



가 :

What Tests Miss: Hard Evidence on Soft Skills

**James J.
Heckman**

Asan Memorial Lecture

05

**James J.
Heckman**

Contents

제임스 헤크먼 약력	•	6
강연	•	9
질의 응답	•	53

James J. Heckman Profile	•	76
Lecture	•	79
Q & A	•	123

제 5 회 아산기념강좌

학력평가지험이 측정하지 못하는 것:
소프트스킬에 대한 확증

제임스 헤크먼

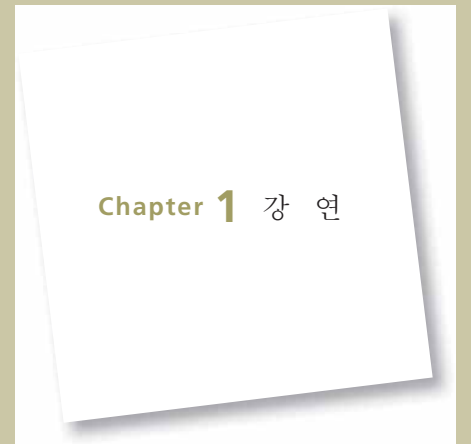
2011년 8월 8일

제임스 헤크먼

제임스 헤크먼(James J. Heckman)은 1973년부터 재직해 온 시카고대학에서 헨리 슉츠 경제학 석좌교수(Henry Schultz Distinguished Service Professor of Economics)를 지내고 있다. 그는 경제학과 미시경제학의 신기원을 이룬 연구로 2000년에 노벨 경제학상을 받았으며, 전 세계에서 가장 영향력이 큰 10대 경제학자 중 한 명으로 인정받고 있다.



헤크먼 박사는 현재 경제학과의 경제연구센터와 해리스 공공 정책 스쿨 사회보장프로그램 평가센터에서 센터장을 맡고 있다. 시카고대학 경제학과에 부임하기 전에는 컬럼비아대학에서 조교수로 재직했으며, 더블린 대학의 유니버시티 칼리지, 베이징대학, 예일대학, 뉴욕대학 등 많은 대학에서 강의한 경력이 있다. 그는 수많은 책과 논문을 발표했으며, 연구의 공로를 인정받아 많은 상을 받기도 했다. 헤크먼 박사는 콜로라도 대학에서 수학과를 졸업하고, 프린스턴 대학에서 경제학 석사와 박사 학위를 받았다.



학력평가
시험이
측정하지
못하는 것:
소프트스킬에 대한 확증

저는 오늘 강연에서 학력평가지험이 무엇을 놓치고 있으며 이런 시험이 놓치고 있는 것에 대해서 아는 것이 왜 중요한지를 이야기하고자 합니다. 학력평가지험은 학교나 사회내의 성과를 조사하고 검토하기 위해 한국을 포함한 많은 국가에서 널리 실시되고 있습니다. 미국도 낙제학생방지법, 학력평가지험, 아이오와 테스트 등 다양한 평가를 실시하고 있습니다. 하지만 이런 시험들은 비인지적 특성인 인성적 특성을 간과한다는 문제점이 있습니다.

당연한 지적이라고 여기실 수도 있겠지만, 사실 이 부분에 대한 고려가 늘 부족했던 것은 사실입니다. 소프트스킬 혹은 성격 기술로 불리기도 하는 이러한 인성적 특성의 예로는 인내심, 성실성, 성취 동기, 확고한 계획수립 등이 있습니다. 경제학계와 교육학계의 최근 연구에 따르면, 이러한 특성들이 인간의 각종 중요한 행동 양상을 예측할 수 있게 한다고 합니다. 따라서 이러한 비인지적 특성과 일반적인 성격적 요소들을 경시하는 것은 매우 위험한 관행이 될 수 있습니다. 이는 결국 사회 문제를 평가하고 취급하는데 있어서는 물론 사회정책의 성패를 평가하는 데 있어서도 많은 대가를 치르게 할 수 있습니다.

한가지 예를 들어 설명해 보겠습니다. 미국에는 아주 큰 규모로 치러 지는 GED라는 시험이 있습니다. 이 시험에 대해 여기서 너무 자세히 설명할 필요는 없겠지만, GED 시험은 중등학교 중퇴자가 치를 수 있는 시험으로 이 시험을 통과하면 일반 중등학교를 졸업한 학생들과 같은 지식 수준을 습득한 것으로 인정받습니다. 미국에서는 중등학생의 칠분의 일정도가 GED 시험을 치르고 있습니다. 또한 다른 여러 나라들도 이와 유사한 시험을 많이 시행하고 있습니다. 예컨대 브라질과 캐나다도 GED와 유사한 시험을 시행하고 있고, 그 외에도 이러한 류의 시험을 시행하는 많은 나라들이 있습니다. 저는 이 GED 시험을 토대로 하되 좀 더 포괄적으로 교육과 교육정책, 그리고 교육 평가에 관한 몇 가지 중요한 교훈을 이끌어 내고자 합니다.

인생의 성공을 위해서는 학업성취 이상의 것이 요구되고 있습니다. ‘소프트 스킬’은 측정할 수 없다고 생각되기 때문에 “소프트” 기술이라고 불리곤 합니다. 하지만 이는 사실이 아닙니다. 소프트 스킬은 측정될 수 있으며 인성의 경우 육성도 가능합니다. 훌륭한 학교, 훌륭한 교육, 훌륭한 가족은 이러한 기술 육성을 촉진할 수 있습니다. 많은 사회들은 이러한 기술을 촉진하는 프로그램을 제대로 탐구하지 않았고, 이는 곧 경제정책과 사회정책을 위한 강력한 잠재적 도구들이 간과되는 결과를 낳았습니다.

이 밖에도 제가 오늘 이끌어 내고 싶은 교훈들이 더 있습니다. 우선 ‘성적 책임(accountability)’ 움직임이란 것에 대해 이야기해 봅시다. 이는 교사와 학교가 주어진 임무를 잘 수행하고 있는 지를 알아보기 위한 시도입니다. 하지만 이런 움직임은 그릇된 동기를 발생시킬 수 있습니다. 점수에만 초점을 맞추는 시험과 인증서 같은 것들은 문제를 일으키기도 하고, 한편 문제를 감춰 버리기도 합니다. 이러한 행태는 교육의 목표를 왜곡하고 창조성을 억압하기도 하며, 사실 학생들의 학업 성과를 정확히 예측하는 데 있어서도 그다지 성공적이지 못합니다.

소프트 스킬은 또한 많은 결과물을 창출하기도 합니다. 우리는 소프트 스킬을 가지고 일반적인 행동, 건강과 관련된 행위, 십대 임신, 중등학교 졸업, 임금, 범죄 행위, 고용, 그리고 사회보장제도에 대한 의존성 등 다양한 결과를 예측할 수 있습니다. 실제로 소

소프트 스킬에 대한 몇 가지 확실한 증거에 따르면, 소프트 스킬이 인지적 기술보다 결과를 도출하는 데에 더 큰 역할을 하는 분야도 많이 있습니다. 그렇다면 이토록 중요한 소프트 기술은 왜 오랜 세월 동안 공공정책에 관한 논의에서 무시되어 왔을까요?

많은 사람들은, 심지어 전문 교육기관에서 계량심리학자라는 직함을 달고 있는 이들조차도, 소프트 스킬을 애매하고 약한 것이라고 묵살해 버리고 있습니다. 그들은 소프트 스킬을 매우 약하고 애매한 예측력을 지닌 개념으로 간주하며, 측정 또한 어려운 것으로 인식하고 있습니다. 다시 한 번 말씀 드리지만, 소프트 스킬은 분명 측정될 수 있으며, 높은 예측력을 지니고 있습니다.

우리는 인생의 성공에 있어서 여러 가지 다양한 기술들이 중요하게 작용한다는 것을 알고 있으며, 이는 직관적 수준에서도 명백합니다. 각기 다른 분야에서 성과를 내기 위해서는 다양한 수준에서의 다양한 재능이 필요합니다. 예컨대 세계적으로 유명한 발명가인 토마스 에디슨(Thomas Edison)은 19세기 미국에서 전구와 축음기를 비롯해 많은 물건들을 발명했습니다. 그는 열정적이었고 끈기도 있었습니다. 그는 두 가지 측면에서 칭송 받곤 하는데, 첫째는 수많은 실험을 수행하면서 보여 준 독창성 때문이고, 둘째는 자신이 이루고자 하는 바를 달성하기 위해 끊임없이 노력했다는 사실 때문입니다. 그는 스스로를 “천재는 1%의 영감과 99%의 노력으로 만들어 진다”는 말로 묘사한 바 있습니다. 토



마스 에디슨보다 조금 덜 유명하지만 비교적 최근의 인물을 예로 들자면, 미국 코미디언인 우디 앨런(Woody Allen)이 있습니다. 그는 에디슨 이야기의 초점을 조금 바꿔서, “성공의 80%는 일단 참여하는 것”이라고 말한 바 있습니다. 이는 곧 사회에의 참여가 대단히 중요하다는 의미입니다.

사회에 참여하려고 하고, 무엇이든 시도해 보려고 하는 사람들은 사회에서 훨씬 더 중요한 역할을 수행합니다. 이는 단순히 얼마나 똑똑한가의 문제가 아닙니다. 인생에서 일어나는 크고 작은 일들에 잘 대처하며 성공적인 삶을 살기 위해서는, 단지 추상적인 문제를 해결하는 능력보다 더 많은 능력이 필요합니다. 물론 맡은 임무나 살고 있는 사회에 따라 어느 정도의 인지 능력이 필요한가는 달라질 수 있습니다만 보통은 극기심, 자제력, 동기 같은 능력들은 재능을 육성하는 데에 있어 필수적입니다. 사람들은 직관적으로는 다양한 기술이 필요하다는 것을 알면서도 소프트 스킬을 무시하고 있습니다.

심지어 학술 토론회에서도 소프트 스킬이 주목 받지 못하고 있습니다. 한국에서도 그렇다고 들었습니다. 많은 토론이나 모임에서 지능, IQ, 인지능력, 혹은 시험을 주제로 삼고 그 외의 요소들은 뒷전으로 합니다. 사람들은 학교가 인지 능력을 가르치는 것은 당연하다고 생각하지만, 성격이나 인성까지 배양해줄기를 기대하지는 않습니다. 물론 늘 그래 왔던 것은 아닙니다. 과거에는 교육

이라는 개념의 폭이 더 넓었고 교육이 유교적인 개념으로 이해되었습니다. 그러나 지금의 교육은 학생이 사회에서 얼마나 잘 활동하고 잘 적응하는지가 아니라 시험에서 얼마나 좋은 성적을 거두는지에 초점을 맞춥니다. 한국만 봐도 학원이 학생의 성격을 형성하거나 기본적인 사회 활동 능력을 키워 주는 것이 아니라, 시험만 통과할 수 있도록 지도하고 있다는 것을 알 수 있습니다.

보통 인격 배양과 기타 전통적인 가치관의 형성은 가족이 책임져야 할 부분으로 여겨 집니다. 하지만 다른 나라들에 비해 가족을 더 중시하는 한국에서조차 많은 가정이 이 부분 때문에 스트레스를 받습니다. 물론 어떤 가족은 다른 가족보다 자녀의 소프트 스킬을 잘 육성해 내기도 합니다. 그러나 결손 가정들은 이런 부분에서 자녀들을 충분히 지원해주지 못한다고 합니다. 저의 제자인 문성혁 군은 시카고 대학에서 결손 가정과 양부모 가정의 차이점을 보여 주는 중요한 연구를 해 왔습니다. 그에 따르면 약 18%의 한국 어린이가 편모 혹은 편부 가정, 이혼 가정, 혼외 출산 가정들을 포함한 결손 가정에서 자라고 있으며 그 수치가 점차 증가하고 있다고 합니다.

한국에는 아동들에게 주어지는 교육과 성공을 촉진하는 환경이 불평등하게 존재한다고 보고하는 자료가 있습니다. 효과적인 사회 정책을 고안하려면, 인지력과 인지력 시험에 대한 현재의 일차원적 초점을 넘어서서 바라볼 필요가 있습니다.

시험의 역사는 고대 중국으로 거슬러 올라가며, 실제로 한국에서는 수백 년 전부터 시험제도를 통해 관료를 임용했습니다. 그러나 오늘날 전 세계적인 토론을 불러일으키고 있는 현대 교육 운동은 1840년대 미국의 호러스 만(Horace Mann)이라는 교육가에 의해 시작되었습니다. 호러스 만은 처음으로 표준화된 학력평가의 개념을 소개했고, 이것은 세계 교육사에 한 획을 그었습니다. 그는 시험이란 것을 도입한 장본인이었지만 곧 시험만으로 학교의 본 목적을 측정하기에는 부족하다는 사실을 깨달았습니다. 그의 관점은 부분적으로 유교와 비슷했으며, 학교의 주된 기능을 도덕과 인격을 배양하는 것이라고 했습니다. 그는 1867년에 “학교를 교육의 질이 아닌 학습 기간으로 평가하는 것은 가장 어리석은 일이다. 산수, 문법 등 기타 기초 과목들은 학교 교육의 아주 작은 부분에 불과하다. 감정의 기초는 사고의 기초 못지않게 교육되고 있으며, 정서와 열정이 지능보다 더 많은 교훈을 얻게 한다. 이 부분에 대해 눈에 보이는 부분은 더 적을 수 있지만, 가려져 있는 훈련은 더 많다”고 했습니다. 이런 식의 사고는 늘 존재해 왔습니다. IQ테스트, 학력평가, 실제로 시스템의 근저를 이루는 시험 등을 개발한 모든 사람들이 시험의 한계효용을 인식하고 있었습니다. 그러나 적절한 기술이 개발되지 않았기 때문에 호러스 만의 생각이 19세기에는 채택되지 않았던 것입니다.

20세기에 들어와 존 듀이(John Dewey)를 비롯한 진보주의자들과 능력 중심의 사회를 만들고자 한 자들은 학교 교육을 보다 많

은 사람들에게 개방하기 위해 전 조직을 대상으로 하는 시험을 제도화하고자 했습니다. 20세기 초반 존 듀이와 같은 진보주의자들은, 비록 존 듀이 자신은 IQ테스트를 개발하지 않았지만, IQ테스트에 의지했습니다. IQ와 인지 시험이 입학생들을 선별하는 역할을 할 수 있으리라 기대했던 것입니다. 초기의 IQ테스트는 학교 부적응자를 걸러 내는 정도의 용도로만 고안되었고, 학교 내에서의 학업성취도를 예측하기 위한 것은 아니었습니다. 하지만 IQ테스트의 역할이 빠르게 확산되어 학교 내에서까지 학생들을 분류하는 용도로 사용되게 된 것입니다.

프랑스인이 최초의 IQ테스트 중 하나를 개발했다는 것은 매우 흥미로운 사실입니다. 그 테스트를 다시 비네(Binet)라는 프랑스인이 개량했습니다. 비네는 스탠퍼드 대학의 심리학자 루이스 터먼(Lewis Terman)과 팀을 이루어 스탠퍼드-비네 테스트를 만들었는데 이것이 오늘날에도 널리 쓰이는 표준 IQ테스트로 남아 있습니다. 이 테스트는 거의 100년 전에 개발되었는데, 그 당시 비네는 학교 생활에서 성공이란 “지능 이외의 요소를 인정”하는 것이며, 학교 생활에서 성공하려면 집중력, 의지력, 온순함, 올바른 습관, 지속적인 노력 등의 요소를 지녀야 한다는 점을 지적했습니다.

심지어 100년 전에도 이미 ‘g’인자라고 불리는 것에 대한 강조가 있었고, 이에 대해 연구하는 이들도 많이 있었습니다. g인자는 지능을 측정하기 위해 고안된 단일 요소입니다. g인자를 살펴봄으

로써 성실성과 같은 비인지적 특성이 성공적인 학교생활을 측정하는 데에 주요한 역할을 한다는 사실을 알게 되었습니다. 최근에는, 지능의 힘을 지지한 아서 켄슨(Arthur Jensen)이 g인자에 대한 책을 저술하였습니다. 그는 성실성, 책임감, 의존성, 배려, 조직화 및 인내심과 같은 특성이 사람들을 성공으로 이끄는 중요한 요소라고 설명합니다.

그럼에도 불구하고 이 모든 개념이 쉽게 무시되고 있는 이유는, 시험을 개발할 당시에 과학적이고 객관적인 자세를 유지해야 한다는 생각이 팽배했기 때문입니다. 19세기 후반과 20세기 초반, 미국은 프레더릭 테일러(Frederick Taylor)라는 인물의 활동으로 인해 효율성과 생산성이라는 화두에 사로잡히기 시작했습니다. 제 생각에는 오늘날의 한국도 당시의 미국과 마찬가지로 효율성과 생산성에 집착하고 있는 것 같습니다. 테일러주의는 당시 대단한 반향을 불러 일으켰습니다. 이 사상은 과학적인 관리, 측정, 책임, 그리고 어떤 의미에서는 공정성을 보장하자는 취지로 전개됐습니다. 그러나 과학적 관리의 원칙을 어떻게 학교에 적용해야 하는가 라는 문제가 남아 있었습니다.

그러다가 교육 철학자이자 이론가인 시카고대학 교수 보비트(Bobbitt)가 등장했습니다. 그는 사람들에게 동기를 부여하고 가르치는 일에 매우 큰 영향력을 행사했습니다. 그에게 영향을 받은 이들은 이후 PISA, SAT, 학력평가지험과 같이 현재에도 널리 사

용되고 있는 시험을 개발하기도 했습니다. 호러스 만, 아서 켄슨, 비네의 글을 읽으면서 교육이란 무엇인가 하고 질문을 던져 보는 것은 흥미로운 것입니다. 보비트는 교육을 공장과 같이 생각했는데, 그렇다면 그 공장은 무엇을 생산했을까요? 그는 교육을 철도 제작 공정과 비슷하다고 여겼지만, 철도 대신 성격(character)과 인성(personality)을 만들어 내는 것이라고 생각했습니다. 보비트는 자신의 이론으로는 학교의 실질적인 활동을 제대로 측정할 수 없다고 생각했고 호러스 만과 같이 인성을 매우 중요한 특성으로 여겼습니다.

보비트의 제자들이 해 온 활동을 보면 흥미롭습니다. 20세기 초에 IQ 테스트가 개발된 이후 새로운 개념이 하나 등장했습니다. 이는 보다 근래의 개념이었고, PISA시험과 많은 학력평가의 기반이 되었습니다. 이 개념은 시카고 대학에서 개발되었으며, “일반 지식”이라 불리고 있습니다. 이는 가공되지 않은 지능을 측정하는 데에 그치지 않고, 개인이 특정한 과정을 통해서가 아니라 일반적인 방법으로 학습하고 있다는 점을 포착하였습니다. 이것은 매우 새로운 개념이었습니다. 50년 전에는 이런 발상을 하는 사람이 없었습니다. 이런 발상은 IQ나 문제 해결 속도를 측정할 뿐 아니라 중요한 생활 기술까지도 측정했습니다. 아이오와 테스트, GED시험, 낙제학생방지 시험 그리고 PISA는 모두 이러한 유형의 시험들이었으며, 특정 지식보다는 실생활의 중요한 기술을 포착하도록 만들어 졌습니다.

이런 시험들은 점수 매기기가 비교적 쉬웠지만, 광학주사장치를 사용해서 채점하는 기술이 등장한 1950년대에 들어서서야 비로소 완전해 졌습니다. 여기에 현대 시험의 양식을 창출해 내었다는 점에서, 현대 시험에 있어서 아마도 가장 중요한 인물이라 불릴 만한 테일러가 필기시험의 가치에 대해 서술한 것이 있습니다. 그는 이러한 시험의 대부이긴 하지만, “우리는 필기시험, 몇 가지 형태의 객관적인 시험, 그리고 때로 교사의 주관적인 인상에 크게 의존하고 있다. 그러나 여기에 한정되지 않고 활동 기록, 질문지, 일화 기록, 관찰 기록과 같은 다른 많은 평가 장치를 사용할 수도 있다”고 말합니다. 또한 그는 이런 기타 장치들을 활용하도록 격려했습니다. 사실 그는 그의 생애 말기에, 즉 이런 시험들이 표준이었고 PISA 시험은 여전히 개발되는 중이던 시기에, 현재 시행되고 있는 것보다 훨씬 더 다양한 유형의 시험을 지지했습니다.

1940년대와 1950년대에 이런 일련의 운동에 이어서 개발된 GED 시험을 예로 들어 보겠습니다. GED 시험은 군인이 제 2차 세계대전 때 배웠던 일반 지식을 인증해주기 위해 개발된 학력평가 시험입니다. 미국은 제 2차 세계대전 당시 많은 고등학교 남학생들을 징집했고 이렇게 징집된 이들은 전쟁에 참전하느라 중등교육조차 제대로 마칠 수 없었습니다. 그러나 2~3년 간 군대에 복무하면서 수업을 들었을 경우 중등학교 졸업자와 유사한 기술을 습득했다는 추정이 가능했습니다. 그들은 엄격한 군대 규율에 따라 선발되었고 혹독한 군대 생활에서 살아남았을 뿐 아니라, 지시에

잘 따르고 자제심과 협동심이 강했습니다. 따라서 이 시험이 탄생한 시기에 이러한 추정은 매우 당연한 것이었습니다.

원래는 군인에게 일반 고등학교 졸업자와 동일한 자격을 주기 위해 개발된 이 시험이 점차 미국 사회에서 훨씬 폭넓은 용도로 사용되었다는 것은 흥미로운 점입니다. 현재의 GED 시험은 엄격한 군생활을 했던 군인이 아닌, 중등학교를 중퇴한 민간인을 대상으로 실시되고 있습니다. GED 시험은 미국 교육에 광범위한 영향을 미쳤으며, 확산하건대 한국에도 현존하고 있는 확장 운동의 일환이었습니다. 이는 평등주의, 능력주의에 대한 갈망입니다. 미국에서 널리 활용되고 있는 SAT 시험은 본질적으로 학연 중심주의를 깨고 가난한 집안 출신의 똑똑한 어린이들도 사회의 최상층까지 진입해 출세할 수 있도록 길을 터주고, 실력에 따라 공정하게 평가 받는 사회를 만들기 위해 고안되었습니다.

그밖에 1950년대와 1960년대 미국 사회에서는 테일러리즘을 부활시키고 정부가 책임을 잘 수행하고 있다는 점을 보여 주고자 하는 움직임이 강하게 일어났습니다. 그 점수만 봐도 정부가 얼마나 성공적이었는지를 가늠해 볼 수 있었습니다. 테일러주의의 극단적인 형태는 베트남 전쟁 중의 미국 정책에서 명백하게 드러났습니다. 예를 들어 로버트 맥나마라(Robert McNamara) 국방장관은 국방부에서 테일러주의를 부활시켰으며, 손익계산서의 사회적 버전을 만들려고 그에 맞는 경제 원칙을 창출했습니다. 이

는 베트남에서의 전사자 수를 다루는 극단적인 유형으로, 사람들은 미국 사망자 수 대비 베트남 사망자 수에 대해 언급하곤 했습니다.

보다 넓게 표현하자면, 시험 등 일정한 형태의 측정 방법을 사용하면 사회의 책임감을 높일 수 있을 것이라는 생각이 있었습니다. 1960년대에 시작된 ‘위대한 사회’가 그런 경우였습니다. IQ 테스트와 학력평가시험은 학교가 학생들에게 얼마나 많은 가치를 제공해 줄 수 있는지를 측정하는 성적 책임 운동에서 아주 중요한 부분이 되었습니다. 이런 운동은 낙제학생방지법을 실시한 부시 정부 시절에 미국에서 열풍을 일으켰습니다. 낙제학생방지법은 근본적으로 모든 학생들이 읽기와 기초 수학 시험에서 아주 좋은 결과를 낼 수 있게 하기 위해 학교를 부추겼습니다. 이로 인해 교사가 시험 문제 풀이 중심으로 학생을 가르치고, 시험에 나올 만한 내용만을 학생들에게 이해시키는 문화가 형성됐습니다.

이에 따라 1955년부터 21세기 초반까지 표준화 시험의 판매는 꾸준히 증가했습니다. 한때 GED 시험을 통해 고등학교 졸업 자격을 인정받은 사람의 수가 총 고졸자수의 거의 20%에 달했습니다. 매년 약 70~80만 명의 학생이 중퇴했으며, 이들은 4년 동안 고등학교에 다니면서 열심히 공부하는 대신 GED 시험을 쳤습니다.

여기에서 시험의 예를 들어 보겠습니다. 간단한 대수방정식 문제

입니다. $8x + 16 = 32$ 라면 x 는 무엇인가? 정답은 2입니다. 이는 중학교 2학년(8학년) 수준입니다. 더 어려운 문제도 있습니다. 하루에 나무 15그루를 심을 수 있다고 하면 200그루를 심기 위해서는 며칠이 걸리는가? 이것이 대략적인 GED의 수준이라고 할 수 있습니다.

그렇다면 GED 자격증을 취득한 이들은 어떤 사람들일까요? GED 취득자는 고등학교 졸업자와 중퇴자의 중간에 속하지만, 대학에 진학하지 않은 고등학교 졸업자와 지적 수준이 비슷합니다. 그림 1은 교육 수준에 따른 여성의 인지적 스킬을 보여주는 그래프이고, 그림 2는 남성의 인지적 스킬을 보여주는 그래프입니다. 전형적인 종형 곡선에서 볼 수 있듯이 GED 취득자와 고등학교 졸업자 사이에 인지 능력의 차이가 거의 없습니다.

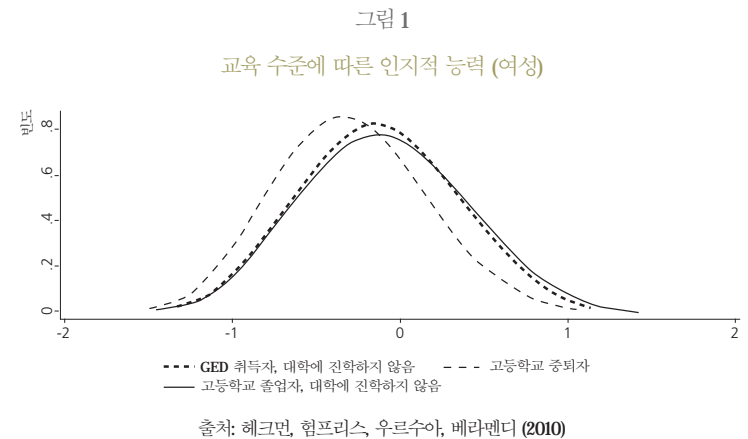
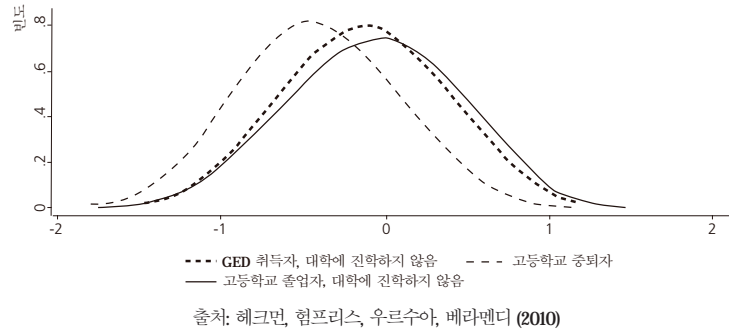


그림 2

교육 수준에 따른 인지적 능력 (남성)



GED 시험을 통과한 사람의 지적 수준이 일반 고등학교를 졸업한 사람의 지적 수준과 거의 동일하다는 점에는 의문의 여지가 없습니다. 그러나 지적 수준이 비슷하다 할 지라도 GED 시험 취득자는 일반 고졸자에 비해 소프트 스킬이 부족합니다. 이러한 소프트 스킬의 총합을 살펴보면 GED 취득자는 중퇴자의 소프트 스킬을 지니고 있으며, 고등학교 졸업자의 소프트 스킬과는 격차가 크다는 점이 드러납니다. 여성(그림 3)과 남성(그림 4)에서 동일한 결과가 나타납니다.

여기에서 주의를 기울여 반드시 이해하고 넘어가야 할 점이 있습니다. 바로, 실질적 수행성과를 고려할 때의 상황, 즉 시험에서만 이 아니라 실제로 사람들이 노동시장에서 어떻게 일하고 있는지를 보는 것입니다. GED 취득자들의 임금은 고등학교 중퇴자와

그림 3

교육 수준에 따른 비인지적 능력 (여성)

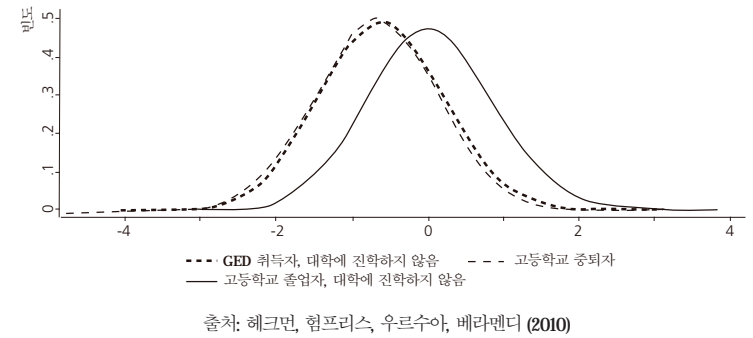
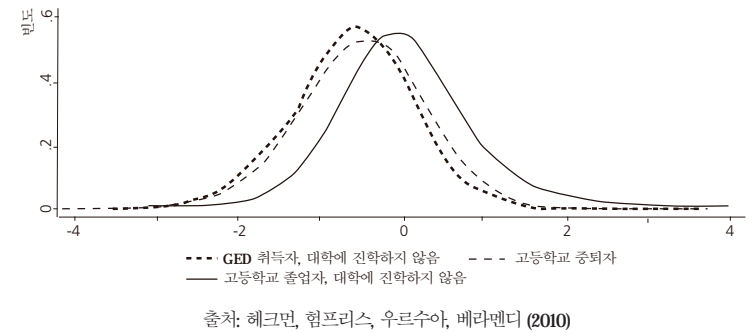


그림 4

교육 수준에 따른 비인지적 능력 (남성)



같은 수준입니다. 따라서 그 결과가 인지능력에 따라 조정된 것이 아니라면, GED 취득자는 일반적인 고등학교 졸업자와 지적 수준이 동일하지만 그에 비해 수입이 적고 사회에서의 수행성과가 낮습니다.

한 사람은 자격을 인정 받지만 실제로 임무를 지속하거나 제대로 마무리 할 수 없을 때 **GED**는 흥미로운 사회적 실험이 됩니다. 따라서 **GED**가 제 역할을 하는지 여부가 의문입니다. **GED**는 능력을 보여 주는 중요한 신호로 노동 시장에서의 가치를 보여줍니다. 따라서 **GED**를 취득하기 전과 후의 임금을 비교하면서 실제로 **GED** 시험이 그런 신호의 역할을 하고 있는지 여부를 검증할 수 있습니다. 그러나 이를 검증해주는 어떠한 신호 값도 존재하지 않습니다. 때문에 **GED** 자체의 효용성은 미심쩍다 하겠습니다.

그러나 **GED**에만 초점을 두지는 맙시다. 아주 많은 사람이 의존하고 있는 각종 시험들이 포착하는 것은 정확하게 무엇일까요? 시험의 타당도는 어떻게 입증되는 것일까요? 이를 위해 산업계 종사자는 대개 검사를 수행하려 합니다. 자동차 회사는 자동차를 만든 후에 도로에서 주행 검사를 합니다. 선박 회사는 배를 만든 후에 항해 검사를 합니다. 사람들은 시험에 대해서도 이와 같이 검사를 할 수 있습니다. 이러한 시험의 타당성은 어떻게 입증하는 것일까요? 어떤 이는 예측 타당성에 주목할 수 있습니다. 이는 이미 결정된 것입니다. 시험은 정확히 무엇을 예측하는 것일까요? **IQ**, 등급, 학력평가지험의 예측력은 얼마나 정확할까요? 흥미롭게도, 학교에서 좋은 성과를 거둔 학생과 그렇지 못할 학생을 예측하도록 만들어 졌다고 표방하는 각종 시험들은 그다지 효과적이지 못합니다.

일반 지식체계에서는 인지 시험에서 포착되는 인지능력 및 각

종 기능이 복잡한 임무에 대하여 실질적인 예측 능력을 가진다고 알려져 있습니다. 복잡한 임무를 완수하기 위하여 정신적 능력이 필요하다는 점에는 의문의 여지가 없습니다. 그러나 육체노동을 하는 직업의 다양한 임무에 대해서도 소프트 스킬의 예측력이 대단히 높다는 점에서, 임무의 복잡성 여부를 떠나 모든 임무에서 인성이 보다 높은 예측능력을 보유하고 있음을 알 수 있습니다. 그러나 복잡한 임무에서도 소프트 스킬의 예측능력은 대단합니다. 대체로 시험은 실생활의 행동에서가 아니라 다른 시험과의 상관관계를 통해서 그 타당도를 인정 받습니다. 인지 시험의 타당도를 생각해 보십시오. 매우 공통적인 타당성 측정 방법은 결과의 변동 부분을 위와 같은 시험으로 설명하는 것입니다.

예를 들어 표 1과 같이 **IQ** 테스트, 평균학점(**GPA**) 그리고 필기시험 점수에서 교사 평가의 정확성을 생각해 보십시오. 이는 학력평가지험과 비슷합니다. 사실상 학력평가지험은 35세가 될 때 받게 될 임금을 어느 정도 예측할 수 있습니다. 총 예측 수준의 정도는 상당히 낮습니다. 1은 완벽한 예측을 말하고 .05는 약 5%의 변동을 의미합니다. 복지 분담 혹은 근무 시간이 고려될 경우 시험의 예측수준이 약간 올라가지만, 여전히 예측 수준은 그리 높지 않습니다. 여기에서는 매우 흥미로운 패턴이 발견되는데, 학력평가지험이 사실상 **IQ** 테스트보다 더 높은 예측 수준을 보인다는 것입니다. 이는 매우 흥미로운 결과입니다.

표 1

노동 시장에서의 타당도 (미국 청소년 종단 조사, 1979)

결과	남성			여성		
	IQ	평균 학점 (10학년)	미군 입대 자격 시험	IQ	평균 학점 (10학년)	미군 입대 자격 시험
35세 시간당 임금	0.03	0.05***	0.05***	0.11***	0.10***	0.13***
35세 근무 시간	0.10***	0.12***	0.21***	0.02	0.10***	0.17***
35세 복지 상태	-0.09***	-0.11***	-0.23***	-0.20***	-0.23***	-0.36***

주: 미군 입대 자격 시험 = AFQT (Armed Forces Qualification Test)
*** $P < .001$

이 결과가 의미하는 바는 무엇일까요? 또 이와 같은 결과가 나오는 이유는 무엇일까요? 학력평가시험이 완벽하다고 할 수는 없겠지만 인성을 부분적으로나마 포착합니다. 실생활의 수행 능력을 측정한다는 면에서 인성이라는 척도가 더 나올까요 아니면 더 나쁠까요? 이 점이 진짜 관심거리입니다.

최근 성격심리학에서는 빅 파이브(Big Five)라고 불리는 것이 개발됐습니다. 이는 사람의 성과에 어떤 차원이 있는지를 판별할 수 있게 하는 척도입니다. 그간 예측의 유효성에 대해 많은 연구가 이루어졌습니다. 다섯 가지 특성인 개방성(Openness), 성실성(Conscientiousness), 외향성(Extraversion), 우호성(Agreeableness), 신경증적 경향(Neuroticism)의 머리말을 따서 만든 OCEAN이란 용어가 사용되기도 합니다. 이 요소들은 인지

표 2

성실성과 SAT 점수의 대학 GPA 상대 예측력

출처	표본	측정 및 결과의 시기	통제 요인	단위	결과	
코나드 [2005]	미국 대학생 (N=186)	대학 재학생이 대학 GPA와 SAT 점수를 직접 보고했다. 성격은 대학에서 측정됐다.	수업 출석	표준화 회귀계수(β)	SAT 총계 성실도	0.27 0.30
노프트 및 로빈스 [2007]	미국 대학생 (N=10,497)	대학 재학생이 대학 GPA와 SAT 점수를 직접 보고했다. 성격은 대학에서 측정됐다.	성별, 기타 빅 파이브 특성	표준화 회귀계수(β)	SAT 언어 영역 SAT 수학 영역 성실도	0.19 0.16 0.24
울프 및 존슨 [1995]	미국 대학생 (N=201)	대학 재학생이 대학 GPA와 SAT 점수를 직접 보고했다. 성격은 대학에서 측정됐다.	고등학교 GPA	표준화 회귀계수(β)	SAT 총계 성실도	0.23 0.31

주: (a) 학생이 직접 보고한 SAT 점수와 대학 기록에서 제공받은 SAT 점수의 상관관계가 대단히 높다($r = 0.92$).
(b) 학생이 직접 보고한 GPA와 대학 기록에서 제공받은 GPA의 상관관계가 대단히 높다($r = 0.89$).

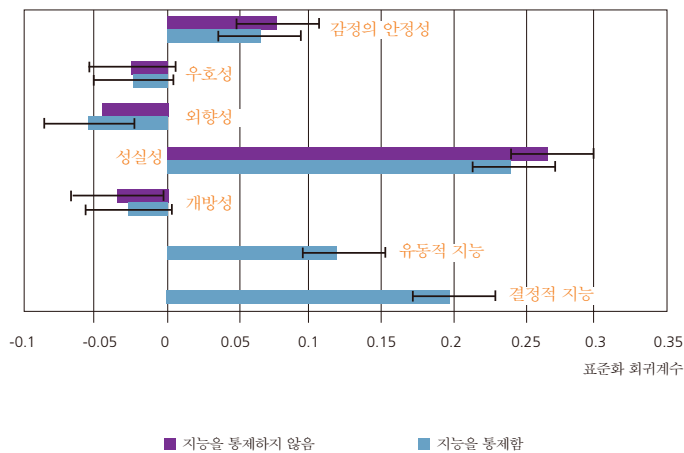
에서 IQ가 가지는 영향력, 그리고 대학에서의 성공가능 여부를 예측과 관련되어 연구되어 왔습니다. 이는 심리학자, 교육학자, 경제학자가 실시한 연구입니다. 표 2는 각기 다른 세 가지 연구를 보여 주고 있습니다. 그런데 여기서 주목해야 할 핵심 질문은 “한

가지 인성, 즉 성실성의 상대적 예측력은 무엇인가?’입니다. 이번에도 SAT 시험과 같은 것을 고려해 보십시오. 교육 평가의 중심 기관인 미국 ETS (Educational Testing Service)조차 성실성 하나만으로도 성공, 점수, 대학에서의 성과를 SAT 점수보다 훨씬 잘 예측할 수 있다고 인정합니다. 이 점은 교육 평가 공동체의 관점에서 봤을 때 당혹스러운 일입니다.

성실도는 측정할 수 있을 뿐 아니라, 성실도를 측정하면 예측 능력이 훨씬 좋아지는 것으로 드러납니다. 예를 들어 그림 5는 독일에서 실시된 연구 내용입니다. 빅 파이브의 척도를 살펴보고 이 척도를 같은 결과에 적용하여, 평가 대상자의 문제 푸는 속도를

그림 5

빅 파이브 및 지능과 학교 교육년수의 관계 (남성)



측정하는 ‘유동적 지능’과 학력평가시험과 유사하게 지식을 측정하는 ‘결정적 지능’을 더한 다음 실질적으로 학교에 다닌 년수를 확인해 보면, 성실도가 보다 훌륭한 예측 기준이라는 것을 알 수 있습니다. 완벽한 예측 기준이란 없겠지만, 성실도는 IQ나 학력평가시험보다 월등히 나은 예측 기준인 것입니다.

학과 성적을 고려하면(그림 6) 인성 중에서 성실성이 더욱 확실한 예측 기준이라는 점이 다시 한 번 분명해 집니다. 표준 학력평가 시험 점수와의 연관성 같은 것을 고려하면(그림 7) 다양한 인성적 측면이 예측 기준의 역할을 한다는 점도 분명해 집니다. 그렇다면 이것은 무엇을 의미할까요? 사망률에 대한 요소를 살펴 보겠습니다(그림 8). 일반적으로 교육 정책에서 사망률과의 연관성을 고려하지는 않지만, 사실 이것은 중요하게 고려돼야 할 요소입니다. 이 경우에도 성실도가 다른 특성보다 훌륭한 예측 기준이라는 점이 확실해 집니다. 인간이 언제 사망할지 확실히 예측할 수 있는 기준은 없지만, 성실성이나 극기심, 자제력은 IQ보다 우수한 예측 기준입니다. 또한 그래프에서 인지 기능과 비인지 기능을 살펴보면 사회적 소프트 스킬이 아주 낮은 사람들은 감옥에 들어간 경험이 있을 가능성이 훨씬 높다는 점이 분명해 집니다(그림 9).

그렇다면 여기에선 무슨 일이 일어나고 있는 걸까요? IQ는 성취도와 동일하지 않습니다. 학력평가시험은 인성을 통해 부분적으

그림 6

빅 파이브 및 지능과 학과 성적의 상관관계

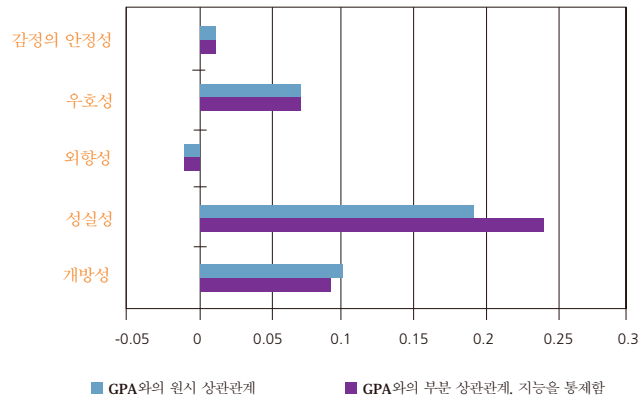


그림 8

인성, IQ 및 사회경제적 지위 (SES)와 사망의 상관관계

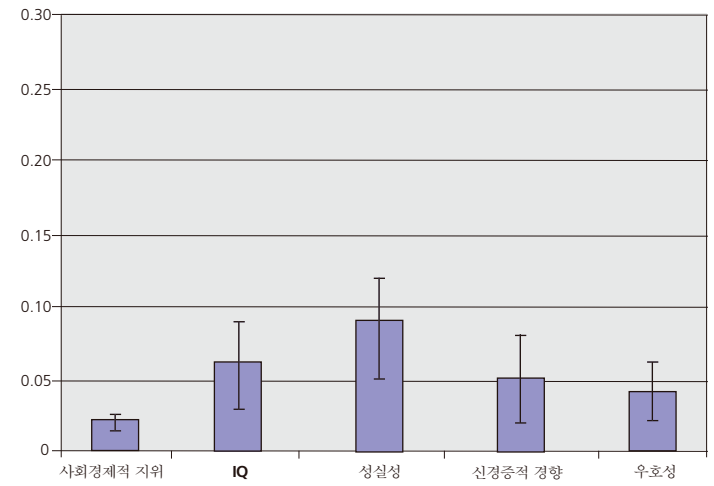


그림 7

표준 학력평가시험 점수와의 관계

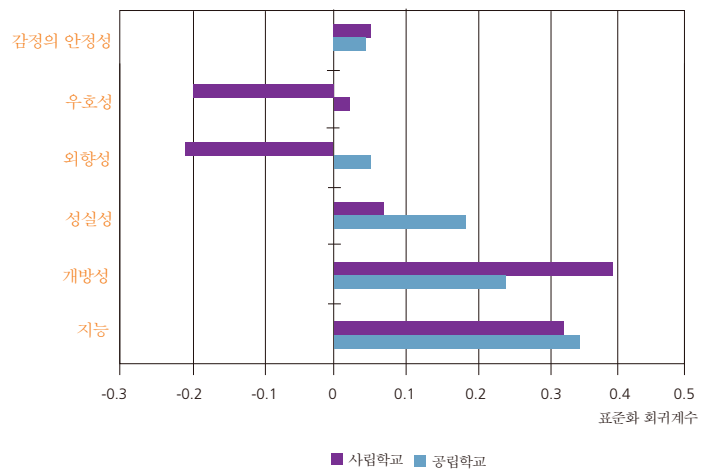
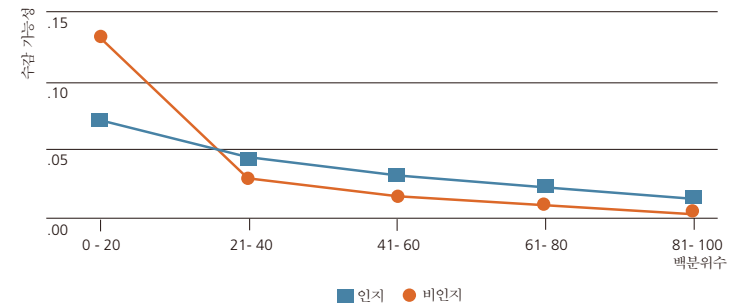


그림 9

능력에 따른 30세까지 수감 가능성 (남성)



출처: 헤크먼, 스티스러드, 우르수아 (2010)

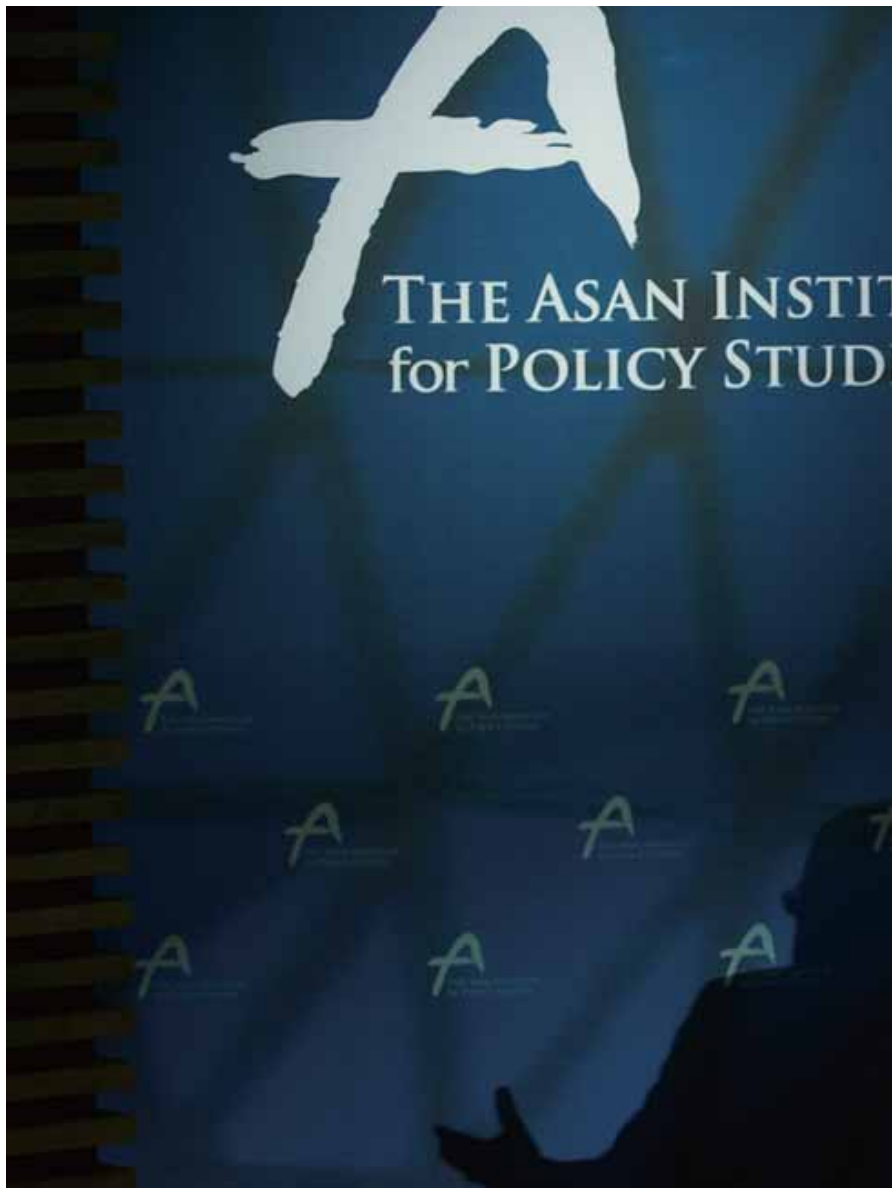
로 설명될 수 있는데, 이에 대해 아직 완전한 연구가 이루어지지 않았습니 다. 학력평가시험은 의도치 않게 인성의 일부 측면을 포착합니다. 예를 들어, 최근 몇몇 연구는 학력평가시험 성적에서 얼마나 많은 변동이 IQ로, 혹은 인성으로 설명될 수 있는지를 살펴 보았습니다. 이는 학력평가시험 점수로 보자면 좋은 소식인 동시에 나쁜 소식입니다. 학력평가시험 점수에서 변동에 관련된 상당 부분이 인성의 단 두 가지 요소로 설명됩니다. 이 모든 요소를 감안하면 설명이 더욱 확실해 집니다. 그렇다면 과연 무엇이 사용되고 있는가라는 의문이 생깁니다. 무비판적으로 사용된 학력 평가 시험 점수인가요? 학력평가시험 점수는 정확히 무엇을 측정하는 것일까요? 같은 맥락에서, IQ 테스트 점수는 무엇을 측정하는 것일까요?

여기에서 악명 높은 예를 하나 들어 보겠습니다. 미국 흑인과 백인 간 IQ 테스트의 평균 점수에는 상당한 차이가 있다는 이야기는 잘 알려져 있습니다. 100점을 기준으로 할 때 약 16점 차이가 납니다. 40여 년 전에 실시된 한 실험에서 사회적·경제적 혜택을 받지 못하는 흑인 가정의 어린이들은 IQ 시험에서 정답을 말할 때마다 사탕을 받았습니 다. 그 날 테스트의 인센티브라 할 수 있는 사탕을 받은 어린이들은 12점이 높은 점수를 받았습니 다. 이는 학력평가시험 혹은 적어도 IQ 테스트에서는 흑인과 백인의 격차가 근본적으로 제거되었다는 사실을 의미합니다.

따라서 연구자들은 시험 성과도 부분적으로는 인성적 측면에 의해 동기 부여를 받는다는 점을 이해하게 됐습니 다. 사실, 이후의 연구에서 연구자가 매우 성실한 아동들을 살펴보았을 때에는 이러한 인센티브에 대한 아이들의 반응이 아주 약하게 나타났습니 다. 이유가 무엇일까요? 이는 성실한 학생들은 이미 열심히 노력하고 있었기 때문이었습니다. 반면 학습에 대한 동기 부여가 낮은 학생들은 시험 성적이 좋지 않았습니 다. 결과적으로 시험은, 부분적으로는 동기를, 부분적으로는 순수한 지능을 측정하고, 또 부분적으로는 지식을, 부분적으로는 욕구를 측정하고 있습니다. 인센티브는 동기부여 수준이 더 낮은 사람들에게 보다 큰 효과를 가 집니다. 연구자들은 동기부여를 통해 시험 점수가 오를 수는 있어도 이것이 ‘지불된 성과’라는 점에서 볼 때 인센티브가 지속적인 효과를 내지는 않는다는 점을 발견했습니다.

그렇다면 학력평가 시험이 치르는 기회비용은 무엇일까요? 미국 사회와 한국 사회 모두 학력평가 시험에 일정한 대가를 치르고 있습니다. 미국 사회가 치르고 있는 대가의 예를 살펴 보겠습니다. 학생들이 비록 학력평가 시험에서 높은 점수를 얻는다 하더라도, 인지적 특성이 동일한 사람들 중 일부는 보다 큰 사회에서 상당히 부족한 수행 성과를 낼 수 있다는 점을 확증해주고 있습니다. 이는 미국의 공공정책을 기만하는 결과로 나타 납니다. 똑똑한 사람들에 대한 공공정책의 일차원적 초점은 교육기관과 개인의 노력을 잘못된 방향으로 인도함으로써 주요 문제들을 감

소프트 스킬은 중요합니다.
인지적 기술과 비인지적 기술 모두 유전적으로
정해진 것이 아닙니다. 이런 기술은
심지어 청소년기나 성인기에도 형성될 수 있습니다.
이런 기술을 향상시키는 것이야말로
사회 정책의 생산성을 높이는 길입니다.



추고 있습니다. GED 취득자는 중퇴자와 비슷한 소득 수준을 갖습니다. 미국의 경우 GED를 취득한 사람들을 고등학교 졸업자와 동일하게 분류할 경우 고등학교 졸업 자격을 취득한 이의 수가 급격히 과장돼 버립니다. GED 취득자를 제외하고 보면 지난 30년 동안 고등학교 졸업률은 사실상 하락세를 보이고 있습니다. 이는 미국사회의 각종 문제점을 설명하는데 도움을 주고 있습니다.

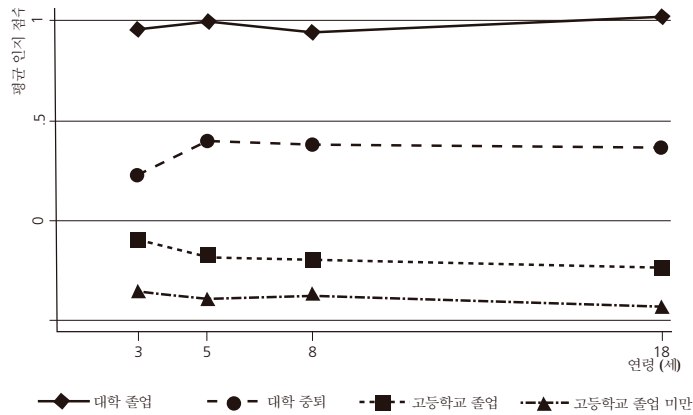
물론 GED가 꼭 나쁜기만 하다는 것은 아닙니다. GED가 수행하고 있는 역할은 무엇일까요? GED는 학력평가 시험입니다. 어떤 이는 이 시험을 통과해서 이득을 얻을 수도 있습니다. 실제로 GED 자격증을 이용해 이득을 얻는 사람들은 그들의 삶에서 몇 가지 실수를 저질렀던 이들입니다. 그들은 GED 시험을 보았고 인지적, 비인지적 기술의 수준이 모두 높습니다. GED를 취득하면 선택의 여지가 생깁니다. 그러나 대부분의 GED 취득자들이 대학에 진학할 정도로 똑똑할지라도 실제 대학에 입학하는 사람들은 40% 이고, 그 중 3%만이 대학을 졸업합니다. 따라서 이들이 보는 이득의 수준은 대단히 적습니다. 다른 기준도 있습니다. GED는 유혹에 약한 청소년들, 기본적으로 10대 후반에 속하는 청소년들에게 제공되며 이들로 하여금 나쁜 선택을 하게 유도합니다. 청소년들이 학교에 가지 않고 대신 시험을 치를 수 있는 또 다른 선택을 갖게 된다면, 그들은 현명하지 못한 선택을 하게 될 수도 있는 셈입니다.

예를 들어, 최근의 몇몇 심리학 연구는 아동의 지적 능력이 훨씬 더 급속히 성장한다는 것을 보여 주고 있습니다. 지적 기능은 보다 어린 시기에 형성되며, 정신적 성숙, 즉 인성 기능은 그 이후에 형성됩니다. 이 시기의 어린이들에게 GED 시험을 제공하는 것은 학교를 중퇴하도록 유도하는 것이며, 이는 사실상 그들이 평생 짊어지고 살아야 할 지도 모르는 심각한 문제로 이끄는 셈입니다.

제가 제안하고자 하는 것은 인지적 기술 외의 다른 여러 기술이 필요하다는 것입니다. 그러한 기술이란 무엇일까요? 인지적 기술과 비인지적 기술은 구체적으로 무엇이며, 그런 기술을 실제로 육성하려면 어떻게 해야 할까요? 무엇보다도 연구자들이 미국의 경제학 및 심리학 연구에서 배우게 된 한 가지는 중등학교 졸업자 이하의 학력을 가진 어머니를 둔 어린이는 인지 시험 점수가 낮다는 사실입니다(그림10). 절대 척도가 인지 점수를 상당히 정확하게 예측해 주기는 하지만, 이는 그다지 중요하지 않습니다. 그러나 갖가지 사회적·경제적 혜택을 누리는 상위 계층 가정 출신과 그렇지 못한 하위 계층 출신 사이에는 커다란 차이가 존재합니다. 18세에 커다란 격차가 있습니다. 그러나 이 표의 가장 중요한 특징은 이 격차가 3세 때에도 존재한다는 점입니다. 따라서 이러한 인지 점수, 시험 점수의 격차, 특권층과 소외 계층 간의 격차 등이 매우 중요해 집니다. 능력주의와 평등주의에 관심 있는 사람이라면 19세에 존재하는 격차가 3세에 등장하고 있다는 점에 주의를 기울여야 할 것입니다.

그림 10

어머니의 교육 수준에 따른 평균 인지 점수 추세

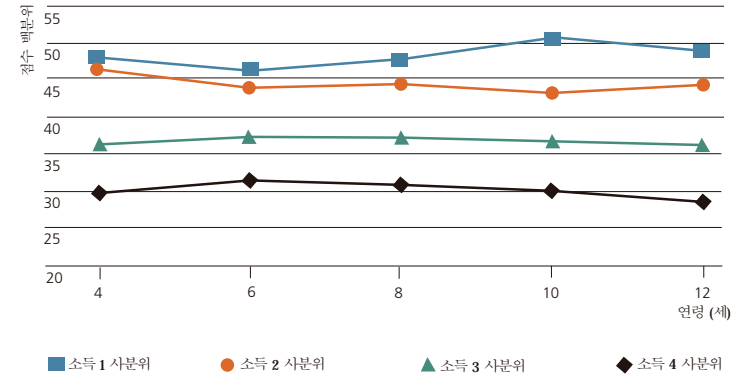


출처: 브룩스 건 등 (2006)

그림 11은 소프트 스킬과 관련한 동일한 종류의 기준을 살펴보고 있습니다. 여기에서는 순서가 반대로 되어 있습니다. 높은 점수는 행동상의 문제가 많다는 것을 의미합니다. 사회적·경제적 혜택이 가장 낮은 가정에서 행동상의 문제가 가장 많이 발생하며, 사회적·경제적 혜택이 가장 높은 가정에서 행동상의 문제가 가장 적게 나타났습니다. 따라서 이러한 스킬들은 단순히 유전적일 수도 있다는 의문이 제기됩니다. 이러한 스킬이 이미 3세나 4세에도 존재한다면, 아마도 이러한 스킬은 부모 혹은 유전자의 결과물에 불과할 수도 있습니다. 하지만 연구자들이 이해하게 된 바로는 이러한 인지적 특성과 비인지적 특성이 유전학 하나만으로 결정되지

그림 11

소득분위에 따른 반사회적 행동 점수의 평균 백분위 점수 순위 (평가 기준-수입 백분위)



는 않는다는 사실입니다. 그런 특성에서 변동성의 약 50%는 유전에 기초를 둔 것으로 판명되었지만, 조기 아동 프로그램에서 가족의 투자는 인지적 기술과 비인지적 기술을 창출합니다.

연구자들은 IQ가 10대에는 거의 변하지 않지만 아주 어린 나이에 변화가 촉진될 수 있다는 것을 알아냈습니다. 학력평가시험에서 측정되는 것은 결정적 지식이고, 결정적 지식은 평생 습득할 수 있습니다. 바로 그것이 학력평가 시험에서 측정되는 것입니다. 연구자들은 또한 인성 기술이 나이를 더 먹을 때까지 훨씬 많이 변한다는 것을 알아냈습니다. 인성 기술을 형성하는 데는 학교와 가정 환경이 중요한 역할을 합니다.

그림 12는 미국 조사에서 발견된 증거를 보여 줍니다. 문제는 교육 기간의 증가가 산술 추론 능력을 얼마나 향상시키는지입니다. 당연한 말이지만, 교육 기간이 길수록 산술 추론의 시험 점수가 높습니다. 단어 지식(그림 13)과 다른 특성에서도 동일한 결과가 나옵니다. 그러나 그림 14에서 볼 수 있듯이 사회·정서적 기술의 평가는 교육의 극적 효과를 보여 줍니다. 따라서 중등교육은 고등학교를 한 해만 더 다닌다 하더라도 자존감과 자신감, 그리고 소프트 스킬의 증가에 실질적인 영향을 미칩니다. 이 부분이 바로 고등학교를 중퇴하고 GED로 졸업장을 받은 사람들이 놓치는 부분입니다. 연구자들은 기술 형성의 역학이 존재한다는 사

실을 알아냈습니다. 풍부한 초기 환경이 사회·정서적 기술을 함양하고 사회의 평등과 생산성을 높일 수 있는 중요한 방법이 되기도 합니다.

이미 나타난 격차가 있다 해도 충분히 어릴 적부터 교육을 시작하면 이 격차를 채울 수가 있습니다. 나는 시카고 대학의 문성혁 군 및 다른 제자들과 함께 약 50년 전에 미시간 대학 근처에서 한 연구진이 개발한 프로그램을 분석했습니다. 페리 프로그램이라고 불리는 이 프로그램의 목적은 저소득 흑인 가정 자녀들의 어린 시절 교육 환경의 질을 높이는 것이었습니다. 프로그램에 참

그림 12

학교 교육이 인지 측정에 미치는 영향 (ASVAB- 군 직업적성시험)

1) 산술 추론

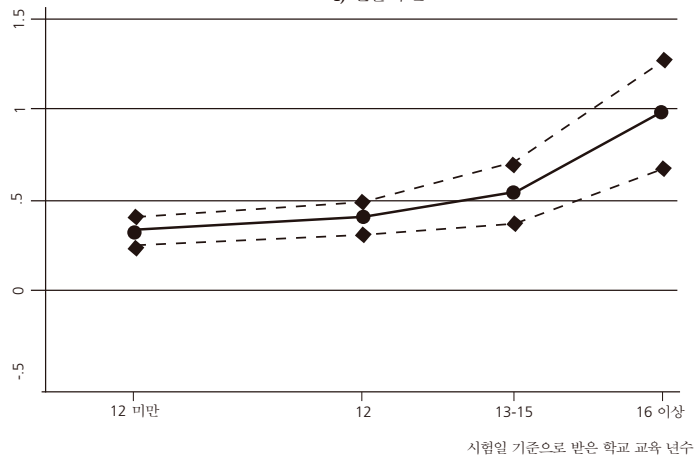


그림 13

학교 교육이 인지 측정에 미치는 영향 (ASVAB- 군 직업적성시험)

2) 언어 지식

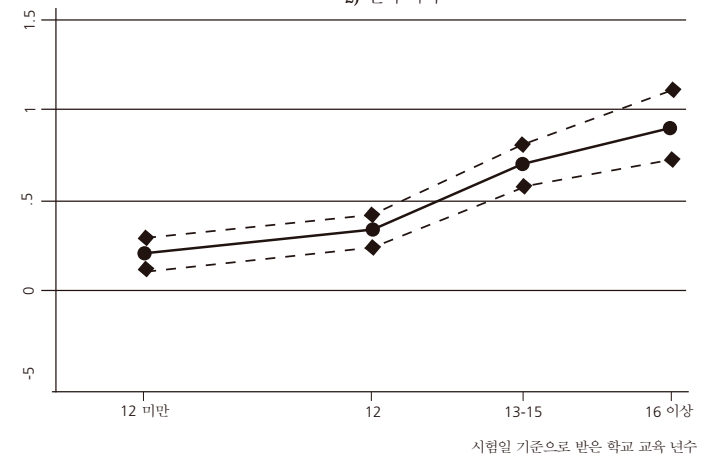
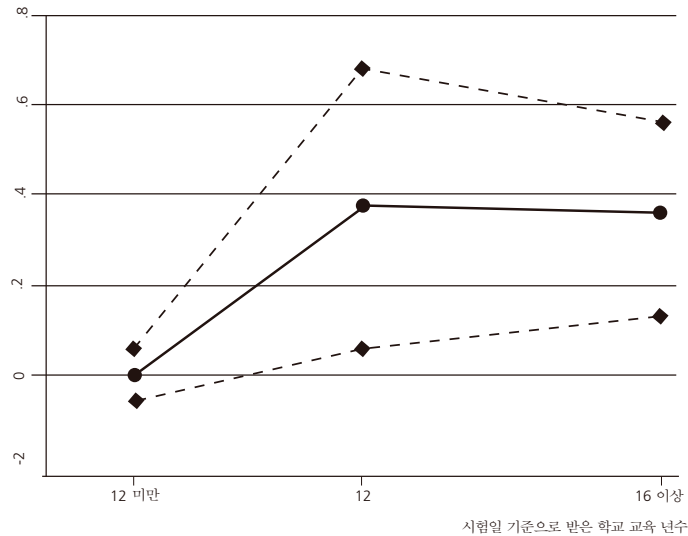


그림 14

학교 교육이 사회·정서 기술의 두 가지 측정에 미치는 영향



가한 모든 어린이의 IQ는 85 이하였습니다. 이 프로그램은 사회적·경제적 혜택을 받지 못하는 어린이들에게 2년 동안 1주일에 5일, 하루에 2시간 30분씩 교육을 제공했습니다. 그렇다면 이 어린이들에게 뭘 가르칠 수 있었을까요? 우리가 교육의 역할을 더 폭넓게 생각해야 하는 이유가 여기에 있습니다. 프로그램 참가자들은 3세 혹은 4세의 아동들이었으며, 따라서 학교에 들어가기도 전이었습니다. 그러나 페리 프로그램은 중단됐습니다. 페리 프로그램은 어떤 역할을 했을까요? 페리 프로그램은 계획을 수립하는 법과 인내력을 가르쳤습니다. 프로그램의 규범을 말 그대로

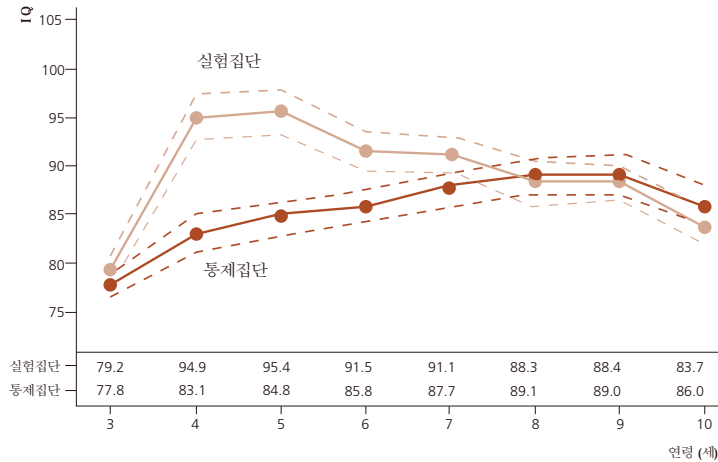
옮기자면 ‘계획하고, 실천하고, 검토하라’입니다. 페리 프로그램은 인성 기술을 가르쳤습니다. 또한 끈기, 분노 조절, 끝까지 임무를 다하는 능력을 가르쳤습니다. 페리 프로그램에서 주목할 만한 점은 어린이들을 두 집단으로 나누어 하나는 실험 집단으로 다른 하나는 통제 집단으로 실험 했다는 점입니다. 실험 집단에는 실제로 프로그램을 적용하였고, 통제 집단에는 프로그램을 적용하지 않았고 임의로 과제를 주었습니다. 그리고 실험에 참가했던 어린이들의 이후 삶을 추적했습니다.

페리 프로그램의 흥미로운 점은 이 프로그램의 경제적 이득이 1년당 7~10%정도로 아주 높다는 것입니다. 그 척도는 무엇이였을까요? 페리 프로그램이 어린이들의 IQ를 높였을까요? 결과부터 말하자면 그렇지 않습니다. 앞서 말했듯이 이 실험은 애초부터 IQ가 낮은 어린이들을 대상으로 한 것이었고, 실험 집단과 통제 집단에서 나타난 초기 결과는 거의 동일했습니다. 그림 15에서 볼 수 있듯이, 프로그램이 진행되는 동안에는 실험 집단의 IQ가 훨씬 높게 나타났습니다. 그러나 8세, 9세, 10세가 되면서 실험 집단과 통제 집단의 IQ차이는 없어졌고, 실험 집단의 IQ는 더 이상 높아지지 않았습니다. 그런데도 이들의 경제적 이득 비율은 1년당 7~10%이었습니다. 이는 제 2차 세계대전 이후 2008년까지 미국의 주식 수익률보다 높은 수치입니다.

페리 프로그램은 학력평가시험의 수행에 영향을 주는 사회·정

그림 15

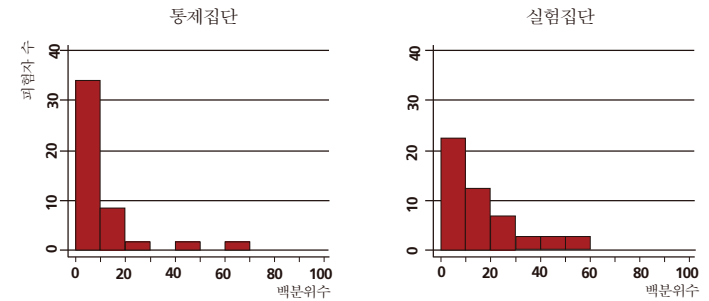
시기별 인지 발달: 남성 인지 역학



서 기술과 여러 성격 요소에 관심을 두었습니다. 학력평가시험 점수가 오르는 이유는 무엇이였을까요? 연구자들은 PISA 시험과 비슷한 CAT (California Achievement Test)를 살펴봤습니다. 점수를 어떻게 올렸을까요? 그림 16에 나온 대로 통제 집단의 성과는 낮았고 실험 집단의 성과는 높았습니다. 앞서 말했듯이 평균적으로 실험 집단과 통제 집단의 IQ는 동일하였지만 소프트 스킬에는 변화가 있었습니다. 실험 집단과 통제 집단 사이의 개인행동을 측정 한 결과는 실질적인 효과를 드러냅니다. 사회· 정서 지수는 실험 집단과 통제 집단 사이의 실제 차이를 보여 줍니다. 페리 프로그램에서 실험의 효과를 일으킨 원인이 인지 요소를 변화시킨 것은 아니지만 비인지 요소, 즉 개인행동과 사회· 정서 상태에 변화

그림 16

페리 프로그램 대상자의 14세 때 CAT 점수



CAT = California Achievement Test

실험집단: N = 49; 통제집단: N = 46

남성과 여성 모두 통계적으로 유의미한 효과 발견 (p - 값이 각각 0.009, 0.021)

출처: 헤르먼, 말로피바, 핀토, 사브예프 (2010)

를 일으킨 것입니다. 이는 통계적으로 봤을 때, 범죄를 줄이고 장기적인 성공을 촉진하는 유의미한 결과를 가져왔습니다.

연구자들은 이런 특성이 변경될 수 있다는 것을 알게 됐습니다. 또한 IQ와 유동적 지능에 대한 투자는 유아기에 하는 것이 가장 생산적이라는 사실도 알게 됐습니다. 반면 어린이가 조금 더 자란 후에는 인성을 육성하는 투자가 더 생산적이며, 이는 아동 발달적 측면과 관계가 있습니다.

그렇다면 이 실험으로부터 우리는 무엇을 배울 수 있을까요? 오늘날 교육의 정의, 교육 정책의 올바른 역할, 교육을 평가하는 방

법에 대한 지식은 훨씬 풍성해 졌습니다. 교육 성취도는 비인지적 기술보다 인지적 기술에 훨씬 많이 의존하는 것이 사실이지만, 비인지적 기술도 나름의 역할을 수행하고 있습니다. 인성은 인지적 기술을 육성하도록 돕습니다. 그 이유는 무엇일까요?

어린이가 학습 동기를 부여 받으면, 보다 많은 지식을 습득하게 됩니다. 이는 페리 연구에서 입증되었습니다. 페리 프로그램에 참가한 어린이들은 특별히 똑똑하진 않았지만 보다 많은 학습 동기를 가지고 있었습니다. 때문에 14세에 학력평가시험을 봤을 때 훨씬 많은 지식을 체화한 결과를 보여 주게 된 것입니다. 게다가 그들은 범죄, 이혼, 그 외 많은 다른 부분에서도 훨씬 나은 결과를 보여 주었습니다. 이는 우리로 하여금 풍부한 스킬에 관심을 갖게 합니다. 스킬은 단순히 어떤 한 가지만을 의미하는 것이 아니고, 인적 자본 또한 단순히 IQ만이 아닙니다. 그리고 교육도 단순히 인지에 대한 것만은 아닙니다.

현재 한국 사회는 인지 시험에 전적으로 초점을 맞추느라 사회적 수행이라는 중요한 차원을 무시하고 있습니다. 소프트 스킬이 중요합니다. 인지적 기술과 비인지적 기술 모두 유전적으로 정해진 것이 아닙니다. 이런 기술은 심지어 청소년기나 성인기에도 형성될 수 있습니다. 이런 기술을 향상시키는 것이야말로 사회 정책의 생산성을 높이는 길입니다.

앞서 설명한 GED 프로그램은 소프트 스킬의 중요성을 보여 줍니

다. 이는 정신력을 측정하는 교육 기관들의 기준을 제대로 만족시키는 시험입니다. GED 취득자의 지적 능력은 일반적인 고등학교 졸업자와 마찬가지로이지만, 그들의 수행 성과는 중등학교 중퇴자의 수준입니다. GED는 상당한 왜곡을 유발하기도 합니다. 저는 미국 경제와 관련해 왜곡된 기본 데이터, 즉 고등학교 졸업률이 얼마나 과장되었는지를 언급한 바 있습니다. 이는 이곳 한국 사회에서도 엄청난 왜곡을 낳을 수 있습니다.

낙제방지 프로그램 때문에, 결과적으로 미국 학교는 물리학이나 기타 다양한 부분을 가르치기 보다는 수학과 읽기에만 초점을 맞추고 있습니다. 시험 점수 향상에만 초점을 두는 정책은 성공에 근본적으로 필요한 것이 무엇인지에 대한 요점을 놓치고 있습니다. 저는 사회 정책을 수립할 때에는 정책의 목표(교육 정책, 인지 및 인성)가 무엇인지에 대한 것뿐 아니라 전인(全人)의 의미에 대해서도 깊이 생각해 보아야 하며, 전인격적 특성이 측정될 수 있다는 사실에 대해서도 인지하고 있어야 한다고 생각합니다. 이것은 매우 중요한 결론입니다. 오늘날 우리는 시험 성적, PISA 성적, IQ 점수 등만 바라보는 교육 정책에 일차원적으로 초점을 맞추는 사회적 경향에서 벗어날 수 있게 돕는 도구를 많이 가지고 있습니다. 근본적으로 이러한 척도에 초점을 두는, 그러니까 본질적으로 학력평가시험이 무엇을 놓치고 있는지를 잘 설명해주는 정책이 미국 사회뿐 아니라 한국 사회에서도 보다 큰 성공을 거둘 것입니다.

Chapter 2 질의 응답

질문1_ 우리 한국인은 표준화된 시험에 대한 과도한 의존이 위험하다는 것을 잘 알고 있습니다. 그러나 그러한 함정에서 어떻게 벗어나야 할까요? 알려진 다른 모든 방법, 그러니까 어린이의 능력이나 수행성과를 측정하는 방법들은 주관적인 기준의 도입을 허용하고 있습니다. 그렇다면 그런 능력을 표준화된 시험 외에 객관적인 방식으로 측정할 수 있는 방법이 있나요?

제임스 헤크먼_ 물론입니다. 앞서 말했듯이, 한국 사회에서 나타나고 있는 현상과 같은 종류의 움직임이 100년 전 미국 사회에도 있었습니다. 존 듀이나 그 밖의 교육자들은 어린이를 선별하기 위해, 또 사회적·경제적 혜택을 받지 못하는 사람에게 교육 받을 기회를 주어 교육의 기반을 넓히기 위해, 표준화된 시험의 사용을 촉진하고자 했습니다. 그러나 우리는 여기에서 두 가지 요점을 인식해야 합니다. 제 생각에 표준화된 시험의 진짜 목적은 성공할 학생과 그렇지 못할 학생을 예측하는 것입니다.

예를 들어, 정제되지 않은 측정법, 그러니까 주관적인 측정에 대해 생각해 보세요. 많은 사람들이 성적은 매우 주관적으로 매겨진다고 이야기하긴 합니다. 하지만 이런 시험이 도대체 무엇인가요? 그리고 이런 시험은 실제로 무엇을 측정하는 걸까요? 심리 측정 저서를 보면 시험의 타당도라 불리는 것이 있습니다. 저는 앞서 임금이나 근무 시간 같은 실제 데이터에서의 타당도가 무엇을 의미하는지 그 예를 제시한 바 있습니다. 우선 상관관계가 0.2

라는 값을 반영하는 경우에도 예측력은 여전히 낮다는 점을 보여 주었습니다. 이는 매우 강력한 상관관계라고 할 수는 없습니다. 변동성의 일부만을 설명할 뿐입니다.

그러나 이러한 주관적인 측정으로도 더 나은 성과를 낼 수 있습니다. 이런 측정법은 그 자체만의 조건 속에서 작동하지는 않으며 시험 개발자에게 특별한 책임의식이 있는 것도 아닙니다. 프린스턴의 ETS (Princeton Educational Testing Service), ACT (American Council on Testing), ACE (American Council on Education) 등은 계량적이고 객관적인 방법으로 만들어 졌기에 공식화될 수 있을 만큼 충분히 합리적이라고 믿어 지곤 합니다. 그러나 앞서 제가 설명 드렸듯이, IQ 테스트의 개발자들, 학력평가 시험의 개발자들, 심지어 표준화된 시험을 처음 주창한 인물들 까지도 모두 훨씬 풍성한 다른 데이터에 의지할 수도 있다는 점을 강조했습니다. 일부 데이터는 주관적인 의견에 불과하겠지만, 어린이의 전 생애에 걸쳐 주관적인 의견이 축적됐다고 생각해 보십시오. 교사가 한 반, 한 과목, 혹은 한 학년의 어떤 학생에 대해 편향된 의견을 가질 수는 있습니다. 그러나 그런 의견이 축적되고, 이렇게 축적해 나가는 괜찮은 방법이 존재한다면, 해당 어린이의 수행 성과 결과가 도출될 수 있게 됩니다. 앞서 제가 여러분에게 제시한 성실도 측정과 페리 연구에서 사용된 측정은 실제로 성실도를 측정한 것입니다. 이러한 측정은 교사가 어린이에게, 얼마나 열심히 공부했는지, 공부를 얼마나 잘했는지, 다른 학생과는

얼마나 협력했는지 등을 말 그대로 직접 물어 봐서 평가하는 방식으로 이루어졌습니다. 수년에 걸쳐서 수많은 교사가 이 과정을 거쳤고, 따라서 주관성을 많이 제거하는 데 도움을 주었습니다.

따라서 저는 두 가지 측면이 있다고 생각합니다. 첫째, 당연한 이야기이지만, 예측을 잘 하는 측정이 가장 바람직합니다. 그런데 사람들의 성공 가능성을 임금, 근무 시간, 삶의 결과에서 3%, 4% 혹은 5%의 가변성만을 설명해주는 시험으로 판단하는 것이 능력주의일까요? 나머지 95%는 어떤가요? 다른 특성들을 살펴보면, 전부는 아니더라도 상당한 부분을 설명할 수가 있습니다. 둘째, 심리학에서 혁신적인 발전이 이루어졌습니다. 최근의 많은 연구는 이러한 성격적 특성들이 비교적 안정적인 요소이며, 장래를 매우 잘 예측할 수 있고, 또한 측정될 수 있다는 점을 보여 주고 있습니다. 어떻게 하는 것일까요? 이는 다소 편향적이라고 할 수 있는 자기 보고 방법뿐 아니라 학생의 행동에 대한 교사의 평가와 제 삼자의 평가를 통해 가능합니다. 따라서 방법은 많다고 할 수 있습니다. 1940년의 상황은 분명 현재와 달랐습니다. 저는 이제 컴퓨터화된 정보와 데이터베이스를 사용하여 성공적인 인성 목록도 만들 수 있을 거라고 생각합니다. 어린이 행동에 대한 기록 또한 존재합니다.

요점을 말하자면, 제가 과거의 시험을 대신할 새로운 시험을 옹호하고 있는 것은 아니라는 겁니다. 저는 인간 기술의 복잡성과 다양

성에 대해 말씀 드리는 것이며, 모든 어린이에 대한 풍성한 데이터 베이스를 만들 역량이 된다는 사실을 인식해야 한다고 이야기하는 것입니다. 저는 이것을 진정한 능력주의라고 생각합니다.

질문2_소프트 스킬의 측정이 중요하다고 강조하셨습니다. 그러나 소프트 스킬을 측정할 방법을 설명하지는 않으신 것 같습니다. 여기에 대해 간단히 설명해 주시겠습니까?

제임스 헤크먼_알겠습니다. 지금까지 많은 목록이 나와 있습니다. 이를 굳이 혁신이라고 부르지는 않겠지만, 성격심리학은 분명 지난 20~30년 동안 상당한 발전을 거듭했습니다. 학생의 목록을 기록하는 데에는 다양한 방법이 있습니다. 예를 들어 한 어린이가 얼마나 집중할 수 있는지에 사용할 수 있는 목록이 있습니다. “실행 기능화”라 불리는 어떤 검사는 “어떤 학생이 주의가 산만해 지지 않고 얼마나 잘 임무에 집중할 수 있는지”를 보여 줍니다. 스트룹 검사(Stroop test)를 비롯한 비슷한 종류의 다른 검사들도 있습니다.

다양한 연령의 어린이에게 과제를 주는 방법으로도 과제 집중력을 측정할 수 있습니다. 어린이에게 이런 저런 자료를 주고 얼마나 잘 정리하는지를 보거나 어린이가 제시된 과제에 얼마나 지속적으로 집중할 수 있는지를 보는 것입니다. 이것이 실행 기능화의 한 측면입니다. 또 어린이가 다른 사람과 어떻게 협력하는지의 관점에서 아이가 무엇을 하고 있는지의 목록도 볼 수 있습니

다. 기록으로서의 교사 보고서와 평가서도 있습니다. 많은 학교가 그런 기록을 보관하고 있는데, 특히 미국이 그러합니다. 우리는 그런 기록을 측정할 수 있고, 그런 기록은 학교 등에서의 성공에 대한 예측력이 매우 높은 것으로 판명되고 있습니다.

이것은 어린이의 행동에 대한 다른 사람의 평가와 어린이의 수행에 대한 객관적인 검사가 어우러진 집합체입니다. 각종 시험을 설명하는 책이나, 오늘 발표한 것과 같은 논문, 그리고 참고 자료가 한도 끝도 없이 많습니다. 소프트 스킬의 측정은 응시자들에게 x 값을 찾기 위한 방정식을 풀라고 요구하는 GED 같은 시험이 아닙니다. 대신, 어린이가 다른 어린이와 얼마나 잘 협력하는지에 대해 묻습니다. 한 교사의 주관성이 좀 개입된다고 하더라도, 다른 교사들 의견의 평균을 내고 어린이에게 교사를 선택할 수 있는 공정한 기회를 주거나, 혹은 적어도 학교 제도의 틀 안에서 어린이를 유심히 지켜 본다면 사적인 기록을 만들어 낼 수 있습니다. 이런 기록은 특정한 학생의 행동과 특정 과제에 대한 수행성과의 조합이라 할 수 있습니다.

보다 구체적인 예를 들어 보겠습니다. 미국 제조업에서 많이 사용되는 마이어스-브릭스 검사(Myers-Briggs Test)가 있습니다. 사실 마이어스-브릭스 검사는 매년 SAT 시험보다도 많이 판매됩니다. 마이어스-브릭스 검사는 사람들에게 그들의 인성과 태도 등에 대하여 묻습니다. 이 밖에 미국의 많은 기업이 사용하는 호

건 인성 목록 검사(Hogan Personality Inventory Test)도 있습니다. 객관적인 질문이나 응시자의 태도 등에 관한 질문이 주어집니다. 이런 검사가 미래의 성공을 얼마나 잘 예측할까요? 기업은 이런 검사를 구입합니다. 매년 이런 검사가 수백만 회 실시되며 고용 결정, 승진 결정, 사원 모집 및 선별을 위한 결정 과정에서 사용되고 있습니다. 따라서 이런 검사는 확실히 성공적이라 볼 만 합니다. 실제로 이런 검사는 기업과 개인의 관점에서 보았을 때 예측력이 있습니다. 때문에 인성, 즉 빅 파이프를 평가하기 위해 창설되는 학문 분야도 있습니다.

과거 성격심리학에서는 각각의 심리학자가 소프트 스킬과 관련된 특성이 무엇인지에 대해 각자 나름대로의 신념을 지니고 있었습니다. 그래서 한 심리학자의 연구 결과와 다른 심리학자의 연구 결과를 비교하기가 아주 어려웠습니다. 그러나 약 30년 전에 빅 파이프라고 불리는 것으로 의견이 수렴되었습니다. 제가 OCEAN이라고 여러분께 소개한 것이 바로 이 빅 파이프의 머리글자를 따온 것입니다. 많은 연구가 빅 파이프의 예측력 즉, 어떻게 성실성이 건강을 증진시키는가, 그리고 왜 신경증적인 경향이 일부 직업에서는 생산적이고 일부 직업에서는 비생산적인가를 살펴보았습니다. 외과의에게 신경질적인 성격은 장점입니다. 외과의는 아주 강박적이고 성공하기를 절실히 원합니다. 때문에 대단히 불안해 하고 임무 중심적입니다. 그러나 다른 분야에서는 신경질적인 성격이 커다란 단점이 됩니다.

일련의 시험, 선별 장치, 보고서뿐만 아니라 성과와 예측력을 보여 주는 결과도 많이 있습니다. 저는 측정 척도가 존재한다는 것과 그런 척도들의 예측력이 뛰어나다는 점을 이야기하고 싶습니다. 그리고 그런 척도가 어떻게 많은 경우에 보다 훌륭한 예측 요인이 되는지를 앞에서 설명했습니다. 따라서 이러한 특성에 대해 더욱 폭넓게 생각할 여지가 많이 남아 있습니다.

질문 3_ 이런 연구를 통해 정책적 차원에서 달성하고 싶은 무언가가 있다면 그것은 무엇인가요? 연방 정부나 지방 정부 차원의 정책 입안자에게 제안하고 싶은 점이 있으신가요? 그리고 한국의 정책 입안자에게 가장 우선적으로 제안하고 싶은 점이 있다면 무엇인가요?

제임스 헤크먼_ 제가 이 모든 연구를 하는데 있어서의 목표는 불평등과 빈부 격차의 문제를 해결하는 것입니다. 문제는 사회적·경제적 혜택을 못 받는 어린이의 삶을 증진시킬 방법이 무엇이나 하는 것입니다. 저는 오늘날 미국을 비롯한 세계 여러 나라 사회 정책의 많은 부분에 유전자나 IQ 같은 특성이 성공을 예측하는 데 중요하게 작용한다는 강한 믿음이 존재한다고 봅니다. 약 15년 혹은 17년 전에 찰스 머리(Charles Murray)와 리처드 헤른스타인(Richard Herrnstein)이 쓴 《중형 곡선(The Bell Curve)》이라는 책이 있습니다. 이 책은 흑인과 백인 사이의 불평등을 IQ로 설명하려고 했으며, 본질적으로 IQ는 유전적으로 결정되는 것이라고 주

장했습니다.

제 연구 프로그램의 목적은 인생에서 성공하기 위해서는 인지 능력 그 이상의 특성들이 필요하다는 점을 사람들에게 이해시키는 것입니다. 그 특성은 단순한 IQ 이상의 것입니다. 앞서 저는 헤른스타인과 찰스 머리가 사용한 학력평가지험의 예측력이 인성 특성을 통한 예측력과 마찬가지로였다는 점을 보여 드렸습니다. 이를 굳이 강조하지는 않았지만, 앞에서 보여 드린 그래프 중 하나에



설명되어 있습니다. 인성은 매우 큰 역할을 수행했습니다. 따라서 IQ의 측정조차도 실제로는 인성의 일부 측면을 기술한 것이라 할 수 있습니다. 우리 연구자들의 목표 중 하나는 이런 검사가 의미하는 점과 부족한 점을 제대로 이해하게 하는 것입니다. 두 번째 연구 목표는 인지와 인성은 둘 다 측정될 수 있는 성질의 것이며, 우리 연구자들이 실제로 인지와 인성을 정확하게 측정하는 방법을 찾아낼 수 있다고 말하는 것입니다. 세 번째 목표는 이러한 특성이 유전적인 것이 아니라 나중에 형성될 수 있다는 것을 보여주는 것입니다. 사회적·경제적 혜택을 받지 못하는 어린이와 성장기에 가족에게 제대로 된 지원을 받지 못한 사람들의 인지적 특성과 비인지적 특성은, 혜택을 받고 성장한 사람들의 경우와 비교해 보았을 때 큰 차이를 보입니다. 그러나 그런 특성은 변화될 수 있습니다. 우리는 이에 대해 무언가 행동을 취할 필요가 있습니다. 교육 정책 입안자는 보다 많은 특성에 대해 폭넓게 생각해야 하며, 스킬 형성의 역학을 이해해야 합니다. 뿐만 아니라 스킬이란 것은 다른 스킬을 기반으로 발전된다는 점과 미리 형성된 특성들은 이후 또 다른 특성의 발달을 촉진한다는 점에 대해 이해해야 합니다. 페리 프로그램에 참가한 어린이들을 보면, 특별히 똑똑하지는 않았지만 학습 동기는 높게 나타났습니다. 따라서 14세가 되어서 지식을 측정하는 시험을 봤을 때에도 훨씬 좋은 성적을 보였습니다. 이는 본질적으로 사회 정책에 대한 사고방식을 바꾸어 놓습니다. 그리고 더욱 많은 측정법이 사용됩니다. 이러한 특성들은 조기에 형성될 수 있으며, 척도와 측정법의 범위를 넓

히고 스킬 형성의 역학을 어떻게 펼칠 것인가에 대해 생각해 봄으로써 우리는 더 많은 이익을 창출해 낼 수 있습니다.

결국, 이것이 전달하고자 하는 메시지는 보다 포괄적인 척도가 필요하다는 점입니다. 부수적인 메시지로는 시험 성적에만 초점을 두는 것은 교육의 본질을 왜곡하고 있다는 것입니다. 이는 한국의 실정을 반영하는 교훈이며, 정책적인 면에 더욱 적용되어야 합니다. 저는 강연에서 여러 아시아 사회의 전통 가치관인 유교 교육에 대해 이야기했습니다. 유교 교육에서는 시험 점수만이 전부는 아니었습니다. 인지라는 관점에서 단순한 수행 능력뿐 아니라 다른 사람을 대하는 방법, 존경심, 성실성, 인내심 같은 전 범위를 중요시했습니다. 시험 점수에만 초점을 맞추면 이런 많은 측면이 무시됩니다. 그런 의미에서 제가 제안하고 싶은 점은 교육 정책을 확대하는 것입니다. 한국이나 칠레, 혹은 다른 나라의 성패는 어린이 삶의 부분적인 면과 교육 체제의 건전성만을 측정하는 PISA 점수에 따라 평가되는 것이 아닙니다. 교육은 스킬의 벡터, 스킬의 폭넓은 집합체를 총체적으로 생산하는 것으로 생각되어야 하며, 저는 많은 국가의 현 교육 정책이 그렇게 하지 못하고 있다고 생각합니다. 미국은 확실히 그렇지 못하며 제 생각에 한국도 마찬가지인 듯합니다.

질문 4_ 교수님의 연구는 취학 전 아동을 위한 정부의 교육 사업, 그러니까 헤드스타트(Head Start)와 같은 프로그램을 지지하시나요?

제임스 헤크먼_ 헤드스타트 자체는 불완전하게 시행되고 있습니다. 헤드스타트의 일부는 페리 미취학 아동 프로그램과 유사한 형태입니다. 단, 헤드스타트는 이미 시행되고 있는 프로그램이며 페리 프로그램에 비해 미흡한 재정 지원을 받는다는 점만 다릅니다. 페리 프로그램에 비해서 집행력도 많이 부족합니다.

미국의 많은 어린이들은 편부, 편모 슬하에서 자라고 있습니다. 문성혁 군의 연구는 편부, 편모 가정의 인지적 자극과 비인지적 자극 측면에서 일반 가정보다 적은 자원을 제공하고 있다는 점을 보여 줍니다. 이는 어린이의 전 생애에 축적되는 효과를 지니고 있습니다. 이런 체계가 조기에 형성되고 시작될 수 있다면, 문제가 발생한 후에 치료하는 것보다는 그런 문제가 발생하기 전에 미리 방지하는 것이 더 수월하게 됩니다. 따라서 저는 가족과 함께 하는 초기의 몇 년에 대해 다시 한 번 방점을 찍고 싶습니다. 특히 전래되는 가치관에 대해 제대로 배우지 못하는 사회적·경제적으로 불우한 어린이를 위해서라면 더욱 그러합니다. 이렇게 되면 정책 토론의 초점도 변화할 것입니다. 학교에서의 수행 능력에 아주 중요하게 작용하는 이러한 스킬은 사실 취학 전에도 생길 수 있다는 점을 이해하는 것이 중요합니다. 이는 교육의 개념을 확장시킵니다. 따라서 이는 헤드스타트와 유사한 것이지만, 헤드스타트의 농축된 하나의 유형 정도로 볼 수 있을 것입니다.

질문 5_ 정책 입안의 관점에서 볼 때 딜레마가 하나 있습니다. 바로

구체적인 정책이나 중재책, 그리고 소프트 스킬을 효과적으로 증진할 만한 도구를 찾아내기가 어렵다는 것인데, 특히 학생의 나이가 많을수록 이것이 어려워 집니다. 어린이의 소프트 스킬을 높이기 위해서 아주 중요하다고 알려진 몇 가지 요소가 있습니다. 훌륭한 육아법이나 좋은 가정환경 등이 그것입니다. 그러나 정책의 관점에서 볼 때, 그런 부분은 구체적인 정책이나 도구, 프로그램에 영향을 주기가 쉽지 않습니다. 사회적·경제적 혜택을 받지 못하는 어린이의 소프트 스킬을 향상시킬 수 있는 구체적인 정책이나 도구 혹은 프로그램의 예를 몇 가지 들어 주시겠습니까?

제임스 헤크먼_ 한 가지는 이미 언급했습니다. 바로 페리 미취학 아동 프로그램입니다. 그러나 다른 프로그램도 많이 있으며, 활발히 사용되기도 했습니다. 현재 제가 주시하고 있는 프로그램은 1970년대에 노스캐롤라이나에서 시행된 ABC 프로그램입니다. 이 프로그램은 태어날 때부터, 혹은 생후 8~9주 때부터 사회적·경제적 혜택을 거의 받지 못하는 어린이를 대상으로 시작됐으며 8년 동안 지속되었습니다. 간호사-가족 협력 프로그램처럼 개입에 성공한 프로그램도 있습니다. 이와 같이 임의적 실험을 통해 장기적으로 평가된 성공적인 중재책이 많이 있습니다. 이런 프로그램은 양육 기술의 향상, 부모와 자녀의 애정도 증가, 성공적인 학교 생활 장려, 학업 성취도 신장뿐 아니라 전반적인 성취도를 증진시키면서 제삼자가 부모나 어린이 혹은 양측 모두에 개입하는 것이 가능하다는 사실을 보여 주었습니다.

따라서 우리는 효율적인 개입이 가능하다고 말할 수 있습니다. 사실 이는 전통적으로 측정된 빈곤만의 문제는 아닙니다. 빈곤에 대한 최악의 척도, 가장 불완전한 척도 중의 하나는 빈곤을 단순히 돈 문제로 생각하는 것입니다. 어린이의 복지 향상을 위해서는 단순히 돈이 아니라 양육의 질이 중요하게 작용합니다. 각종 프로그램은 개입을 통해 양육의 질이 향상될 수 있다는 것을 보여 주며, 부모의 참여도 가능합니다.

현재 여러 프로그램이 실시되고 있으며, 일부 프로그램은 장기적 평가를 지속해 오고 있습니다. 또한 자제심과 같은 특성을 가르치는 많은 프로그램이 현재 제안되거나 평가되고 있습니다. 세계적으로 여러 국가에서 활발하게 사용되고 있는 정신의 도구(**Tools of the Mind**)라는 프로그램이 있습니다. 이 정신의 도구는 프로그램은 자제심과 분노 조절 능력을 가르칩니다. 앞서 제가 설명 드린 페리 프로그램이 정신의 도구에서 몇 가지 원칙을 차용했다고 생각하셔도 무방합니다. 정신의 도구라는 프로그램은 최근에 적용됐기 때문에 평가 결과가 다소 단기적이지만 칠레, 미국과 유럽을 비롯한 전 세계에서 연구되고 있습니다. 임무에 오래 집중하는 법, 친구들과 상호작용을 잘하는 법 등을 가르치면 어린이의 역량이 증가한다는 결과가 입증됐습니다. 따라서 저는 이러한 프로그램에 대한 이해가 이미 어느 정도 존재한다고 생각합니다.

저는 이런 프로그램들이 개선될 수 없다고 말하는 것이 아닙니다. 모든 사회 정책은 개선될 수 있습니다. 정책의 개선이 없다고 생각하면 오산입니다. 시행되고 있는 정책 중에서 아주 성공적인 정책도 많습니다. 그런 정책은 스킬을 향상시키기 위한 방법에 대해 일종의 약속을 내놓고 있습니다. 앞으로 그런 정책은 교육을 개선시킬 것이고, 다양한 차원에서 성공을 촉진할 것입니다.

질문 6_ 저는 소프트 스킬과 두 개 이상의 언어를 구사하는 능력과의 관계에 대해 질문하고자 합니다. 미국 엄마들 사이에 자녀를 중국에 유학 보내서 중국어를 가르치는 열풍이 불고 있고, 이런 현상이 ‘타이거 맘(Tiger Moms)’으로 표현되곤 합니다. 그런 양육 혹은 교육 방법이 소프트 스킬의 개발에 정말 도움이 된다고 생각하시나요?

제임스 헤르먼_ 타이거 맘은 그 자체로 하나의 현상입니다. 효율적으로 자녀를 양육하면서 자녀의 길을 이끌어 주는 부모와 과도하게 아이를 지배하는 부모 사이에는 차이가 있습니다. 예컨대 타이거 맘을 주제로 한 토론에서도 바로 이런 차이가 도출되었다고 봅니다.

양육의 역할이 아주 중요하다는 점은 잘 알려져 있습니다. 그런 말은 반복할 필요도 없습니다. 그러나 요점은 많은 사회에서 수많은 어린이가 결손 가정에서 자라고 있다는 점이며, 저는 최근 한국 사회에서도 그런 현상을 목격하고 있습니다. 물론 한국 사

회는 미국 사회보다 결혼 가정의 문제가 덜 심각하고, 편부, 편모 가정의 수가 비교적 적지만, 이러한 가정의 수가 꾸준히 증가하고는 있습니다. 집단 실험을 통해 알려진 바로는 이러한 환경에서 아이들은 자극을 덜 받게 됩니다.

이를 측정할 직접적인 방법이 있습니다. 예를 들어, 문성혁 군은 자녀에게 주어지는 격려의 양에 관심을 기울이면서 이에 대한 연



구를 수행해 왔습니다. 예컨대 일방적으로 자녀들에게 동물원에 가라고 지시하는 것이 아니라 자녀를 동물원에 데리고 가는 것, 자녀에게 책을 읽어 주는 것과 같이 여러 가지 방법으로 자녀를 격려하는 것에 대해 연구하는 것입니다. 저는 여러분이 어떠한 것을 과도하게 할 수도 있다는 점을 알고 있습니다. 그러나 여러분은 효율적인 부모를 가질 수도 있고 이러한 자원을 제공해 주는 효율적인 양육을 실천에 옮길 수도 있다고 생각합니다. 저는 중요한 것은 바로 이러한 자원이며, 이러한 자원은 측정될 수 있다고 생각합니다. 타이거 맘은 아이들을 숨 막힐 만큼 억제함으로써 아주 강한 반발심을 유발하기도 합니다. 그러나 한편으로는, 시카고나 미국의 다른 대도시처럼 범죄가 빈번하게 일어나는, 지구상의 혜택 받지 못한 많은 가정환경과 비교해 볼 때, 아예 엄마가 없거나 엄마가 있더라도 양육 기술이 부재한 경우보다 타이거 맘이 있는 환경이 훨씬 나은 대안이라고 생각합니다.

그러나 저는 그 중간 수준의 환경이 필요하다고 생각합니다. 타이거 맘이라고 비난 받았던 한 중국계 미국인 어머니는 미국의 한 주요 신문 논설에서 과도한 통제를 하고 있는 자와 풍요로운 환경을 제공하는 자의 차이점을 분명히 설명했습니다. 저는 양육이 아주 어려운 일이라고 생각합니다. 양육이 어렵고 중요한 일인데도 양육에 실질적으로 도움이 되는 책은 그다지 많지 않습니다. 그러나 자녀와의 상호작용을 통해서라면 양육에 성공할 수 있습니다. 어머니가 특히 아이가 어린 시기에 아이와 관계를 가

지고 아이를 격려하는 것은 무척 중요합니다. 이것은 아주 중요한 기본 원칙입니다

질문 7_ 우리 한국인들은, 한국의 수학 수준이 33개국 중 1위라든지 쓰기수준이 2위라는 이야기를 종종 듣습니다. 이런 측면에서 볼 때, 국제기관은 좋은 교육이 무엇인지 혹은 어떤 국가가 학교 수행능력에 있어 몇 위를 하고 있는지의 개념을 조장하고 있으며, 각국 정부 또한 그런 순위를 국가의 위상과 교육 수준을 평가하는 기준으로 삼고 있습니다. 국제기관이 제안한 정책 조언이나 구상 중에 성공에 영향을 미치는 비인지적 특성의 중요성을 다루는 내용이 있나요?

제임스 헤크먼_ 성적의 객관적인 측정이 중요하다는 분위기가 조성된 데에는 많은 경제학자에게 전적인 것은 아니지만 분명 부분적인 책임이 있고 그들이 주요한 역할을 해 왔다는 점에 저도 동의합니다. 앞서 저는 베트남 전쟁의 사상자 수를 산출한 발상에 대해 특히 극적이었다고 언급한 바 있습니다. 많은 사람이 양적으로 평가할 수 없는 것은 가치가 없다는 강한 믿음을 가지고 있습니다. 저 역시 그렇게 생각하는 경향이 있습니다. 그러나 다른 한편, 중앙집권화된 체제에 대한 각종 연구에서 볼 수 있듯이 사회의 자원과 인센티브가 잘못 이용되면 사람들은 더 많은 잘못을 저지르게 될 수 있습니다.

유명한 예가 하나 있는데, 주제에서 조금 벗어나지만 인센티브와 관련된 이야기입니다. 과거에 소련의 한 공장에 무게가 1,000파운드에 달하는 못이 있었습니다. 당시 소련의 공장은 일정량의 톤 혹은 킬로그램의 못을 생산하면 인센티브를 주었습니다. 어떤 공장은 그래서 무게가 1,000파운드나 되는 못을 만들었는데 그것이 바로 할당량이었기 때문입니다. 그들이 유용하게 쓰일 수 있는 못을 원하였는가는 문제가 아니었습니다.

인센티브는 중요하며, 잘못된 측정은 왜곡을 일으킵니다. 예를 들어, 기본적으로 학교와 교사의 성과를 측정할 목적으로 시작된 낙제학생방지법에서 교사가 시험에 부정행 방법을 동원했다는 스캔들에 대한 비난이 꼬리에 꼬리를 물고 이어지고 있습니다. 이는 교사와 교장, 그리고 학교가 시험 성적을 향상시키기 위한 자금 지원을 받고 있기 때문입니다.

잘못된 측정 혹은 왜곡될 가능성이 있는 측정은 실제로 활동을 엉뚱한 방향으로 유도할 수 있습니다. 예를 들어, 저는 수년 전에 레이건 행정부에서 연구를 수행한 적이 있습니다. 당시에 직업훈련협력법(Job Training Partnership Act)이라는 것이 있었습니다. 직업훈련협력법에는 노동 인력, 즉 성인 혹은 최소한 성년에 가까운 청소년을 훈련하는 프로그램이 포함되었습니다. 교육 업체는 프로그램 참가자의 수행 성과와 취업 성과에 기반해 자금을 지원 받았습니다. 이 경우 최종 측정 기준은 취업률, 즉 시험 점수

가 아니라 일자리였습니다. 따라서 교육 업체는 일단 좋은 결과가 나오게 하기 위해 프로그램 참가자를 선별했습니다. 이는 ‘크림 스키밍’이라 불리고 있습니다. 둘째, 그들은 프로그램에서 낙제한 참가자들을 사실상 몇 년 동안 프로그램 일지에 보존해 두어 새롭게 참여할 수 없게 함으로써 늘 성공적인 결과를 도출해 냈습니다. 많은 정부 관료는 인센티브에 대해 보이는 인간의 세밀한 반응을 제대로 이해하지 못합니다.

PISA 시험이나 낙제학생방지 시험과 같은 시험 성적에 대한 오래된 농담이 있습니다. 다른 데서 열쇠를 잃어버리고서 가로등 아래 와 열쇠를 찾는 취객에 관한 농담입니다. 잘못 사용된 인센티브는 아예 없는 경우보다도 더 나쁠 수 있습니다. 제가 제안하고자 하는 것은 그렇게 극적인 것은 아니고, 다만 부분적으로 극적인 것입니다.

저는 지금은 인간의 성취라는 한 측면에만 초점이 대체로 맞추어져 있지만, 꼭 고려되어야 함은 물론이고 인생의 성공에 대해 높은 예측력을 지닌 다른 중요한 차원이 많이 존재한다는 점을 말씀드리고 있는 것입니다. 그런 의미에서, 저는 정책의 결론이 분명해야 한다고 생각합니다. 보다 광범위한 기록이 존재해야 합니다. **PISA** 시험의 모순은 인지 능력, 즉 수학과 읽기 능력만을 측정하는 대신, 불완전하긴 해도 동기와 인성의 척도를 포착해 낸다는 점입니다. 그럴 바에야 아예 처음부터 비인지적 특성을 직접

적으로 측정하는 것이 낫지 않을까요? 성실성이나 과제를 집중해 수행할 수 있는 능력과 같은 예측력 있는 특성의 측정 기준을 검토해 보는 것이 더 낫지 않을까요? 그렇게 되면 우선은 현재의 측정 시스템이 본질적으로 보강될 것이고, 학교 차원에서는 호러스만과 비네, 테일러를 비롯해서, 공자도 그랬죠? 교육을 중요한 현상으로 생각한 거의 모든 교육자가 이야기해 온 완전한 인간을 육성하려는 책임의식이 더욱 더 잘 형성될 수 있습니다. 그렇기에 시험 개발자에서부터 공자에 이르기까지 다양한 분야와 시대의 많은 사람이 교육은 훨씬 더 포괄적인 활동이며, 교육의 결과가 완전하게 측정돼야 한다는 똑 같은 말을 한 것입니다.

많은 사람들은 한국의 인센티브 제도가 과도하게 초점이 맞춰진 주의를 제공한다고 말하며 이에 대해 반대할 것입니다. 이런 세계에서 살아남은 사람도 있을 것입니다. 지금 이 강연장의 객석에 앉아 있는 많은 사람이 그러하리라고 저는 확신합니다. 살아남은 사람들은 인성의 미덕과 끈기의 미덕 등을 통해서 여기에서 벗어날 수 있었습니다. 그럼에도 불구하고 왜곡 현상은 명백히 나타나고 있으며, 더 많은 불평등과 빈부 격차가 생기고 전통적 인 가족의 가치가 위협받는 미래의 한국 사회에서도 이는 분명히 나타날 것입니다.

저는 결손가정이 있는 사회에서 비인지적 특성을 향상시키는 방법, 학교의 올바른 역할, 가정의 올바른 역할, 비인지적 스킬이 실

제로 생성되는 과정에 대해 고민하고 있습니다. 저는 **PISA** 점수가 대단히 불안정한 회계 시스템이라고 생각합니다. 어떤 의미에는 시험 점수에 대한 집착이 사람들을 교육하고 효율적인 교육체제를 고안하려고 하는 사회의 노력을 잘못 인도하기도 한다는 주장에 전적으로 동의합니다. 그리고 이것이 중요한 메시지라고 생각합니다.

The 5th Asan Memorial Lecture
“What Tests Miss:
Hard Evidence on Soft Skills”

James J. Heckman

August 8, 2011

James J. Heckman

James J. Heckman is the Henry Schultz Distinguished Service Professor of Economics at the University of Chicago where he has served since 1973. In 2000, he won the Nobel Memorial Prize in Economics for his groundbreaking work in economics and microeconomics. He is considered one of the ten most influential economists in the world.



Dr. Heckman currently directs the Economics Research Center in the Department of Economics and the Center for Social Program Evaluation at the Harris School for Public Policy. He served as an assistant professor at Columbia University prior to joining the Economics Department at the University of Chicago. He has also taught at various schools including University College Dublin, Peking University, Yale University, and New York University. In addition, Dr. Heckman has published an extensive number of books, articles, and has received numerous awards for his work. Dr. Heckman received his B.A. in mathematics from Colorado College, an M.A., and Ph.D. in economics from Princeton University.

Chapter **1** Lecture

What Tests Miss: Hard Evidence on Soft Skills

Today's talk is about achievement tests, what these tests miss, and why they miss what is important to know. These tests are widely used not just in Korean society but in many societies, to screen and monitor the success of schools, and society itself. In the United States, there is the No Child Left Behind Act, there are achievement tests, Iowa Tests, and other tests of this sort. What these tests all miss are non-cognitive traits, or personality traits.

This may sound obvious, but in fact much of the discussion really does not consider this point. These traits are sometimes called “soft skills” or “character skills,” and examples of these traits include perseverance, conscientiousness, motivation, and willful planning. Recent research in economics and education shows that these traits predict a range of important behaviors and that ignoring these traits—and ignoring character in general—is a very dangerous practice. It can lead to costly mistakes in assessing and addressing social problems, and evaluating the success and failure of social policies.

I want to illustrate this by way of an example. There is a large-scale testing program in the United States known as the “GED.” The details of this test are not so important, but it can be taken by secondary school dropouts to certify that they are the equivalents, in terms of this test, to ordinary secondary school graduates. About one-seventh of all U.S. secondary school students who graduate use this program and it is also widely implemented in other countries. For example, Brazil uses it, it is used in Canada, and there are other countries that have versions of this test. Drawing on this particular test, but speaking generally as well, I want to try to draw some important lessons for education, educational policy, and educational assessment.

More than academic achievement is required for success in life. Soft skills are frequently called “soft skills” because it is thought that they cannot be measured, but that is simply not true. Soft skills can be measured, and in addition, personality can be fostered. A good school, a good education, and a good family can foster these skills. Programs that promote these skills are unexplored in many societies, and potentially powerful tools for economic and social policy are thus overlooked.

There are also other lessons that I want to draw out today. First, there are movements to increase accountability, to see whether or not schools and teachers are doing their jobs. These movements often create perverse incentives. In fact, tests and test certification—this kind of focus on test scores—can create and conceal problems. It warps educational goals, stifles creativity, and is not that successful in terms of predicting success in school, even though that is what it was designed to do.

There is a wide array of outcomes that these soft skills produce. They predict behavior, health behaviors, teenage pregnancy, secondary school graduation, wages, criminal activity, employment, and welfare dependency. In fact, some hard evidence on soft skills show that in many areas

soft skills play a greater role in determining outcomes than cognitive skills. So the question is, if these skills are so important, why have they been ignored in public policy discussions for so long?

Many people, even some people in the professional educational establishment with the title of psychometrician, dismiss these skills as being fuzzy and soft. They are viewed as concepts that have a very weak, tenuous predictive power, and they are thought to be difficult to measure. In fact, they can be measured and they are highly predictive.

It is known—and it is just obvious at an intuitive level—that many different skills are important for success in life. Achievement in different fields requires different bundles of talent at different levels. Here is an example: a very famous American inventor, Thomas Edison, invented many things in the 19th century, including the light bulb and the phonograph. He was hard-driving and persistent. Thomas Edison is celebrated both for his creativity in terms of the number of experiments he ran and also for the fact that he worked endlessly under his inspiration to achieve what he accomplished. He described himself in the following terms: “the genius is 1% inspiration and 99% perspiration.” A more recent person, less famous as an inventor than Thomas

Edison, is Woody Allen the American comedian. He put the point slightly differently, saying that 80% of success was just showing up, meaning engagement with society is extremely important.



Those who show up, who attempt, who try, play a far more important role. It is not just a question of being bright. In almost every task in life, more than just the raw ability to solve abstract problems is required for success. Although in different tasks, in different parts of society, the proportion that is required in cognition and character varies across tasks. Cognition is very important in very complex tasks. Things like self-control, self-discipline, and motivation are required to foster and apply talent. Despite the intuitive force that many different abilities are required to succeed, these soft skills are typically ignored.

Academic discussions ignore these skills and I know it is true here in Korea. Much of the discussion is in terms of intelligence, the IQ, cognition, maybe some achievement tests, and everything else is given a back seat. Now, schools are not expected to teach character or these personality traits as much as they are designed to teach cognition, although it was not always so. In an earlier era there was a time when a fuller notion of teaching and education, a Confucian notion, was much broader. But now, school effectiveness is not how well the person performs in society or even adapts in the larger society; it is much more a focus on how well that person performs on a test. In fact, in South Korea, the hagwon system primarily creates a focus on coaching

students to pass tests, not to form their character and these other basic skills that turn out to be so important.

Character education and other traditional values are thought to be the province of the family. Yet even in Korea, which has very strong families compared to many other countries in the world, the family is under stress. Some families are better situated to foster these traits than other families. It is known that single-parent families provide fewer resources for their children. Seong Hyeok Moon, one of my students, has done important work at the University of Chicago showing and documenting how important the differences can be between single-parent families and multiple-parent families. Counting single-parent families, divorced homes, out-of-wedlock families, and the like, roughly 18% of Korean children are now in single-parent families and that figure is increasing.

In Korea, there is evidence of inequality among children in environments that promote schooling and success. If effective social policy is to be designed, it has to be designed to look beyond the one-dimensional focus on cognition, and tests of cognition that dominates current thinking.

The history of testing goes back to ancient China, and in fact in Korea there was a reliance on tests in the civil services

centuries ago. But the modern movement in education, which influences the discussion in the world today, was started in the United States in around the 1840s by an educator named Horace Mann. He was the first to introduce the notion of a standardized test and he played a very important role in the history of education. He introduced the test, but he realized it was very much a crude measure of what schools were designed to do. Mann had a view that could be described at least partly as a Confucian view. He viewed the primary function of schools as teaching morality and character. In 1867, Mann said, “To value schools by length instead of quality is a matchless absurdity. Arithmetic, grammar, and other rudiments comprise but a small part of the teachings of school. The rudiments of feeling are taught not less than the rudiments of thinking. The sentiments and passions get more lessons than the intellect. Although their open recitations may be less, their secret rehearsals are more.” So this idea was always present. Every designer of a test—the IQ tests, the achievement tests, the tests that actually underlie the system—has recognized their limited value. Mann’s ideas were not really adopted in the 19th century because the proper technology was not yet available.

It was not until the 20th century that progressives like John Dewey, and others who sought to create a meritocratic

society, tried to institute the idea of system-wide testing to essentially open up schools to a broader population. In the early part of the 20th century, progressives like John Dewey turned to the IQ, although Dewey himself did not develop the IQ test. The idea was that the IQ and a test of cognition would play a role in essentially determining who is ready for school, and who is not. The original IQ test was not designed to predict success in college, so much as designed to screen out misfits, but the role of the test was rapidly broadened to sort students within schools.

It is interesting that one of the first IQ tests ever developed was created by a Frenchman. That test was further developed by a second Frenchman named Binet, who teamed up with a psychologist at Stanford named Lewis Terman. And together they created the Stanford-Binet test, which is still a standard IQ test. It is interesting to read that almost a 100 years ago when this test was developed, Binet pointed out that success in school “admits of other things than intelligence” and that to succeed in school one must have other qualities that depend on attention, will, character, and for example, a certain docility, a regularity of habits, and a continuity of effort. Even a 100 years ago there was a stress and many people working on, what is called “g.” “G” is a unitary factor designed to measure intelligence, to recognize that conscientiousness, a



Soft skills matter.
These skills, both cognitive and non-cognitive,
are not genetically determined.
They can be shaped, even in the adolescent
and adult years.
Improving them is a productive
avenue of social policy

non-cognitive trait, plays a very powerful role in explaining success in schools. More recently, a person who is very much associated with the idea of the power of intelligence, Arthur Jensen, writes in a recent book on the “g” factor that the traits of conscientiousness, being responsible, dependable, caring, organized, and persistent are important personality traits leading to success.

But all these notions are easily ignored because in the initial wave of creating tests, there was a notion of trying to be scientific and objective. In the late 19th century and the early part of the 20th century, America started developing an obsession for efficiency and productivity—partly because of the work of a man named Frederick Taylor—which I think South Korea shares. Taylorism was highly influential. The idea was to create scientific management, scientific measurement, accountability, and in some sense fairness. But there was a real question about how the principle of scientific management could be used in schools.

It turned out that there was an educational philosopher and theorist, a professor at the University of Chicago, named Bobbitt. He was very influential in motivating and teaching the people who would later develop the tests that are still widely used—the PISA tests, the SAT, and the achievement

tests. It is interesting, as one reads the works of Horace Mann, Arthur Jensen, and Binet, to ask what education is about. Bobbitt viewed education like a factory. But what was the factory actually producing? He saw education as a process much like the manufacture of steel rails, but instead of producing steel rails, it produced character and personality. It is a shaping of more delicate matters but a shaping nonetheless. Bobbitt recognized that he lacked the tools to measure what schools were really doing, but like Mann, he viewed personality as an important trait.

It is interesting to see what his students did. After the development of the IQ test in the early part of the 20th century, a new concept emerged. This is a more recent concept, and it is a concept that underlies the PISA tests and a lot of the achievement tests. It was a concept developed at the University of Chicago, at least partly developed there, called “general knowledge.” The idea was to capture not just the raw intelligence but that individuals were learning, not in any specific course, but in a general way. This was a new concept, an idea no one had 50 years earlier. The idea was not just to measure IQ or some measure of how quickly a person could solve a problem, but important life skills. The Iowa Tests, the GED, No Child Left Behind, and PISA are all versions of this test that are designed to capture important life

skills, not specific knowledge.

These tests were easy to grade, but it was only perfected in the 1950s, when technology emerged that allowed a machine to grade the tests using optical scanning. But here is what Tyler, who is probably the single most important person in modern testing in the sense of having created the form of the modern test, wrote about the value of written exams. Even though he is the godfather of these tests, he says, “We rely heavily on written exams, on a few types of objective tests, and occasionally on the subjective impressions of teachers. Many other appraisal devices could be used, such as records of activities, questionnaires, anecdotal records, and observational records.” And he encouraged that this be done. In fact, late in his life when these tests were prototypes and the PISA tests were being developed, he advocated much richer versions of these tests than are currently implemented.

Here is an example of this by way of the GED, which was created in the wake of this movement in the 1940s and 1950s. The GED is an achievement test that was created to certify the general knowledge of what soldiers had learned in World War II. In the United States during the Second World War, many young males were drafted out of high school and they were not able to finish even their secondary

education because of the requirements of war. But there was an assumption that these soldiers who had been in the army—they took courses in the army—for two or three years possessed similar skills as those who had finished secondary school. Army discipline had selected them, they had survived the rigors of the army, were obedient, self-controlled, and able to work with others. So this was implicit at the time this test was created.

It is interesting that this test, which was originally designed to certify soldiers as being equivalent to ordinary high school graduates, became applied much more generally in American society. It was applied to people who were not subject to the rigors of military discipline, who were civilians who dropped out of secondary school. This test had a widespread influence in American education, but it was a part of a larger movement that I am sure is present here in South Korea as well. It is a desire for egalitarianism, for meritocracy. The SAT and its widespread application in the United States was designed to essentially break up old-boy networks, to allow bright kids who came from poor families to rise to the top and create a more meritocratic society. On top of that, there were powerful forces in American society at the time, in the 1950s and 1960s, to revive Taylorism and to show that governments were being responsible. If one had a

test score, one could see how successful the government had been.

An extreme version of that was evident in American policy during the war in Vietnam. For example Robert McNamara, the defense secretary, actually revived Taylorism in the Defense Department and created economic principles to try to produce a social version of a profit-and-loss statement. It was an extreme version that dealt with body counts in Vietnam, where people would talk about the kill ratio—the number of Viet Cong killed relative to American troops killed.

But more generally there was a notion of using tests and using measurements of some sort to make society accountable. The Great Society, launched in the 1960s, did this. Achievement tests and the IQ tests became a very important part of this accountability movement—the achievement test in particular—which measured how much value was added by a school system. This became a mania in the United States during the [George W.] Bush administration with the No Child Left Behind Act. This essentially incentivized schools to do very well on certain tests that tested reading and basic mathematics. It became an entire culture where teachers were driven to teach to the tests and students were motivated only to understand what was on the tests.

So during the period of 1955 to the early part of the 21st century, there was a steady growth in the sales of standardized tests in real terms. At one time, close to 20% of students receiving a high school degree received it through the GED. About 700,000 to 800,000 students a year drop out and take a test instead of actually doing the hard work of staying through four years of high school.

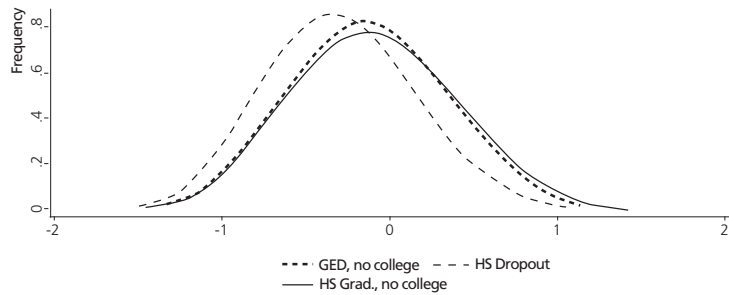
Here is an example of what the test shows. A simple algebraic equation asks: if $8x+16$ equals 32 what is x ? The answer is 2. That is at the eighth-grade level. A more difficult question asks, if you can plant 15 trees a day, how many days will it take to plant 200 trees? That is about the level of the GED.

So who are these GEDs? These are people who are between high school graduates and dropouts, but they turn out to be as smart as ordinary high school graduates who do not go on to college. Figure 1, a graph of cognitive skills for females by educational status and Figure 2 is for males. As is illustrated by these standard bell curves, there is almost no difference in cognitive ability between someone who receives a GED and someone who graduates from high school.

There is no question that the people who pass this test

Figure 1

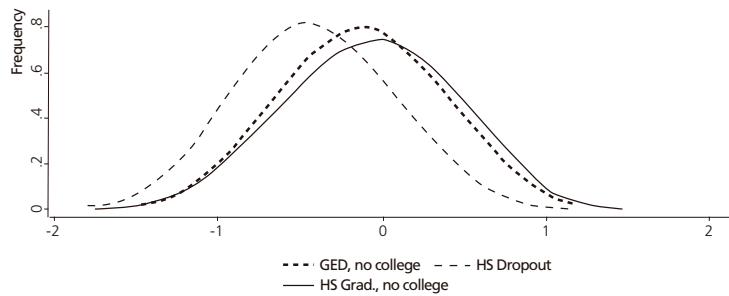
Cognitive Ability by Educational Status (Females)



Source: Heckman, Humphries, Urzua, and Veramendi (2010)

Figure 2

Cognitive Ability by Educational Status (Males)



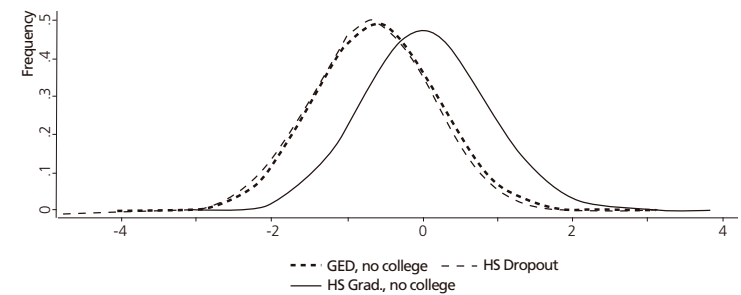
Source: Heckman, Humphries, Urzua, and Veramendi (2010)

are as smart as ordinary high school graduates. But it turns out, even though they are as smart as ordinary high school graduates, these people lack soft skills. Aggregates of these

soft skills demonstrate that the GEDs have the soft skills of dropouts and are far different from those who complete high school. The same thing is true for females (Figure 3) and males (Figure 4).

Figure 3

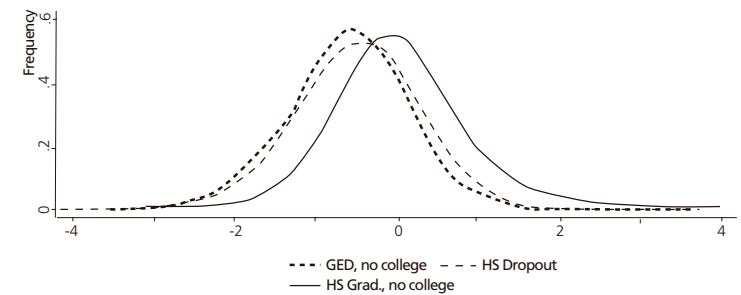
Non-cognitive Ability by Education Status (Females)



Source: Heckman, Humphries, Urzua, and Veramendi (2010)

Figure 4

Non-cognitive Ability by Education Status (Males)



Source: Heckman, Humphries, Urzua, and Veramendi (2010)

What is remarkable and what has to be understood is that when the real performance is considered, not just on a test but how people actually work in the labor market, the wages earned by GED recipients are at the level of high school dropouts. So if the results are not adjusted for cognitive ability, they are just as smart but they earn no more and they perform less well in society than ordinary high school graduates.

The GED is actually an interesting social experiment where individuals who are smart are certified, but they are not able to stay on task and finish tasks. So the question is whether the GED plays a role. It signals ability and this signal may be of value in the labor market. This possibility could be examined by looking at the wages the GEDs earned before and after they received the GED. However, there is no evidence of any signaling value. So the test itself is quite dubious.

But let us focus not just on the GED. What exactly are these tests—which so many people rely on—capturing? How are they validated? Those in industry typically want a test or whatever they are doing to work. An automaker makes a car and asks that it perform well on the road. A shipbuilder builds a ship and asks that it sail. Observers could ask the

same thing of tests. How are these tests validated? One can look at predictive validities. This has been determined. What exactly do these tests predict? How strong is the predictive power of IQ, grades, and achievement tests? What is interesting is that even on their own terms, designed to predict who will succeed in school and who will not, these tests are not that effective.

It is known from a general body of knowledge that for many complex tasks, cognition and the kinds of skills captured by these cognitive tests have real predictive power. Complex tasks require mental ability, no question about it. But for all tasks complex and not, personality has greater predictive power in the sense that, across a wider range of tasks soft skills are highly predictive in blue-collar jobs. But even in complex tasks, soft skills are highly predictive.

Tests are generally validated by their correlation with other tests, not in real-world behavior. Consider the validities of these tests. A very common measure of validity is what fraction of the variability in outcome is explained by these tests. For example, as shown in Table 1, consider the IQ test, a grade point average, and how well the teacher assessment is in terms of a person's scores on written tests. This is something that is like an achievement test. The achievement

tests are actually somewhat more predictive of wages at 35. The total level of prediction is quite low. 1 would be a perfect prediction, and .05 would only explain about 5% of the variability. If welfare participation or hours worked are considered, these tests are a little more predictive, but they still do not predict very much. There is an interesting pattern here, which is that this achievement test is actually more predictive than a straight test of IQ. This is interesting.

Table 1

Validities in Labor Market Outcomes from the National Longitudinal Survey of Youth, 1979

Outcome	Males			Female		
	IQ	GPA (10th grade)	AFQT	IQ	GPA (10th grade)	AFQT
Hourly Wage Age 35	0.03	0.05***	0.05***	0.11***	0.10***	0.13***
Hours Worked Age 35	0.10***	0.12***	0.21***	0.02	0.10***	0.17***
Any welfare Age 35	-0.09***	-0.11***	-0.23***	-0.20***	-0.23***	-0.36***

Note : *** $P < .001$

What does this mean? What does this arise from? These achievement tests, even though they do it imperfectly, partially capture measures of personality. Are measures of personality any better or worse in terms of their performance in real-life behavior? This is a real concern.

Recently, in personality psychology something called the “Big Five” has been developed, which is a group of measures that allow you to determine what the dimensions are in terms of people’s performance. There has been a lot of work about predictive validity. An acronym has been accepted called “OCEAN”, for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These are different traits that have been studied relative to the power of IQ in cognition and predicting success in college. These are studies done by psychologists, educators, and economists. Table 2 shows three different studies. But the key question to ask is, “What is the relative power of just one trait of personality, conscientiousness?” Again, consider something like the SAT test. Even the Educational Testing Service, the central power in educational testing, has to admit that conscientiousness alone is more predictive of success, scores, and outcomes in college than are SAT scores—just one measure of personality. This is an embarrassment for the educational testing community.

Not only can conscientiousness be measured, but when it is measured, it turns out to be much more predictive. For example, Figure 5 is from a German study. If one looks at these Big Five measures and puts on that same scale fluid intelligence, which is a measure of how rapidly a person solves problems, and crystallized intelligence, which is a

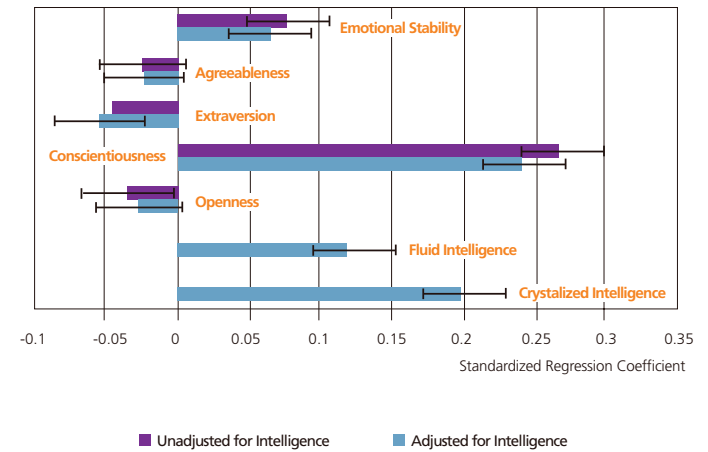
Table 2
The Relative Predictive Power of Conscientiousness and SAT Scores for College GPA

Source	Sample	Timing of Measurement and Outcome	Controls	Metric	Results
Conard [2005]	University students in the US (N=186)	College GPA and SAT were both self reported during college. Personality was measured in college.	Class Attendance	Standardized Regression Coefficient (β)	SAT Total Conscientiousness 0.27 0.30
Noftle and Robins [2007]	University Students in the US (N=10,497)	College GPA and SAT were both self reported during college. Personality was measured in college.	Gender, Other Big Five Traits	Standardized Regression Coefficient (β)	SAT Verbal 0.19 SAT Math-Conscientiousness 0.16 0.24
Wolfe and Johnson [1995]	University students in the US (N=201)	GPA and SAT were provided by the Colleges' Record Office. Personality was measured in college.	High School GPA	Standardized Regression Coefficient (β)	SAT Total Conscientiousness 0.23 0.31

Notes: (a) Self-reported SAT scores and those obtained from college records were highly correlated ($r=0.92$).
 (b) Self-reported GPA and that obtained from college records were highly correlated ($r=0.89$).

measure of a person's knowledge—more like an achievement test—and asks how many years of school a person actually succeeds in, one can see that conscientiousness is a much better predictor. Nothing is a perfect predictor, but it is a

Figure 5
Association of the Big Five and Intelligence with Years of Schooling (males)



much better predictor than the IQ or achievement tests. If course grades are considered (Figure 6), it is clear that conscientiousness is again more of a predictor among all of the personality traits. If things like associations with standardized achievement test scores are considered (Figure 7), it is clear that these personality traits play a predictive role as well. So what is going on here? Look at something like mortality (Figure 8), something that normally is not thought of as a role for educational policy but in fact should be. It is clear that conscientiousness is a greater predictor. Nothing

is a good predictor of mortality, but conscientiousness, self-control, or self-management is a better predictor than the IQ. One can go across a number of social outcomes. As one goes from the bottom of the distribution to the top and compares the power of cognitive skills with the power of non-cognitive skills, one goes from the bottom to the top of the distribution, and it is clear that those people who are very low in these social soft skills are also much more likely to have been jailed (Figure 9).

Figure 6

Correlations with the Big Five and Intelligence with Course Grades

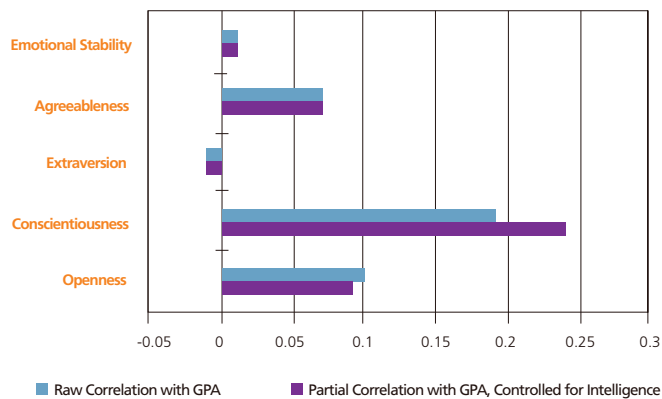


Figure 7

Associations with Standardized Achievement Test Scores

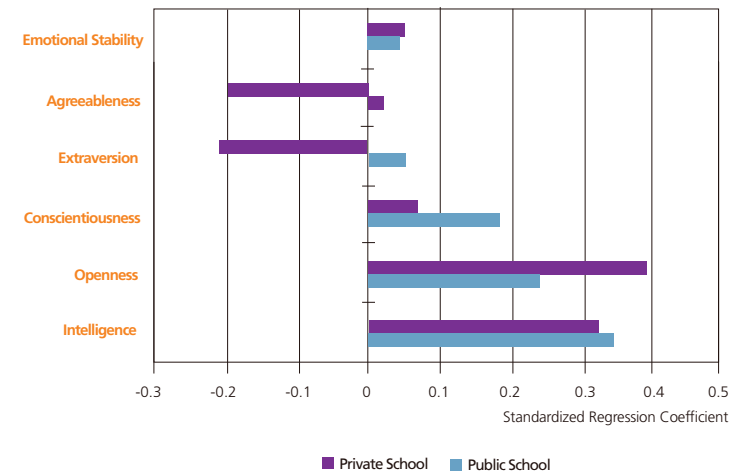


Figure 8

Correlations of Mortality with Personality, IQ, and Socioeconomic Status (SES)

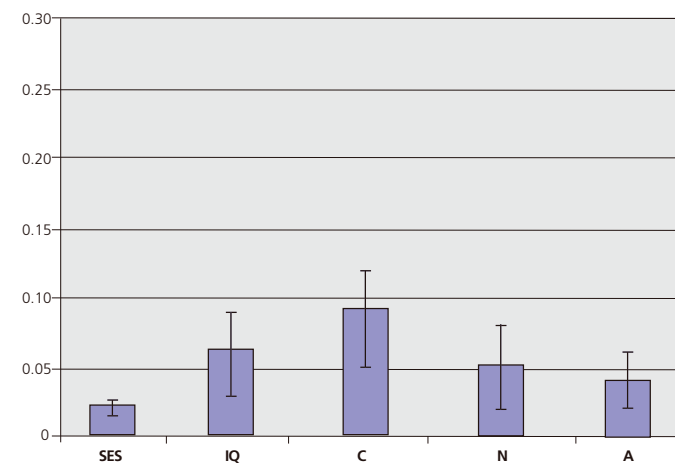
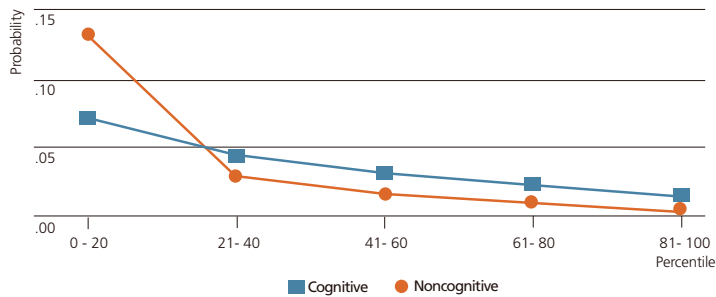


Figure 9

Ever Been in Jail by Age 30, by Ability (males)



Source: Heckman, Stixrud, and Urzua (2006)

So what is going here? IQ is not the same as achievement. Achievement tests are explained in part by personality traits and this is something that people have not fully developed. The achievement test inadvertently captures some aspect of personality. For example, some recent work has actually looked at how much of the variability in the achievement test score can be explained by the IQ and by aspects of personality. This is both good news and bad news for the achievement test score. A substantial part of the variability in the achievement test score is explained by just two personality factors. If they are all put in, the explanation is much greater. So the question then becomes, what is being used? Achievement test scores that are used uncritically? What does the score on the achievement test really measure? For

that matter, what does a score on the IQ test measure?

Here is one notorious example. It is well known that between the African-American population and the white population in the United States, there is a substantial difference on average in the IQ tests. It is about 16 points on a scale of 100. In an experiment done some 40 years ago, disadvantaged African-American children were essentially given candy for each correct answer on the IQ test. It turned out that those kids who were given the candy, given some incentive, actually scored 12 points higher that day on the test. This means that the black-white gap was essentially eliminated in terms of achievement tests, or at least in terms of the IQ tests.

So researchers have come to understand that even performance on the test is partly motivated by aspects of personality. In fact, it turns out in subsequent research that when researchers look at the most conscientious children, their response to these incentives was quite weak. Why? Because, these students were already trying hard. But the individuals who were not as highly motivated did not do that well on the test. So the tests are measuring in part motivation and in part raw intelligence. They are measuring in part knowledge and in part desire. But knowledge itself in part measures motivation. Incentives operate much more effectively on those with lower levels of motivation, and in fact, what researchers found was that even

though the score on a test can be raised, there is no lasting effect of incentives in terms of increasing paid performance.

So what are the costs of the achievement tests? There is a cost for American society and a cost for Korean society. Here is some evidence of what the costs are in American society. Coming back to this test, it certifies that people are equal in cognitive traits but leads to a substantial lack of performance in larger society even though the students get a high score. This turns out to deceive American public policy.

The one-dimensional focus of public policy on smarts conceals major problems by misdirecting the efforts of institutions and individuals. The GEDs are earning at the rate of dropouts. It also turns out, in the United States, that if these people who pass tests are counted as being high school equivalents, the fraction of the population that is actually achieving high school graduation status is radically overstated. In fact, when the high school graduation rate is adjusted, it is clear that it has actually been decreasing over the last 30 years. This helps to explain certain problems in American society.

Well, maybe the GED is not bad. What is it that the GED is doing? It is an achievement test and some people may benefit by passing it. It turns out that the people who actually

benefit, who actually use it, are those who made some kind of mistake in their lives. They took the test and they are high on both cognitive and non-cognitive skills. It does create an option. But even though most of these kids are bright enough to go to college, 40% try college, and only 3% actually graduate. So there is a very low level of benefit. There are other criteria as well. This test is offered to the vulnerable youth, individuals who essentially are in their late teenage years, and made bad choices. If they are given the option not to go to school, but instead to take a test, they actually make choices that can be unwise.

For example, some recent works in psychology demonstrate that the intellectual ability of a child develops more rapidly. Intellectual skills are formed at an earlier age and the psychosocial maturity—personality skills—are formed at a somewhat later age. So offering the test to kids in this period entices children to drop out of school and take the test, and in fact leads to permanent and serious lifetime problems.

What I have tried to suggest is that other skills besides cognitive skills are required. But the question then becomes, what about these skills? What are these cognitive and non-cognitive skills, and how are they actually fostered? First of all, one thing that researchers have come to learn

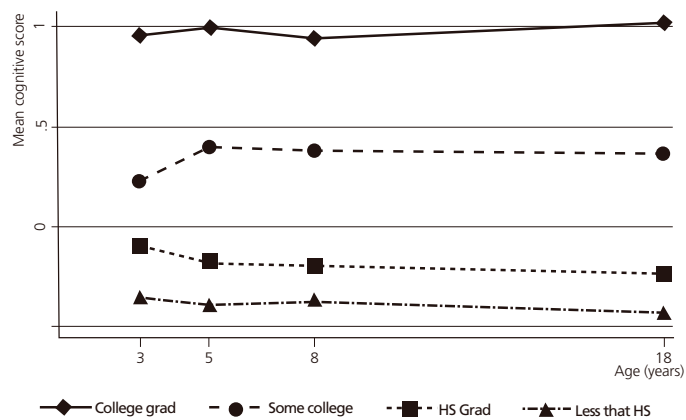
from studies in economics and psychology in America is that children with a mother who is less than a secondary school graduate have low scores on cognitive tests (Figure 10). The absolute scale is not so important, although it is highly predictive of the cognitive scores. But there is a huge difference between people from advantaged families at the top and disadvantaged families at the bottom. There is a huge gap at 18, but the most important feature of this table is that the gap is present at age three. So these cognitive scores, these test score gaps, these gaps between the advantaged and disadvantaged, turn out to be very important. If one is

interested in meritocracy and egalitarianism, these gaps that are present at 18 are emerging at age three.

Figure 11 looks at the same kind of criteria in terms of soft skills—and here it is done in reverse order, a high score means high behavioral problems. The children from the most disadvantaged families have the greatest behavioral problems and those from the most advantaged families have the least behavioral problems. But the gaps that are present at age twelve are more or less present at age four. So the question is open then—maybe these skills are just genetic. If they are present already at age three or age four, maybe these skills

Figure 10

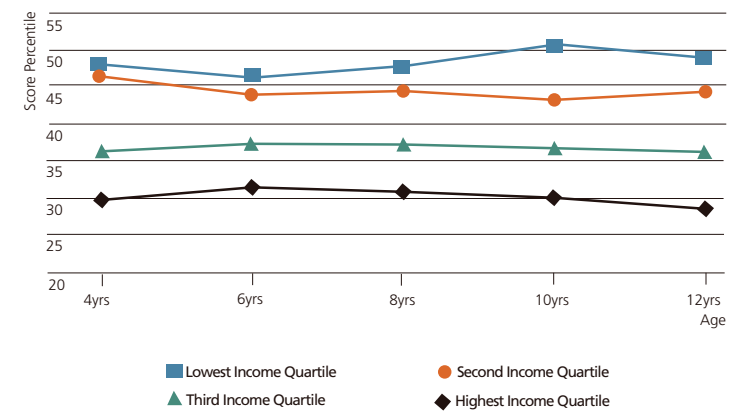
Trend in Mean Cognitive Score by Maternal Education



Source: Brooks-Gunn et al. (2006)

Figure 11

Average Percentile Rank on Anti-Social Behavior Score, by Income Quartile



are just the result of who your parents are and what their genes are. But what researchers have come to understand is that these cognitive and non-cognitive traits are not solely determined by genetics. About 50% of the variability in these traits turns out to be genetically based, but family investment in early childhood programs creates both cognitive and non-cognitive skills.

Researchers have come to learn that IQ can be fostered in the very early years even though it becomes roughly stable by the teenage years. Crystallized knowledge can be acquired throughout one's lifetime and that is what is measured on an achievement test. Researchers have also come to understand that personality skills are more malleable at later ages. Schools and family environments provide important ways to shape these skills.

Figure 12 presents evidence from the United States. The question is, how much does an extra year of schooling promote one's ability to make arithmetic reasoning? Not surprisingly, the more education one has, the higher the test score in terms of arithmetic reasoning. The same is true for word knowledge (Figure 13) and for other traits. But an examination of socio-emotional skills, shown in Figure 14, reveals the dramatic effects of education. So secondary

Figure 12

Causal Effect of Schooling on Measures of Cognition (from ASVAB)

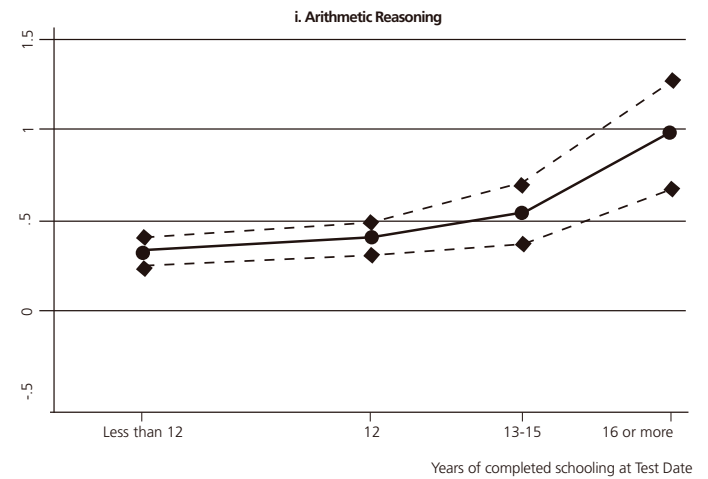
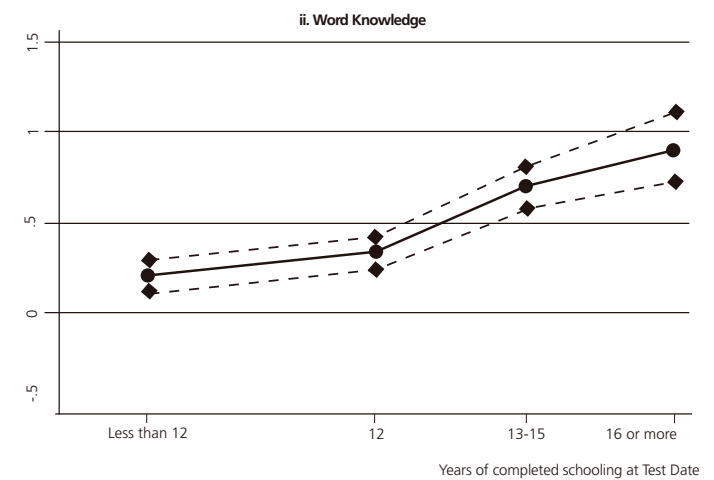


Figure 13

Causal Effect of Schooling on Measures of Cognition (from ASVAB)



education, even an extra year of high school which the GEDs are missing, has a real impact on producing self-esteem, self-confidence, and soft skills. What researchers have come to understand is that there is a dynamics of skill formation. Enriched early environments foster socio-emotional skills and they are important avenues for promoting equality and productivity in society.

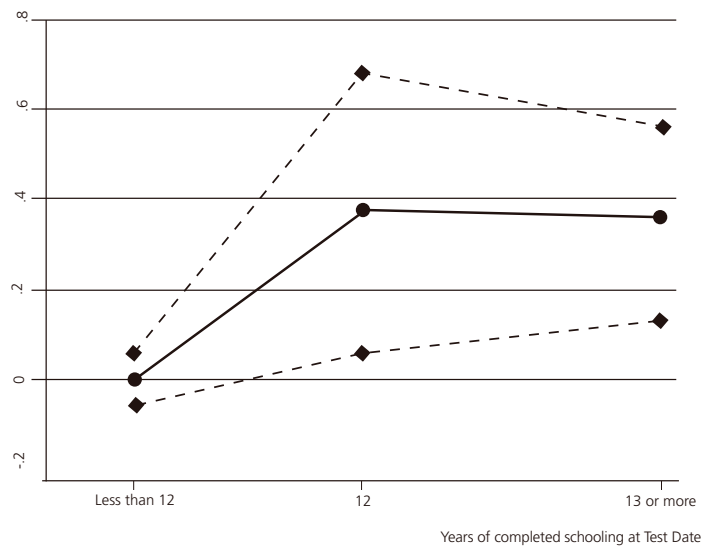
Those gaps that emerged can actually be filled if the process starts early enough in the lifecycle. I, along with Seong Moon and others at the University of Chicago, analyzed a program

developed some 50 years ago by a group near the University of Michigan. Called the “Perry Program,” it aimed to enrich the early lives of low-income African-American children; all of them had IQs below 85. The program took these disadvantaged children and gave them two and a half hours of instruction, five days a week, for two years. But what was the nature of instruction? This is why one must think more broadly about what education does. The kids were three and four years of age, so they had not yet started to attend school. However, the program was terminated. What did the program do? It taught planning and persistence. The criterion, if taken literally, was “plan, do, and review”. It taught personality skills, it taught perseverance, anger management, and the ability to stay on task. What is remarkable about this program is that two groups were created, a treatment group and a control group. The treatment group was actually given the program and the control group, identical otherwise just by random assignment, was not given the program. These children could be followed over their lifetimes.

What is interesting about the program is that the economic return to this program is 7 to 10% per annum. That is a very high economic return. But what were the criteria? Did the program raise IQs of the children? No. Remember, the IQ was selected to be low, so in the early years the treatments

Figure 14

Causal Effect of Schooling on Two Measures of Socioemotional Skills

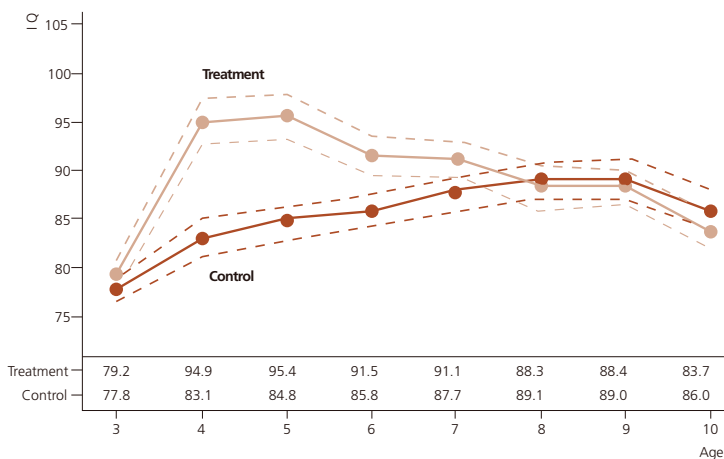


and controls were more or less the same. When the children were in the program the treatment group IQ, the upper graph, was much bigger as shown in Figure 15. So there was a real gain in the IQ. But by age eight, nine, and ten there was no difference between the treatment and control groups. IQ was not boosted. Yet, the rate of return was 7 to 10% per annum, which was above the U.S. return to equity until 2008, post-World War II.

The channel by which it worked was through socio-emotional skills and these personality factors that affect

Figure 15

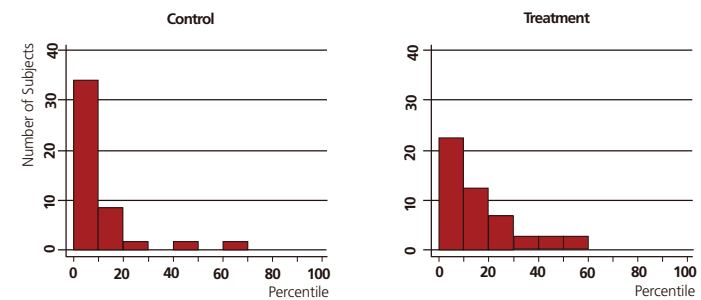
Cognitive Evolution Through Time,
Perry Males: Male Cognitive Dynamics



performance on the achievement tests. What was the source of growth of the achievement test? Well, the researchers looked at something called the California Achievement Test (CAT), which is not unlike the PISA tests. How could they promote that? The control group's performance was lower and the treatment group's performance was higher, as shown in Figure 16. Remember, they had no higher IQ so the treatment group actually had the same IQ as the control group on average, but the soft skills were changed. So an examination of the measures of personal behavior between the treatment group and the control group reveals real benefits. The socio-emotional index shows real benefits between the treatment and control groups. Most of the

Figure 16

Perry Age 14 Total CAT Scores



CAT = California Achievement Test
 Treatment: N = 49; Control: N = 46
 Statistically Significant Effect for Males and Females (p-values 0.009, 0.021 respectively)
 Source: Heckman, Malofeeva, Pinto, and Savelyev (2008)

sources of the treatment effects in this program were not changing the cognitive factors but instead the changes were in personal behavior, in socio-emotional state, that actually created the effects that were statistically significant, that reduced crime, and promoted long-term success.

What researchers have come to understand is that these traits are not fixed in stone. Researchers have also come to understand that investment is most productive in the early years for the IQ and fluid intelligence. Investment is relatively more productive in the middle years for fostering personality and this is associated with certain developmental aspects of childhood.

So what can be learned from this? There is a much richer knowledge these days about what education is, what educational policy should be, and how education can be evaluated. Educational attainment really depends much more on cognitive skills than non-cognitive skills, but non-cognitive skills play a role. Personality helps to foster cognitive skills. Why?

When a young child is motivated to learn, that young child will acquire more knowledge, and this was shown in the Perry study. Here are these kids, no brighter, but even though

they are no brighter they are more motivated. So when they were tested at the age 14 on the achievement test, they showed greater embodied knowledge as captured by the achievement test. But also in terms of other dimensions, in terms of their social performance on crime, in terms of divorce, in terms of a number of other traits, they are significantly improved in terms of their performance. This really leads to a focus where a wealth of skills is recognized. Skill is not just a single word, human capital is not just IQ, and education is not just about cognition.

The current exclusive focus on cognitive tests in Korean society ignores important dimensions of social performance. Soft skills matter. These skills, both cognitive and non-cognitive, are not genetically determined. They can be shaped, even in the adolescent and adult years, and improving them is a productive avenue of social policy.

The GED program I drew on illustrates the importance of soft skills. It is a test that genuinely satisfies the criteria of the psychometric educational establishment. These people are just as smart as ordinary high school graduates and yet they perform at the level of secondary school dropouts. It also leads to substantial distortion. I have mentioned the distorted basic data on the American economy, how the high

school graduation rate is overestimated. It also leads to huge distortions that are present here in Korean society as well.

As a result of the No Child Left Behind program, U.S. schools are focusing only on mathematics and reading rather than teaching physics, or other various traits. Policies focusing on promoting test scores miss a basic point about what is required to succeed. I think this is a very important conclusion, to understand that social policy should think more comprehensively about not only what its goals should be—educational policy, cognition, and personality—but to think of the whole person and to recognize that the whole person can be measured. There is a rich body of tools now that allow society to move beyond the one-dimensional focus on educational policy of looking only at a test score, a PISA score, an IQ score, and the like. A policy that essentially focuses on these measures, that essentially takes a wider account of what tests miss, will be a much more successful policy not only for American society but also for Korean society.



Chapter 2 Q & A

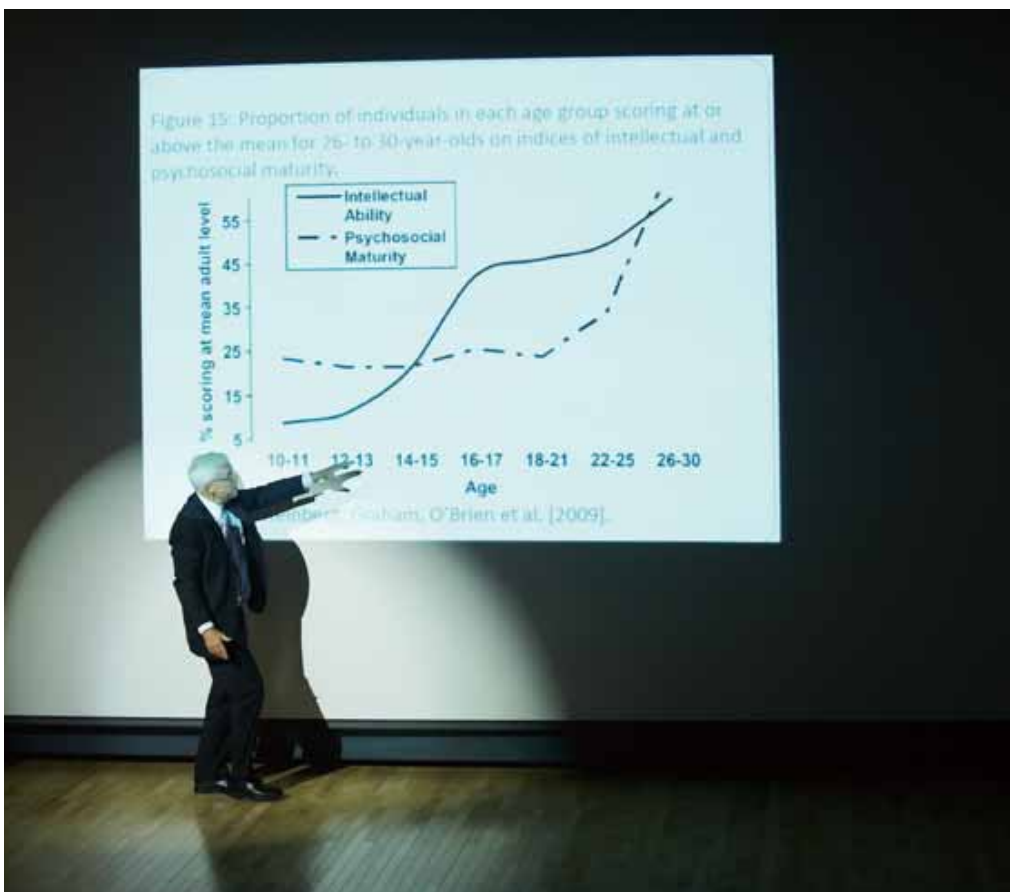
Question 1_ We Koreans are fully aware of the dangers of overreliance on standardized tests. But how do we get out of that trap? All the means that are known, that supposedly measure a child's skill or performance, allow the injection of subjective criteria. So are there any other measures or means by which these skills can be measured in an objective manner, other than the standardized tests?

James Heckman_ Well, sure. As I mentioned earlier, the same motivation that appears in Korean society was present

in American society 100 years ago. People like John Dewey and others were trying to promote the use of these tests to screen children, to allow people who were disadvantaged to gain access to education, and to broaden the educational base just as it has been broadened in Korea. But one has to recognize a couple of points here. I think what is really wanted is to predict who succeeds and who fails.

For example, consider the crude measure of the so-called subjective measures of the grades. Many people say grades are very subjective, but what are these tests? What do these tests actually measure? In the psychometric literature there is something called the validity of the test. I provided some examples of what the validities were in real-world data, like wages, hours worked, and so forth. I showed you that, first of all, the predictive power is still low where the correlations reflect the value of 0.2. These are not very powerful correlations. They explain only a fraction of the variance.

But secondly, you can do even better with these subjective measures. They are not working in their own terms and there is no sense of accountability of the test makers. There is a belief that somehow, the formalized tests such as Princeton Educational Testing Service, American Council on Testing, and the American Council on Education are psychometrically



and objectively formulated in which it is highly reasonable to distribute the tests to people. And yet, as I tried to show, every creator of an IQ test, every creator of an achievement test, and even the first person who advocated standardized tests have all pointed out that there is richer data to draw on. Some of the data are subjective opinions, but think about subjective opinions cumulated over the whole life of a child. A teacher may be biased against a student in one class, maybe in one course, in one year, but if these are cumulated, and there are good ways to do that, a profile on the performance of the child can be created. These measures of conscientiousness that I gave you and those that were used in the Perry study were actually measures of conscientiousness. They were created by assessments of school teachers literally asking how hard the child worked, how well they worked, how they cooperated with others, and so forth. These covered multiple schoolteachers over multiple years. That helps to eliminate a lot of the subjectivity.

So I think there are two aspects. First, surely a measure that predicts well is desirable. So, is it meritocratic to base people's potential success on a test that only explains 3, 4, or 5% of the variability in wages, hours, and lifetime outcomes? What about the other 95%? A big chunk of that, not the whole amount, can be explained by looking at those other

traits. Secondly, there has been a revolution in psychology. A lot of recent work shows that these personality traits are relatively stable objects, they are very predictive of what goes on in life, and they can be measured. How? Not only by self reports, which really are somewhat biased, but by teacher assessments and other third-party assessments of the child's behavior. So the means exist. It certainly was not true in 1940 that this could be done. With computerized information and databases, inventories of personality can be created that I think would be as successful. There are also records of the child's behavior.

So the point is, I am not advocating a new test to take the place of the old test. I am suggesting that the multiplicity of human skills, the variety of human skills, be recognized and that the capacity to create a much richer database of all those children exists. I think that is true meritocracy.

Question 2_ You emphasized the importance of measuring soft skills. But as far as I understand, you did not explain how to measure the soft skills. Could you explain this briefly?

James Heckman_ Yes. A large inventory has been created. I wouldn't call it a revolution, but personality psychology has undergone substantial development in the last 20 or 30 years.

There are various ways of charting the inventory of students. So, for example, there are inventories that can be taken about how focused the child can be. There is something called “executive functioning”, which essentially involves how well a student can stay on task, so there is less distractibility. There is also the Stroop test and other tests of that sort.

One can actually measure by giving tasks to children of various ages. If I give some measurements here and some data there, how well can the child organize it, how well can the child focus on the task at hand? That is one aspect of executive functioning. But one can also look at inventories, for example, of what children are doing in terms of how they are cooperating with other people. There are teachers’ reports and assessments, which are records. Many schools will keep those kinds of records, certainly in the United States. And so those can be measured and they turn out to be highly predictive of success in schools, and so forth.

It’s a combination of things like objective tests of a child’s performance as well as assessments by other people of the child’s behavior. I can direct you to some books, showing these tests, or some papers that I posted for the presentation today and show you what those references are. So these are not tests like the GED that ask the test taker to solve a

certain equation for “x”. It’s going to ask instead, how well did this child get along with other children? And again, even though there may be subjectivity with any one teacher, if you average over teachers and you give children some fair chance to sort of choose among teachers, or at least follow the child through the school system, you can come up with these inventories that are personal. It’s a combination of the behavior of the student as well as the performance on particular tasks that are inventoried.

I will give you a more specific example. There is something called the “Myers-Briggs Test” that is used a lot in U.S. manufacturing. In fact, there are more sales of the Myers-Briggs Test each year in the United States than of the SAT test. It’s an inventory where people are asked questions about personality, their attitudes, and so forth. There is another test called the “Hogan Personality Inventory Test”, which many firms in the United States use. People are asked objective questions or questions about their attitudes and so forth. How well do these tests predict future success? Well, businesses buy these tests. Millions of these tests are taken each year, and they are used a lot in hiring decisions, in promotion decisions, and in decisions for recruiting and screening. So in terms of passing a business test, they actually do succeed. They actually are predictive in the sense of firms

and individuals. So there is a whole science that has come into play to assess personality, this Big Five.

It used to be the case in personality psychology that each psychologist would have his own notion of what the relevant traits were. It was very difficult to compare the findings of one psychologist with the findings of another. But about 30 years ago, something called the Big Five—I gave you the acronym for it, OCEAN—essentially was solidified. A lot of research has been conducted about the predictive power of the Big Five, how conscientiousness promotes health, and how neuroticism in some professions can actually be very productive and in other professions, very unproductive. A neurotic surgeon is a good thing. He is very compulsive and he wants to succeed, so he is very anxious and task-oriented. But in some other fields neuroticism may be a very bad thing.

So there is a series not only of tests that have been given, screening devices, reports that are used, but also of successes and how well they predict. So I would say that there are measurements and they do predict. I showed you how in many cases they are better predictors. So there is really a scope for thinking more broadly about these traits.

Question 3_ What are some of the policy-level things that you are trying to achieve through this research? Are there one or two or three things at either the federal level or at the state level that you are trying to suggest to the policy makers, incorporating your findings? And what are your top-priority recommendations to Korea's policymakers?

James Heckman_ I think my objective in doing all of this is actually to try to tackle questions of inequality and disadvantage. The question is: how can one promote the lives of disadvantaged children? I think right now in many parts of American social policy and worldwide social policy, there still remains a very powerful belief in the importance of genetics and in the importance of traits like IQ in predicting success in life. There was a book written some 15 or 17 years ago by Charles Murray and Richard Herrnstein called "The Bell Curve". It basically tried to explain black-white inequality as being a matter of just IQ and essentially said that the IQ is genetically determined.

What this research program tries to understand is that these traits that are required for success in life are more than just cognition. They are more than just IQ. Actually, part of what I was showing to you was that the very achievement test that Herrnstein and Murray used was as much predictive by

personality traits as it was by cognition. I didn't emphasize this, but it was illustrated in one of the graphs I used. Personality played a very big role. So even their measure of IQ was actually also describing some aspect of personality. So that is one goal—that we researchers really understand what these tests are telling us and what the shortfalls of these tests are. A second goal of the research is to say that cognition and personality both can be measured, that the researchers can actually come up with the ways to accurately measure them.

The third goal is to show that these traits are not genetically given; they can be shaped. Disadvantaged children and the people growing up with fewer family resources will show from a very early age gaps in these cognitive and non-cognitive traits, but they are malleable; something can be done about it. Educational policy makers should think more broadly about more traits and understand the dynamics of skill formation, as well as understand that skill builds on skill, that early traits promote later traits. Just like in the Perry program, these kids were no smarter but they were more motivated to learn. So by the time they were 14 years of age and were tested even on their factual knowledge, they knew more. That essentially changes the way one thinks about social policy. More measures are used. It is thought that early on, these traits can be shaped and that advantage can be



promoted by essentially broadening the scale, the measures, and thinking about how the dynamics of skill formation unfold.

The broad message then is to have much more inclusive measures. A subsidiary message though, which I know is appreciated in Korea but should be more implemented in terms of policy, is that focusing only on the test scores really perverts the nature of education. Earlier today I was talking about Confucian education, which was a traditional value in many Asian societies. That education was not just about a test score. It was about a whole range—not only abilities to

perform in terms of cognition but also how to treat others, respect, conscientiousness, and persistence on tasks. Many other aspects of this get neglected if one focuses only on the test score. So that is the sense in which I am trying to suggest a broadening of educational policy. The success or failure of Korea, Chile, or any other country is not rated on PISA scores, which are only measuring one aspect of the life of a child and the health of the educational system. Education should be thought of as producing a vector of skills, a broad collection of these skills, and I don't think the current educational policy in many countries does this. It certainly doesn't in the United States to the extent that it should, and my perception is it also isn't doing so in Korea.

Question 4_ Does your research support programs like “Head Start?”

James Heckman_ “Head Start” itself is actually imperfectly implemented. Parts of Head Start are a version of the Perry preschool program. Except it's done and funded much more poorly. It's much less rigorously enforced than the Perry program.

A large fraction of children in the United States grow up in families with a single parent. Seong's research has shown

that these parents are actually given many fewer resources in terms of both cognitive and non-cognitive stimulation. These have cumulative effects over the lifetime of a child. If the system builds early and starts early, it's much easier to prevent the problem than to remediate the problem later in the lifecycle. So it really would suggest somewhat of a re-emphasis on early family years, especially for children in disadvantaged families where many of the traditional values are not taught or they are taught to a lesser degree. I think in that sense it would refocus the policy discussion. It is important to understand that these skills that are so important for performance in schools actually can be created even before children enter school. It broadens the notion of education. So it's something like Head Start, but an enriched version of Head Start.

Question 5_ One dilemma is that from a policy perspective it is a bit difficult to come up with a concrete policy, interventions, and tools to effectively promote soft skills, especially in the older years. Some factors are known to be very important for promoting soft skills of young children, such as good parenting or caring home environments. But from a policy perspective, those areas are not so easy to affect with a concrete policy, tool, or program. Would you please talk a little bit about some examples of concrete

policy tools or programs, especially in the early years, which improve the development of soft skills, especially of disadvantaged children?

James Heckman: Well, I mentioned one already. This is the Perry preschool program. But there are other programs as well. Many such programs have been put in place. Another program that I am looking at now is something called the “ABC Program”, which was implemented in North Carolina in the 1970s. This program started with very disadvantaged children at birth, first eight or nine weeks of life, and followed through for eight years. There are other programs like the “Nurse-Family Partnership” program. There are many such programs that have had successful interventions that have been evaluated by randomized trials with long-term follow-up. They have shown that intervention is possible both with the parents and the child, and even together, improving parenting skills, improving the attachment relationship between the child and the parent, leading to greater school success, greater academic achievement, and greater levels of achievement generally, not just in academic life.

So it is understood that effective intervention is possible. In fact, it is not just a matter of poverty as conventionally measured. I would argue that one of the worst measures,

well, one of the most imperfect measures, of poverty is thinking of poverty as just simply dollars. For a child’s wellbeing, it is not just a matter of dollars, it is also a matter of the quality of parenting. With interventions that are currently being conducted, it is shown that the quality of parenting can be improved. The parents can be engaged.

There are programs that are now in place. Some of them have undergone long-term evaluations. There are many other programs that are being proposed and currently being evaluated that are teaching things like self-control. There is a program that is very actively engaged in many countries around the world called “Tools of the Mind”. “Tools of the Mind” teaches self-control and anger management. This Perry program that I talked about, you can think of it as an application of some of the same principles in Tools of the Mind. This program has now been shown more in short-term evaluations because it is more recent in its application, but in studies all around the world—in Chile, in the United States, in Europe—it has been found that teaching children to stay on task, to cope with their interactions with their fellow students, and so forth leads to enhanced capacities. So I think there is some understanding about these programs.

It’s not to say that these programs cannot be improved.

Every social policy can be improved. But you see, to think that there are no policy steps is a mistake. There really are some successful policies that have been put in place that can be implemented. They do provide some promise for how to promote skills, will promote education, and they will promote success in a number of dimensions.

Question 6_ I have a question regarding soft skills and their relationship to being multilingual or bilingual. There is a phenomenon now called “tiger moms,” the Americans who are sending their children to China and teaching them Mandarin. Do you think such parenting or such a way of education really helps them develop soft skills?

James Heckman: Well, the “tiger mom” is a phenomenon in itself. I think there is a difference between the parent who is an effective parent and sort of directs the child, and a controlling parent who might over-direct. I think that has been the distinction that has been drawn in the discussion, for example, of the “tiger mom.”

It is known that there is a very important role for parenting. You don’t need me to tell you that. But the point is that in many societies, and I even have seen some evidence recently in Korean society, many children are in one-parent

homes. It is a less of a problem in Korean society than in American society. But still, in Korean society there is a small but growing group of families where children are in one-parent homes. It is known from group work that those environments have traditionally provided less stimulation. There are direct measurements. Seong has done work on this, looking at the amount of encouragement of the child, for example, reading to the child, taking the child to the zoo, not ordering the child to go to the zoo but encouraging the child in various ways. I am sure you can over do anything, but I think you can have an effective parent and have effective parenting that essentially provides these resources. I think it is those resources that are important, and those resources can be measured.

In fact, the “tiger mom” may actually create an amount of hostility, too much hostility, by smothering the child to death. On the other hand, if a “tiger mom” environment was compared to an environment of a disadvantaged inner-city child, say in Chicago or in many U.S. cities where there is almost no parenting whatsoever, I think the “tiger mom” is a much better alternative than no mom at all, or a mom with no parenting skills.

But I think an intermediate chord should be struck. I know

that one mother, a Chinese-American mother who was accused of being a “tiger mom,” drew a very fine distinction in an editorial in a major American newspaper between somebody who is over-controlling and others who are providing a rich environment. I think parenting is a very difficult business. It’s very difficult and there is no good book on it. But parenting actually can succeed by interplay with the child. One of the basic rules is attachment of the parent to the child. It is important that the mother, it’s primarily the mother in the earliest years, essentially have a relationship with the child and encourage the child. Those are very basic principles.

Question 7_ We Koreans often hear that Korea is in first place, in terms of math, among 33 countries, or in second place in writing, etcetera. In that way, I think international institutions are promoting a concept of what good education is or where the countries are in terms of school performance, and national governments are taking that measure as an indicator of where we are [and] how good our education is. Are there any policy recommendations or initiatives made by international institutions to deal with this kind of importance of non-cognitive traits that explain success in one’s lifetime?

James Heckman_ I agree that many economists have been partly but not solely responsible, or have played a major role anyway, in promoting the idea of coming up with an objective measure of accountability. I mentioned earlier the notion of a body count in Vietnam; that is a particularly dramatic gesture. Many people feel strongly that if you cannot quantitatively assess something, it is not worthwhile. I have a tendency to believe that as well. But on the other hand, it has been known for years, as shown by studies of centralized systems, that if a society’s resources and incentives are misdirected people frequently do the wrong things.

There is a famous example, somewhat off-topic but related to incentives. There was the famous case of the thousand-pound nail in Soviet factories. Factories were given the incentive to produce a certain number of tons of nails or kilograms of nails, so one factory produced one thousand-ton nail because that was its quota. It wasn’t a question that they wanted nails that could be useful down the line.

Incentives matter and mismeasurement produces distortion. For example, in No Child Left Behind, where the idea was to essentially measure the performance of schools and teachers, scandal after scandal has occurred where teachers have now been brought to task because they have cheated on tests,

because they, their principals, and their schools are given funds to essentially perform well on the tests.

A bad measure or a measure that can be distorted can actually misdirect the activity. For example, I worked on a study years ago in the Reagan administration and there was something called the “Job Training Partnership Act.” “The Job Training Partnership Act” involved training programs for individuals in the workforce, adults, or at least late-year adolescents. They were paid on the basis of the performance of people in those programs and their placement. Their final measure was a placement rate, not a test score, but a job in this case. So they screened people into the program in the first place so they would be well placed—“cream skimming”, it’s called. Secondly, they made sure that the people who had failed in the program were actually kept on the logs of the program for years so they never had unsuccessful performance. Many government bureaucrats fail to understand the subtlety of human beings in response to incentives.

Now in the case of test scores like the PISA test scores or the No Child Left Behind test scores, there is an old joke; it’s a trite joke about the drunk looking under the streetlamp for his keys even though he lost them somewhere else. A

misdirected incentive can be worse than no incentive at all. What I am suggesting is not something that dramatic, but something that is partially that dramatic. I am suggesting that there is essentially a focus on only one aspect of human achievement and there are many other important dimensions that should be considered, and that actually are highly predictive of success in life. So in that sense I think that policy conclusions should be clear. There should be much broader inventories.

The irony about the PISA tests is that instead of measuring cognition and, knowledge of math and reading, some of these achievement tests, and some of these other tests turn out to be capturing imperfectly the measures of motivation and personality anyway. Why not directly come out and measure these traits? Why not look at measures of conscientiousness, the ability to stay on task, all of which are predictive? The existing measurement systems can essentially be supplemented and more responsibility can essentially be created at the school level to produce the complete person—the person that Horace Mann talked about, that Binet talked about, that Tyler talked about, that virtually every educator who has ever thought about education as an important phenomenon has talked about. Confucius too, right? So there is a very broad sweep of individuals, from test designers to

Confucius, all saying the same thing: education is a much more comprehensive activity and the output of education should really be measured completely.

Many people would object to the Korean incentive system by saying it provides an overly focused attention. Some people may survive this system. I am sure many people in this room have. They have been able to escape it by virtue of personality, by virtue of their perseverance, and so forth. Nonetheless, the distorting quality is apparent, especially in other societies, and maybe in an emerging Korean society where there is more inequality and disadvantage, or where some of the traditional family values are now under threat.

I think about how to promote those traits in a society where there are single-parent families, about what schools should do, what families should do, and how skills really get produced. I think it is a very incomplete accounting system, the PISA scores. I completely agree that in some sense this obsession with the test score is really misdirecting society's efforts to educate and devise an effective educational system. I think that is an important message.

※The views expressed herein do not necessarily reflect the views of the Asan Institute for Policy Studies.

