신경심리 검사, 성인용 Kims 전두엽-관리기능 신경심리 검사, 성인용 Kims 전두엽-관리기능 신경심리 검사-II, K-사병진단검사

마음사랑: MMPI, TCI(Temperament and Character Inventory)

색채도형심리: 색채도형심리 진단검사

아이진로: 성격검사, 정서검사, 학습검사, 진로검사, 인재상검사, 구직 지식 검사, 구직 기술 검사, 인재상 검사, 직업 적성 검사, 직군 적합 검사, 직무 적성검사, 통합인성검사

연우심리연구소: U&I 심리검사, 객관적 성격검사, 투사적 성격검사, 학습검사, 적성검사, 지능검사, 신경심리검사

(주) 찬스원: 키즈큐

중앙적성연구소: 유아지능검사, 그림지능검사, 학습집중력검사, 초등용 검사, 중등용 검사, 고등용 검사, 대학교 검사, 특수검사

테스트온: 학습클리닉검사(LETA), 표준화성격진단(SPI)

테스피아: 영아발달검사, 영아언어검사, 유아발달검사, 유아사회성기술검사, ADHD진단검사, 자폐증진단검사, 청소년진로적성검사, 학습진단검사

프로이드: 종합심리검사, 자기주도적학습검사, 아동/청소년 성격검사, 진로적성검사, 학교생활 적응검사, 다중지능검사

학지사 심리검사연구소: 지능인지능력검사, 창의성검사, 자폐증 진단검사, 신경심리 치매 진단 검사, 신경심리 장애진단검사, 인성불안 대인관계 검사

한국가이던스: 적성검사, 인성검사, 지능검사, 성격유형검사, 학습지도검사, 창의성검사, 신경심리검사, 강점검사

한국심리검사 테스트코리아: 일반지능검사, 수리지능검사, 문화공평식 지능검사, 아동지능검사

한국심리적성검사연구소: 적성 검사, 인성 검사, 학습 검사, 진로 검사, 창의성 검사, 다중지능 검사

한국에니어그램교육연구소: 성인용검사, 간편성격유형검사, 청소년용검사, 진로 및 학습유형검사, 아동용 검사, 아동-청소년용 간편검사, 성격하위유형검사

한국직업능력개발원 Career Net: 진로적성검사, 직업흥미검사, 직업가치관 검사, 진로성숙도 검사, 진로개발준비도 검사, 이공계 전공적합도 검사, 주요능력 효능감 검사

한국평가원: 직장/기업/조직, 유,초,중,고,대학생, 군,경 갈등관리, 부모/교사/상담자, 특사아학습 클리닉

한국학습능력개발원: 아동청소년 종합심리검사, 대학생 종합심리검사, 아동용 인성검사, 학습 유형검사, 기초학습능력검사

한국행동과학연구소: KATB(대졸용), KATB-h(비대졸용), KPDI 성격진단검사, KPDI-E 직무지향

한국심리학회지: 일반

성 검사, Customized Test, KBSS(KIRBS Biodata Screening System), KIT 지능검사(청소년), KAT 적성 검사(청소년), KPI-e 초등인성검사, KPI-a 청소년 인성검사, KAB 학습흥미검사(청소년), 진로검사 (청소년)

한국행동심리연구소: 우울증 검사, 불안증 검사, 자아존중감 검사, 주의력결핍 과잉행동성향 검사, 다중지능 적성검사, 스마트폰중독 진단검사, 온라인 중독 진단검사, 종합심리검사, 학습/지능 검사, 창의인성검사, 진로적성검사

허브적성검사연구소: 영역별 적성검사, 대학별 적성검사

헬로잡: 기업용 온라인 인성검사

휴브알앤씨: SNSB-II, SNSB-C, SNSB, APAC, SVLT-E, SVLT-C, K-RCFT, K-VCIHS-NP

ACG Assessment Consulting: 인성검사, 적성검사, 상황판단검사, 정서역량검사, 흥미검사

ASSESTA: MBTI, STRONG, FIRO-B, CPI

BIRKMAN KOREA: 관계적-직업적 특성 진단검사

EBS표준화심리검사 꿈의 지도: 유아부 심리검사, 초등부 심리검사, 중등부 심리검사

HUNO: 정신건강 검사, 발달/기질 검사, 양육검사, 초등 학습/인적성 검사, 청소년 학습/진로 검사, 대학용 검사, HR 검사

Leaders Insight: 인성검사, 역량검사

LTI 연구소: LTI(Life Type Indicator) 검사

ORP Institute: O-AT(ORPi-Aptitude Test), HBRI(Hogan Business Reasoning Inventory), HPI(Hogan Personality Inventory), HDS(Hogan Development Survey), MVPI(Motives, Values, Preferences Inventory), OPI(ORPi negative Personality Inventory)

PSI Consulting: 조직진단, 팀진단, 개인진단

SOLSO LAB.: 학습진단검사, Solso적성검사, ADHD, 중금속검사, 집중력검사, 자살충동검사, 우울증검사