韓國心理學會誌: 産業 및 組織
Korean Journal of Industrial and Organizational Psychology
1998. Vol. 11, No. 1, 97-112.

신입사원 선발용 실용적 지능·검사의 변별타당도

김 명 소

호서대 경영·산업심리학부 산업심리학전공

본 연구는 전통적인 지능이론의 '확장과 변화'라는 맥락 속에서 실용적 지능(Practical Intelligence: 이후 PI)의 구성개념을 이해하고 선행연구(김명소, 김명언, 이도형, 1996)에서 개발한 PI검사를 실제 인사선발 상황에 적용하여 그 타당성을 살펴보는 데 목적이 있다. 구체적으로 선행연구에서는 Sternberg의 삼위일체 지능이론에 근거한 묵시적 지식(tacit knowledge)을 이용하여 신입사원 선발용 PI검사를 개발하였고 자율채용 응시자 및 현직자를 대상으로 집단간 차이분석 및 상사평가와의 상관관계분석을 통해 구성개념 및 준거관련타당도를 검토하였다. 본 연구는 3차례에 걸친 실제 채용장면에서 얻어진 PI검사와 학업지능(Academic Intelligence: 이후 AI)검사, 성격검사, 면접, TOEIC, 학력 등의 기존선발변인들과의 변별타당도를 중심으로 구성개념 타당도를 재확인하는 데 그 목적이 있다. 연구결과 전통적인 AI검사와는 평균 .26의 약한 상관을 보여, AI검사와 중복되는 면이 크지 않은 반면, 성격검사와의 상관은 기대한 대로 무시할만했고 면접점수 및 TOEIC점수와도 .10미만의 낮은 상관을 보였다. 그 외에도 학력과는 .26의 상관을 보였는데, AI 하위검사들에 비해서는 비교적 학력의 영향을 덜 받는 것으로 나타났다. 또한 출신학교의 등급 및 전공계열별로 차이가 없었다. 따라서 본 PI검사는 종래의 선발상황에서 많이 사용되었던 예측변인들과 구별되는 고유한 특성을 측정하고 있다고 볼 수 있다. 끝으로 이러한 결과를 바탕으로 그 의미와 제한점을 논의하고 앞으로의 PI연구 주제를 제시하였다.

어떤 조직에서든지 생산성을 극대화하기 위해 가장 심각하게 고려하는 요소들중 하나는 '업무에 적합한 적성이나 능력을 소유한 우수한 인재를 어떻게 효과적으로 선발할 것인가'이다. 우수인력확보는 기업의 생존과 발전에 직결되는 것으로 타당한 선발도구를 사용하여 인재를 채용하게 되면 그 조직에 상당한 경제적 이익을 가져다 주지만 예측력이 낮은 선발도구를 사용하여 부적절한 사람들을 선발하게 된다면 그 손해는 막대할 것이다(Hunter & Hunter, 1984).

특히, 최근 우리기업들은 신입사원 채용시험에서 다양한 선발도구들을 사용하고 있는데 종래 암기위주의 상식이나 전공시험 대신 면접과 함께 직무능력(적성)검사위주의 채용방식을 점차 늘려가고 있다. 즉 기업들의 채용경향이 과거 '추려내기' 방식에서

*. 한국심리학회 1997년 연차학술대회 산업 및 조직심리학분과 소 심포지움에서는 발표자들간에 practi-cal intelligence(PI)를 잠정적으로 '실제적 지능'이라고 번역하기로 하였으나, 본 논문에서는 1996년 산업 및 조직학회지에 실린 선행연구에서 사용한 용어와 일관성을 유지하기 위해 '실용적 지능 혹은 능력'이란 용어를 계속 사용하였다.

벗어나 '골라내기' 방식으로 선회한 때문에 대기업들이 앞다투어 업무수행능력을 종합적으로 평가할 수 있는 적성검사를 필기시험의 대체 또는 적어도 보충수단으로 채택하고 있다(중앙일보, '97. 10. 7). 이러한 상황에서 효과적인 인사선발을 위해 어떠한 형태의 적성검사 혹은 적성검사들의 조합이 효과적인 예측력을 극대화할 수 있는지 연구하거나 예측력이 뛰어난 선발용 적성검사를 개발하고 타당화하는 일이 오늘날 산업심리학분야에서 매우 중요한 과제가 되었다.

선발도구들중 특히 어휘력, 추리력, 수리력, 공간지각력 등과 같은 하위검사들로 구성된 전통적인 AI검사는 신입사원 선발도구로 종래에 가장 많이 사용되어왔는데, 이 검사가 본래 목적했던 학교에서의 성공이외에 다른 영역들에서의 성공, 특히 직무수행이란 준거변인에 대한 예측변인이 될 수 있는가에 관한 연구는 수십 년간 활발히 진행되었다. 그 결과 AI검사가 직무성공을 전혀 예측할 수 없다는 입장부터(McClelland, 1973, 1993) 가장 좋은 예측변인이라는 주장까지 있어 (Schmidt & Hunter, 1981; Ree & Earles, 1992) 산업심리학 분야에서 뜨거운 논쟁거리가 되어 왔다. 일반적으로 AI검사의 예측계수는 .20~.25로 보고되고 있는데 (Wigdor & Garner, 1982), Schmidt 와 Hunter(1981, 1992, 1993)를 중심으로 한 연구자들은 타당도 일반화 연구에서 메타분석을 통해 범위의 제한, 표본수, 준거변인의 비신뢰성과 같은 오차요인들을 수정한 결과, 약 .50의 예측계수를 보고하고 있다. 위에서 나타난 예측계수들의 차이는 실제 예측력의 차이이기보다는 오차변인의 수정여부에 의한 것으로 생각할 수 있다. 그러나, 아무리 인위적으로 통계적, 측정적 오차가 없다고 가정해도 준거변인에 대한 설명력은 20~25%를 넘지 못하고 있는데, 선발도구의 예측력을 높이려면 아무리 시간과 비용이 드는 작업이라 할 지라도 직무수행과 같은 준거변인의 나머지 분산을 설명해 줄 수 있는 추가적인 변인들을 밝혀내는 연구는 계속 되어져야 하며 이에 대해서는 양편의 연구자들이 모두 동의하고 있다.

일반적으로 AI검사 외에 기업에서 신입사원들을 채용할 때 사용하는 선발변인에는 성격검사, 흥미검사, 생활사 혹은 전기자료(biodata), 전공시험, 면접, 학력, 학교성적, 모의실험(simulation), 평가센터(assessment center) 등이 있다. 그러나 이러한 다양한 선발도구들은 Hunter와 Hunter (1984)가 지적한 바와 같이 사실 각각 서로 다른 다양한 특성 혹은 내용을 측정하기보다는 동일한 혹은 유사한 내용을 다른 방법으로 측정하고 있을 가능성이 높다. 예를 들면, AI검사에서 측정하고 있는 능력을 모의실험이나 평가센터에서도 중복되게 다루고 있고, 성격검사의 내용들을 생활사와 면접에서 측정하고 있다고 볼 수 있다. 물론 모의실험이나 평가센터와 같은 방법은 개인 혹은 소규모 집단단위로 실시하므로 시간과 경비가 많이 소요되어 매우 신중을 기해야 할 경력사원 및 중견사원들을 채용할 때 주로 사용된다. 한 변인을 측정하기 위해 여러 가지 다른 방법을 개발하는 것도 중요하겠지만, 그보다 더 시급한 일은 기업에서 요구되는 다양한 특성 혹은 내용을 밝혀내어 선발에 반영하는 것이다.

특히 AI외에 직무성공이란 준거변인을 예측하는 다른 능력변인들의 가능성에 관한 논의와 그 구성개념들을 밝혀 보려는 연구는 오래 전부터 연구자들이 관심을 가져온 주제인데, McClelland(1973)로부터 본격적으로 시작되어 1980년대 이후 Sternberg와 그의 동료들에 의해 체계화되었다고 볼 수 있다.

McClelland는 1973년 「American Psychologist」에 실린 논문에서 IQ유형의 AI 혹은 일반지능검사의 제한점을 체계적으로 신랄하게 지적하고 또 6개의 심리검사 개발방향 및 대안을 제시하였다. 그에 따르면 산업장면에서 심리검사의 예측력을 향상시키기 위해서는 선발변인으로 일반지능 혹은 AI란 개념보다는 실용적 역능 혹은 능력(competence)이라는 개념이 더 적절하며 이는 타고난 잠재능력이 아니라 경험에 따라 변화와 성장이 가능한 다차적(higher-order) 기본특성이다. 또한 McClelland는 이를 타당하게 측정하기 위해서는 단순한 반응적 행동(respondent behavior)보다는 조작적 행동(operant behavior)에 초점을 맞추어 준거표집(criterion sampling)을 해야한다고 주장했다. 즉, 검사가 조직에서의 직무수행능력을 제대로 예측하려면 검사의 내용적 틀이나 형식이 구체적인 현실세계와 유사해야 하며 직무에서의 실제문제 해결능력 즉, 준거행동들을 포함시킬 때 예측타당력이 더 높아진다는 것이다. 뿐만 아니라 McClelland는 검사가 준거행동을 직접적으로 측정하고 있다면 거짓반응으로 높은 점

수를 얻을 수 있는 가능성(faking)이 줄어들 수 있다고 주장하였다.

그의 준거표집 원칙을 AI검사에 적용시켜보면 AI검사가 학업성적을 성공적으로 예측할 수밖에 없는데(r = .40~.70), 이는 학교에서 교사가 원하는 혹은 교사가 높은 성적을 주는 능력이나 기술이라는 준거들을 측정하고 있기 때문이다. 반면, 학업적 기초능력은 현실사회나 기업조직에서 요구되는 다양한 지적 능력들 중 일부분에 지나지 않아 AI검사는 충분한 준거표집이 되고있지 못하며 따라서 실제 직무관련수행을 예측하는 정도가 떨어질 수밖에 없다는 의문이 자연스럽게 제기되는 것이며 추가선발변인에 관한 연구의 필요성을 시사하는 것이다. 일반적으로, 학교장면과 좀 더 유사하다고 볼 수 있는 신입사원 교육에서의 수행과 AI검사와의 상관은 실제 직무수행과의 상관보다 약 두 배가 높은데, 이러한 결과는 선발검사의 준거표집 정도에 따라 예측력이 달라진다는 McClelland의 주장을 지지해준다.

McClelland의 제안들은 그당시 학자들뿐 아니라 일반인들에게까지 크게 영향을 미쳤으며 전공서적외에 신문·잡지 등 언론에서까지도 IQ검사의 한계를 지적하는데 공헌했고 IQ검사에 관한 일반여론을 이끌어 갔다. 그러나 McClelland의 제안에 관해 여러 학자들에 의한 비판도 만만치 않은데, 두가지로 요약해 볼 수 있다(예: Barrett & Depinet, 1991; Day & Silverman, 1986; Hunter, 1986). 첫째, McClelland의 제안중 가장 근본적인 문제점은 역능·능력의 개념을 명확하게 정의하는데 실패했다는 것이다. 협의의 역능·능력은 직무분석 혹은 행동분석을 통해 관찰된 각 직무의 개개 업무를 의미하고 있어 본인 자신도 일반적인 역능·능력으로 보완하려 노력했다. 반면, 광의의 역능·능력은 인지능력 및 성격, 동기와 같은 비인지 능력 모두를 포함하고 다른 선발변인들과 구별되기 어렵고 따라서 독특한 기여정도를 파악할 수 없게 된다. McClelland의 연구를 종합하여 논의한 Boyatiz(1982)도 역능·능력이란 개인의 기본특성중 동기, 특질, 기질, 자아상, 사회적 역할, 지식 등이 될 수 있다고 언급하면서 명확한 구성개념 설명을 하고 있지 못하다.

둘째, McClelland는 중요한 검사개발방향의 원칙만을 제시해놓고 이 원칙들을 지지하는 경험적 연구들을 수행하는데는 실패한 것 같다. 1973년 McClelland가 처음 역능·능력검사와 준거표집을 주장한 뒤, 약 25년 정도가 지난 현시점에 이르기까지 그렇다할 역능·능력을 측정하는 타당한 방법이나 도구에 관해 구체적으로 언급이 되고 있지 않다. Boyatiz (1982)는 역능·능력을 밝혀내기 위해서 BEI(Behavioral Event Interview)라 불리우는 중요사건접근법에 의해 현직자들을 인터뷰하고 그들이 직장에서 자신감이나 효용감을 경험한 사건 세 가지와 그 정반대의 경우 세 가지를 나열하게 하였다. 그리고 「Picture Story Exercise」라는 도구를 제작해 역능·능력을 측정하려고 시도했으나 구조화된 상황에서의 객관적 측정방법은 구체적으로 제시하지 못했다. 또한 역능·능력검사에 관한 참고문헌자료로 가장 많이 인용되는 Klemp와 McClelland(1986)의 연구에서도 27명의 중견간부들을 대상으로 간부급에게 필요한 역능·능력들로 자신감, 계획 혹은 기획력, 인과적 사고력 등을 지적했으나 구체적인 선발도구를 개발하거나 타당화하는데는 실패했다.

McClelland의 역능 혹은 능력개념과 관련된 변인들중 수 십년간 개인의 실생활의 적응 및 성공에 꼭 필요한 지능의 영역은 사회적 지능이란 용어를 중심으로 다양한 용어로 쓰여졌고 학자들에 따라서 여러가지 다른 의미로 해석되었다(김명소, 김명언, 이도형, 1996). 그러나 사회적 지능은 그 용어의 중요성에 비해 정의가 체계적으로 확립되지 못하여 그 개념 자체가 모호해져 버렸다. 즉, 좁게는 대인관계능력으로부터 넓게는 도덕적 사고나 건설적 사고까지 포함하는 매우 이질적인 특성을 보이고 있다. 아울러 사회적 지능을 어떻게 측정하고 평가할 것이냐에 관한 연구는 체계적이지 못하고 산발적인 연구만이 존재하였다. 게다가 이와 관련된 대부분의 검사들에 관한 타당성연구나 활용보고서등이 별로 없고, 몇몇 안되는 타당성에 관한 연구결과도 일관성이 없다. 예를 들면, 사회적 지능검사들은 다른 사회적 지능측정치들과 유의미한 상관을 나타내지 않는 반면에 오히려 IQ와 더 높은 상관을 나타내 AI와 구별되는 개념으로 제시되지 못했고, 사회적 수행능력을 측정하는 대부분의 준거변인들과도 상관이 없는 것으로 보고되었다(Cronbach, 1960; Matney & McManis, 1977; Osipow & Walsh; 1978, Sternberg & Smith, 1985).

따라서 90년대 이후부터는 사회적 지능이라는 개념보다는 이를 좀 더 체계화, 세련화시킨 개념으로

실용적 지능(PI), 정서적 지능(Emotional Intelli-gence: 이후 EI)이 더 강조되고 있다. 우리나라에서는 최근 EQ선풍이 불어 EI쪽에 더 많은 관심이 집중되고 있지만, PI는 다른 개념들을 폭넓게 포함하는 개념이고, 특히 직무성공에 대한 예측변인으로 EI에 관한 경험적 연구결과는 보고되고 있지 않으며 따라서 산업심리학분야에서는 PI연구가 EI연구보다 훨씬 더 활성화되어 왔다고 볼 수 있다.

PI연구는 최근에 와서 Sternberg를 중심으로 하는 연구자들에 의해 실행되어 왔는데, 학업성공을 예측하는 전통적인 AI검사가 학교밖에서의 사회적 성공을 예측하는 분산이 아주 적다는 선행연구들을 바탕으로 사회적 성공을 보다 잘 예측할 수 있는 새로운 선발개념으로서 실용적 혹은 실무적 지능(PI) 또는 상식(common sense)의 개념을 도입하면서 활발해지고 있다.

Sternberg(1985, 1995)에 의하면 다음의 세 가지 이유 때문에 PI와 AI는 구별된 개념으로 간주되어야 한다. 첫째, 학업적 문제와 실제생활의 과제 혹은 문제 자체가 서로 다른 특성을 보인다는 것이다. 학교장면에서 요구되는 능력이나 전통적인 지능 검사에서 다루는 문제는 (a) 타인에 의해서 형성되고 (b) 일반적으로 잘 정의되어 있고 (c) 시작부터 문제해결을 위한 정보가 모두 주어지고, (d) 일반적으로 하나의 정답이 있으며, (e) 일반적으로 정답의 해결 방법이 하나이고, (f) 개인의 일상 경험과는 동떨어져 있고, (g) 개인에게는 내재적 관심이나 흥미가 결여되어 있는 상황이다. 반면에 직무와 같은 실제 상황에서 부딪치는 실용적인 문제들의 특징은 (a) 각자가 문제가 무엇인지 인식하고 규정해야 하며 (b) 대체로 체계적으로 정의되지 못하고 (c) 문제해결을 위한 정보가 부족하여 많은 정보를 얻는 과정을 필요로 하며 (d) 일반적으로 각각 장단점이 있는 여러 가지 정답이나 대안이 있고 (e) 일반적으로 정답해결방법이 여러 가지이고 (f) 일상생활에서 얻은 사전 경험을 이용할 수 있으며, (g) 개인이 관심을 가지고 동기화되며 그 문제에 몰두하게 된다(Neisser, 1976; Wagner & Sternberg, 1985).

둘째, 사람들이 생각하는 지능의 개념을 조사해 보면 PI와 AI는 구분된다는 점이다. Sternberg, Conway, Ketron, 및 Bernstein (1981)은 암묵적(implicit)[1]지능개념 연구에서 지적인 사람들의 원형적(prototype) 행동을 구하기 위해 일반인이 생각하는 지적인 행동특성들을 나열하게 한 후, 이를 바탕으로 설문지를 작성하여 각 문항에 제시된 행동이 이상적으로 지능이 높은 사람의 행동을 얼마나 잘 나타내는지 지능연구 전문가와 일반인들의 평정(9점척도)을 요인분석한 결과, 전문가 집단에서는 언어력, 문제해결능력, 실용적 능력요인이, 일반인 집단에서는 실용적 문제해결능력, 언어력, 사회적 혹은 실용적 능력이 도출되어 전문가뿐만 아니라 보통사람들도 PI와 AI를 구분한다는 것으로 밝혀졌다.

셋째, 생애에 걸친 발달패턴에서 PI와 AI는 차이가 있다는 것이다. 성인들의 경우 나이가 들수록 AI는 감소하지만 일상적인 문제해결능력이나 재정관리 능력과 같은 PI는 계속 증가하는 경향이 있으므로 (Williams, Denny, Schadler, 1983), PI와 AI는 같은 능력으로 볼 수 없다고 한다. PI의 발달단계에 관해서 전통적인 지능이론의 대표주자인 Horn과 Cattell (1966)의 인지요인들과 연결시켜 보면, 특정 문화권에서 축적된 지식을 반영하는 결정지능(crystallized intelligence)이 학업적 문제해결 보다는 실제 문제해결과 더 관련이 깊고, 반대로 추리나 새로운(novelty) 문제해결에 필요한 유동지능(fluid intelligence)이 학업적 문제해결에 더 관련이 깊다. 많은 연구 결과, AI를 대표하는 유동지능은 성인가운데 나이가 들수록 줄어들지만 결정지능 및 PI는 성인후기에 가서도 경험·연령에 따라 유지 혹은 증가되고 있어, 이 두 지능을 구분할 필요가 있다는 주장이 제기되었다(Horn, 1982; Schaie, 1978; Sternberg, Wagner, Williams, & Horvath, 1995).

Sternberg와 Wagner(1993)는 PI를 체계적으로 정의하고 측정하기 위해서 Sternberg의 삼위일체 이론(특히, 상황적 하위이론)의 틀 안에서 묵시적 지식(tacit knowledge)이라는 용어를 사용했다. 즉, 산업장면에서의 PI란 자신(self), 타인(others), 업무(task)와 관련된 묵시적 지식을 가지고 산업장면에서 타인의 행동의 의미를 파악하고 주어진 상황에 적응하고

1) 영어단어 'implicit'와 'tacit'이 우리말로 '암묵적' 혹은 '묵시적'으로 함께 번역되어 사용되고 있어 혼돈이 되기도 하는데, 본 논문에서는 'implicit'은 '암묵적'으로, 'tacit'은 '묵시적'으로 번역하였다. 1996년도 선행연구에서는 'implicit'란 단어가 사용되지 않았고 'tacit'이 '암묵적'으로 번역되었었다.

(adaptation), 방해가 되는 긴장이나 갈등의 요소를 적극적으로 해소하기 위해 새로운 선택을 하며(selection), 이를 바탕으로 주어진 상황을 개선하고 건설적으로 만들어 가는 데(shaping) 적절하게 사용하는 능력'이라고 정의될 수 있다. Sternberg의 이러한 PI접근은 지식을 바탕으로 한 입장(knowledge-based approach)을 반영하고 있는데, 이는 문제해결에서 일반적 인지능력이나 인지과정의 효율성보다 특정영역의 지식이 상대적으로 더 중요하다고 보는 것이다. 즉 개개인의 지능의 차이란 그들이 소유한 지식구조의 차이 때문에 발생한다는 생각이다.

묵시적 지식이란 Polanyi(1976)에 의해 처음 언급되었는데, PI가 AI와 구별되는 또 하나의 중요한 이유로 생각될 수 있다. 즉, AI는 형식적 지식을 주로 다루는 반면, PI는 비형식적, 묵시적 지식을 다룬다. 학교장면에서 중요하게 생각하는 AI와 관련있는 지식은 내용(content)과 법칙(rule)에 관한 것으로 형식적(formal)이다. 이 지식은 읽기와 듣기 같은 과정을 통해 주로 학습되고 학교장면에서 높이 평가되며 전통적 지능검사를 통해 측정된다. 반면 실용적 문제들을 해결하는데 필요한 지식은 학업적이기 보다는 실용적이며, 형식적이기 보다는 비형식적(informal)이며 사회규범(norm)과 관련되어 있다. 따라서 직접적으로 교육받기보다는 간접적으로 일상생활의 경험(즉, 관찰이나 모델링)을 통해 얻어지는 절차적 지식으로 대부분의 학교장면에서 중요하게 생각되지 않는다. 따라서 개인이 자신이 사용하고 있는 묵시적 지식이 무엇인지 분명히 기술하고 전하는 데는 어려움이 많고 자신들에게 어떤 묵시적 지식이 필요한지 등을 잘 이해하지 못한다. Boyatiz (1982)는 McClelland와 동료들의 연구를 정리하면서 직무역능 혹은 유능이란 개념안에 개인이 사용하는 지식을 포함하고 있는데, 이러한 개인의 특성들은 본인이 의식하고 있는 것이 아닐 수도 있어, 묵시적 지식처럼 그것이 무엇인지 혹은 어떻게 습득되었는지를 분명히 기술하거나 설명할 수 없다고 한다.

Sternberg와 Wagner는 그들의 대표적 PI검사인 TKIM(Tacit Knowledge Inventory for Managers)을 개발하는데 PI측정방법으로 유능자·초보자 차이방법을 사용하였다. 직업에 따라 묵시적 지식을 이용한 PI검사의 자세한 절차들은 약간씩 달라질 수 있지만, 일반적인 유능자·초보자 차이방법의 제작 및 활용 절차는 다음과 같다(Sternberg, 1995).

첫째, 특정직업 혹은 직무에 필요한 묵시적 지식들이 무엇인지 대체적으로 살펴보기 위해 그 직무에 관한 문헌조사(예: 직무기술서, 직무분석)를 실시하여 관련된 이론적 틀이나 경험적 연구결과가 있는지 조사한다. 둘째, 추천과정을 통해 각 직무영역에서 경험이 풍부하고 성공적인 우수자 집단을 선별한 뒤, 그들과의 인터뷰를 통하여 조직에서 빈번히 일어날 수 있는 상황들과 대처방법들을 알아본다. 즉 그 직무에서 성공적으로 적응하는 과정에서 그들이 겪었던 힘들었던 상황들, 성공담, 교훈을 얻었던 실수, 동료들을 관찰한 경험, 신입사원이나 부하직원에게 하고 싶은 충고 등에 관해서 묻는다. 때에 따라서는 성공적인 사람들뿐 아니라 조직에 잘 적응하지 못하는 사람들도 인터뷰하여 직무성공에 방해가 되는 부정적인 묵시적 지식의 요소들도 파악하여 오답 혹은 바람직하지 못한 반응으로 사용할 수 있다. 셋째, 문헌조사와 인터뷰를 한 결과를 토대로 직무의 여러 영역(즉, 자신, 타인, 업무를 포함)에 걸쳐 적용되는 예비 시나리오(상황)와 반응대안들을 작성하여 예비검사를 제작한다. PI검사의 각 문항은 직무와 관련된 시나리오와 5~20개의 반응대안들을 포함하게 된다. 응시자는 각 대안에 대해 효율성 정도에 따라 7~9점척도(전혀 바람직(중요)하지 않다 ~ 매우 바람직(중요)하다 혹은 자신의 행동과 매우 비슷하다 ~ 전혀 다르다)로 평정하게 된다. 넷째, 위 예비검사를 전문가들 집단에 보내어 피드백을 얻는다. 이 전문가들 집단안에는 인터뷰 대상이었던 성공자들외에도 특정 직무영역과 관련있는 학자들, 심리학자들이 포함된다. 다섯째는 예비조사를 하는 단계인데, 여러가지 준거측정치(예: 상사평가)를 바탕으로 해당 조직이나 직무에서 우수자 집단과 우수하지 못한 집단을 선발한 후 실제 직무장면에 적용하여 검사문항들 중 우수집단과 우수하지 못한 집단을 적절히 변별하는 대안들을 골라낸다. 즉 변별력이 없는 시나리오 문항이나 반응대안들은 수정하거나 삭제한다. 여섯째, 본 조사 작업을 실시하여 검사의 신뢰도나 타당도를 조사해 본다. 가능하면, 준거관련 타당도뿐 아니라 수렴 및 변별타당도에 관한 정보를 얻는다. 마지막으로, 실제 선발상황에 적용하기 위해 필요한 규준자료(우수자 집단의 프로파일)와 통계적 자료들을 수집한다.

이제까지 기술한 Sternberg와 그의 동료들을 중심으로 한 PI의 구성개념 및 검사제작 방법들은 1970년대에 McClelland가 제시한 원칙들에 대체적으로 부합된다. 특히 검사제작절차 및 문항특성들은 McClelland가 구조적 틀안에서 조작적 행동을 측정할 수 있는 대안으로 제시한 것을 그대로 실천한 경우에 해당하며 따라서 McClelland의 제안중 Sternberg의 연구에서 가장 근접하게 실행된 것이 바로 이 부분이라고 할 수 있다. McClleland가 미래 검사 개발방향에 대한 아이디어를 제의했지만 뒷받침할 만한 경험적 연구를 보고하는 데는 실패했다면, Sternberg와 그의 동료들은 McClleland의 제안을 약 20년이 지난 후 PI이론 및 구체적 측정도구 개발에 적용하여 체계적으로 실천하고 발전시켰다고 결론내릴 수 있겠다. 이제까지 논의된 McClelland의 검사개발방향에 관한 제의와 Sternberg의 수행 및 실천과정을 다음의 표로 요약해 보았다(표 1).

이상에서 논의된 PI에 관한 관점들은 최근 지능이론 분야의 두드러진 특징과 변화인 지능을 현실세계와 밀접한 구체적 상황내에서(Cantor & Harlow, 1994; Cantor & Kihlstrom, 1987, 1994) 그리고 비교적 독립적인 것으로(Gardner, 1983, 1987; Sternberg, 1996) 이해해야 한다는 점을 잘 반영하고 있다. 따라서 종래의 IQ유형의 AI검사에서 측정하고 있는 개인의 학업적 지능차이를 일상생활이나 직무와 같은 다른 영역에서의 능력차이로 일반화할 수 없다는 것이 자명하고 기업조직의 선발상황에서 실용적 능력의 필요성이 더욱 부각되었다. 특히 수 십년간 신입사원 채용에 활용되어온 다양한 선발변인들을 검토해본 결과 예측력이 있거나 그 필요성을 인정받은 선발도구중 능력을 측정하는 도구들에서는 AI유형의 검사 외에는 특별히 언급되지 않고 있어(김명소, 1997) 요사이 지능개념이 확장되면서 새롭게 제안된 지능개념들(예 실용적 지능, 정서적 지능, 도덕적 지능, 창의적 지능 등)이 다수 존재한다는 점에 비추어 보면, 기존의 선발도구들은 능력의 한정된 측면만을 다루고 있었다는 것이 거듭 밝혀졌다. 따라서 실제업무에서 높은 수행과 적응을 보일 수 있는 잠재력을 가진 인재를 좀 더 효과적으로 선발하기 위해서는 PI검사를 AI유형의 검사에 추가하여 보완하는 것이 요망된다.

이러한 연구의 필요성에 의해 선행연구에서는 21세기 모기업에 걸맞는 인재를 선발하기 위해 기본 학업적성외에도 기업조직상황에서의 직무수행 및 적응능력을 함께 측정할 수 있는 종합적인 직무적성검사를 개발하려는 계획하에 앞에서 논의된 검사제작절차에 따라 신입사원용 PI검사를 개발하였고 자율채용 응시자 및 입사한 지 3년미만 되는 현직자를 대상으로 신뢰도 및 예측 타당도, 특히 공인타당도(concurrent validity)분석을 실시하였다(김명소, 김명언, 이도형, 1996). 그 결과 본 PI검사는 우수자/비우수자를 성공적으로 변별하였고 또 상사평가와의 상관계수가 통계적 교정없이 .32~ .37로 나타나 본 PI검사에서 측정하고 있는 실무적 능력은 모기업 신입사원의 업무수행이나 성공과 밀접한 관련이 있었다고 할 수 있다.

본 연구는 선행연구에서 개발된 PI검사를 실제 채용장면에 적용하여 기존에 사용되고 있는 선발변인들(AI검사, 성격검사, 면접, 학력 등)과의 관계성 혹은 독립성, 즉 변별타당도를 조사함으로써 묵시적 지식을 바탕으로 하고 있는 PI에 관한 구성개념 설명(construct explication)을 시도하려고 하는데 그 목적이 있다.

방법 및 절차

연구대상

95하반기, 96상반기, 96하반기의 모그룹 공채채용 응시자들(5개 직군)을 대상으로 본 PI검사를 적용하여 PI검사와 함께 실시된 AI검사, 성격검사, 면접, TOEIC, 학력, 학교군 등과의 관계를 분석하여 변별타당도 증거를 수집하였다. 총지원자수는 각 공채때마다 약간의 차이가 나지만, 약 18,000~20,000명 정도이다. 응시자 구성비율은 남녀간에 대략 7:3이었고 고졸이하가 약 10%, 전문대졸이 약 25%, 대졸이상이 약 65% 정도 되었다.

도구 및 채점방식

본 PI검사는 선행연구에서 사용되었던 35개의 상황 시나리오중 변별력 높은 25개의 상황을 포함하고 있다. 이 검사는 아직 계속적으로 개발중에 있는 상

표 1. McClelland의 제의와 Sternberg의 수행/실천

	McClelland의 제의(1970년대)	Sternberg의 수행/실천(1990년대)
측정변인 (구성개념)	학업지능 혹은 일반지능과 구별되는 **역능,능력**: 직무의 성공적 수행에 기여하는 ┌ 개인의 동기, 자아상, 적절한 목표 설정 능력, 자기개발 ├ 대화기술, 대인관계기술 └ 읽기, 쓰기, 업무지식 등	학업지능 혹은 일반지능과 구별되는 **실용적 지능(PI)** 특히, **묵시적 지식**을 가지고 상황에 적응, 선택, 개선해 나가는 능력: **역능개념의 체계화** ┌ 자기 ┐ ├ 타인 ┤ 과 관련된 묵시적 └ 업무 ┘ 지식의 사용능력
검사의 표집	· 직무에서의 성공적 수행을 위한 준거행동들을 수집하여 문항들을 작성해야 함: **준거표집** · Boyatiz(1982)는 **중요사건방법**을 이용한 BEI (Behavior Event Interview)를 통해 직무에서 성공하기 위해 필요한 행동특성들을 파악	· **유능자,초보자 접근방법**을 사용하여 실제조직에서 빈번히 일어나는 상황과 대처반응들로 검사문항이 구성되므로 **준거표집**을 하고 있음 · 대처방법들은 특정직무에서 성공적으로 적응하기 위해 필요한 혹은 불필요한 암묵적 지식들을 반영함 · 인터뷰, 문헌조사를 통해 묵시적 지식파악
문항형태	· 구조화된 상황, 즉 명백한 자극(즉, 문항과 보기)이 있고 거기에 단순하게 반응하는 행동보다는 직무와 같은 실제세계에서 빈번히 일어나는 **조작적 행동**을 측정할 수 있는 형태여야 함. 즉, 실생활처럼 주어진 문제해결을 위해 개인이 여러 가지 대안들을 제시하게 하고 그 안에서 가장 효과적인 해결책을 선택하거나 효율성 정도를 평가하게 할 수 있는 형태가 되어야 함 · 조작적 행동이 조작적 행동을 더 잘 예측함	· 각 문항세트안에는 문제상황과 조직에서 우수자 혹은 비우수자가 제시한 여러 가지 가능한 해결대안들이 주어지는데, 이 대안들은 정답 혹은 오답으로 구분되기 보다는 효율성 혹은 바람직성이라는 측면에서 그 정도가 다르므로 응시자들이 어떤 대안을 선택했는지 혹은 각 대안마다 바람직성의 정도를 어떻게 다르게 평정하고 있는지(7점 혹은 9점척도)에 따라 점수가 부여되고 있어, **조작적 행동**을 측정하고 있음 · 응시자의 반응이 특정직무에서 우수자들의 반응과 얼마나 유사한 지에 따라 응시자를 평가함
변화가능성	· 타고난 혹은 변화하지 않는 정신능력을 측정하기보다는 **성장,변화**가 가능한 능력을 측정해야 함 · 검사가 실생활의 행동들을 표집했다면 개인의 검사수행정도는 그의 경험, 지혜, 또는 다양한 삶의 현장에서의 수행능력이 증가함에 따라 올라가는 것이 당연함 · 따라서 검사의 타당성은 이러한 성장, 변화를 반영할 수 있느냐에 달려있음	· 실제 PI검사는 경험(경력), 직급, 부하직원수 등과 유의미한 상관을 나타내고 있어 **성장, 변화**를 측정해 내고 있음
거짓반응성의 경향(faking)	· 응시자가 거짓되게 사회적으로 바람직한 패턴을 따르려는 경향을 줄이기 위해 검사의 목적, 특성, 내용들을 명백히 응시자에게 알려야 함. · 거짓반응경향은 검사내용과 준거의 관련성이 적을수록 증가될 가능성이 있음 · 그러나 일반적으로 안면타당도가 높으면 거짓반응이 증가되는 경향이 있어 이 제안의 적용범위는 제한적임	· PI검사의 문항특성들을 명백히 밝히고, 검사실시 전에 예비문항을 제시하여 하나의 정답 혹은 바람직한 보기가 있는 것이 아님을 강조하면 거짓반응을 줄일 수 있음 · PI검사는 준거표집하여 안면타당도도 높으면서도 특수한 문항형태 때문에 거짓반응을 방지할 수 있어 일거양득.
영역일반화 가능성	· 역능 · 능력이란 단순한 세부적 기술이기 보다는 어느정도 일반화시킬 수 있는 **다차적 잠재속성**을 나타냄	· 묵시적 지식이란 특정영역에 국한되지도domain- specific) 탈영역적이지도(domain- general) 않은 어느정도 영역간 공통적으로 요구되는 영역중심적(domain-centered)인 자기, 타인, 업무의 **3차원으로 구성된 잠재속성**을 나타냄

태이므로 96하공채때는 95하공채와 96상공채때와는 다른 채점방식을 택하였다. 선행연구에서는 한 문항당 여러 개의 응답선지(대안)가 있었고 선지별 점수(7점 만점)가 경험적으로 미리 부여된 것이 있었으므로, 그에 따라 응답결과에 점수를 부여하였다. 즉 선지별 점수와 모기업에서의 우수자·비우수자를 나타내는 가변수와 유의미한 상관을 보이는 문항들에 대해 각 응시자별 합산점수를 구하였다.

96년 하반기 공채에서는 Sternberg가 제시한 세 가지 방법중 경험적 법칙방법(rules-of-thumbs)의 변형된 형태를 사용하였다. 즉, 하나의 상황에 변별력이 우수한 4~5개 응답선지가 주어지고 그 가운데 상황을 해결하기 위한 가장 바람직한 대안(best answer)과 가장 바람직하지 못한 대안(worst answer)이 있다. 이 대안들은 모기업 간부사원들과 면접관들의 평정에 의해 결정되었다. 응시자는 가장 바람직한 대안과 그 반대의 대안을 각각 제대로 고르면 가장 높은 점수를 받게 되며, 거꾸로 응답을 하면 가장 낮은 점수를 받게 된다. 그러나 둘다 옳은 선택도, 정반대의 선택도 아니면 중립선택을 한 경우로 중간 점수를 받는데, 그 채점기준은 다음과 같다.

- 둘 다 옳은 선택: 4점
- 하나는 옳은 선택, 하나는 중립선택: 3점
- 둘 다 중립선택이거나 하나는 옳은 선택, 하나는 정반대로 선택: 2점
- 하나는 중립선택, 하나는 정반대로 선택: 1점
- 둘다 정반대로 선택: 0점

따라서 1개 상황에 대해서 위와 같이 0점에서 4점까지 받을 수가 있고 총 25개 상황에서 100점 만점이 가능하다.

특히 직무적성검사 전체는 AI검사, PI검사, 그리고 모기업에서 자체 개발한 사회상식검사로 구성되었다. 그 중 AI검사는 언어력, 수리력, 추리력, 공간지각력, 정보검색력(95하반기와 96상반기에는 직관력검사) 등 5개 하위검사로 구성되어 있다. 성격검사는 MMPI중 반사회성, 편집증, 경조증, 정신분열증, 타당도 척도, 허위성 척도와 Cattell의 16성격요인검사중 도덕성, 동기왜곡척도, 또 자체 개발한 5요인(정서적 안정성, 외향성, 성실성, 호감성, 지적 개방성) 검사와 사회적 책임감, 작업곤란도 척도를 포함하고 있다. 면접은 두 부분으로 나뉘어졌는데, 인성면접과 프리젠테이션 면접이다.

결 과

AI검사와의 관계분석

PI검사의 변별타당도를 알아보기 위해 우선적으로 AI검사들과 상관계수를 산출하였다. 95년 하반기에 AI검사와 PI검사를 실시한 결과 직군별로는 .22~.29의 상관을 나타내었고 평균 상관계수는 .26이었다. 하위검사별로는 결정지능을 대표하는 언어력검사와의 상관이 제일 높았고, 유동지능을 대표하는 수리력, 추리력, 공간지각력 검사들과는 비교적 낮은 상관을 나타냈다(표 2).

표 2. 95년 하반기 공채시 AI검사들과 PI검사의 상관계수

	언어력	수리력	추리력	공간지각력	직관력	AI총점
PI	.22	.10	.17	.10	.15	.26

아래 표 3은 96년 상반기에 AI검사와 PI검사와의 상관을 직군별로 나타내고 있다.

표 3. 96년 상반기 공채시 AI검사들과 PI검사의 상관계수

	언어력	수리력	추리력	공간지각력	직관력	AI총점
PI	.28	.16	.21	.08	.18	.31

전체적으로 AI검사와는 .31의 상관을 보이고 있는데 직군별로는 .29(기술지원직)~ .36(경영지원직)의 상관을 나타냈다. 하위검사별로 살펴보면 언어력검사와 상관이 제일 높았고, 수리력, 공간지각력 및 추리력검사와는 비교적 상관이 낮아 95년 하반기 공채시와 동일한 패턴이 도출되었다.

또한 96년 하반기 공채적용시에 나타난 AI검사들과 PI검사의 상관계수는 다음과 같다(표 4).

표 4. 96년 하반기 공채시 AI검사들과 PI 검사와의 상관계수

	언어력	수리력	추리력	공간지각력	정보검색력	AI 총점
PI	.37	.21	.22	.09	.25	.40

96년 하반기 결과를 살펴보면, 95년 하반기와 96년 상반기의 결과와 마찬가지로 PI검사는 AI하위검사들중 주로 결정지능을 측정하고 있다고 가정되는 언어력이나 정보검색력과 .25~.37의 비교적 높은 상관을 보이고, 유동지능의 일반적 측정치인 수리력, 추리력, 공간지각력과는 .09~.23까지의 낮은 상관을 보이고 있어 PI와 결정지능 모두 실생활의 문제해결과 더 깊은 관련이 있음을 시사하고 있다.

3번에 걸친 공개채용을 통해 PI와 AI검사간 상관계수는 .26, .31, .40까지 폭넓은 관계성을 제시하고 있어 양 구성개념의 구별성에 대한 결론을 쉽게 내리지 못할 것 같다.그러나 96년 상반기에는 언어력 하위검사안에 어휘력, 문장이해력외에 언어적 감수성, 96년 하반기에는 언어적 감수성외에도 정보검색력 하위검사와 같은 전통적인 AI검사영역에서 벗어난 하위검사들을 첨가함으로써 AI와 PI의 관계성이 예상보다 증가하고 있는 것으로 판단된다. 뿐만 아니라 수리력 검사도 계산능력 위주의 전통적인 문항형태가 아니라 통계자료 해석력이 주도적으로 측정되고 있으므로 전통적인 AI검사와는 어느정도 측정하는 내용이 다르다고 볼 수 있다. 그러므로 3번 실시된 경우중 그래도 가장 전형적인 AI영역들을 포함한 95후반기의 AI검사와의 상관관계로 두 검사의 구별성을 검토해 보는 것이 더 합리적인 듯 싶다. 이 관계성은 5개 직군별 평균이 .26으로, 위의 경우보다 낮아졌으나 여전히 Wagner와 Sternberg (1985, 1990)의 연구에서 제시된 평균 .10의 상관관계에는 못미친다.

그 이유는 아마도 Ree와 Earles(1993)가 지적한대로 표집수가 너무 적은 표본(예: 60명의 Yale대학생)이 Sternberg와 그의 동료들의 연구에 참여했고 결과적으로 매우 제한된 능력범위의 집단을 대상으로 실시되었기 때문에 상관계수가 범위의 제한요인에 영향을 받은 것 같다. 따라서 본 연구대상자의 수가 약 20,000명에 가까운 것을 생각해 볼 때, 이 연구에서 산출된 상관계수의 크기는 PI가 전통적인 AI와는 크게 중복되지 않고 어느정도 질적으로 구별되는 개념이라는 것을 시사해주고 있다.

성격검사와의 관계분석

성격검사는 96년 하반기 공채시부터 임상적 검사가 신규도입되면서 본격적으로 채용에 적용되었다. 능력과 성격은 일반적으로 서로 변별되는 변수들이므로, PI검사는 물론 AI검사도 성격검사와는 상관이 낮을 것으로 기대되었다(표 5).

표 5. 96년 하반기 공채시 AI 및 PI검사와 성격검사간의 상관계수

	AI	PI
반사회성	-.09	-.12
편집증	-.07	-.07
경조증	-.09	-.08
정신분열증	-.08	-.09
사회적 책임(부족)	-.05	-.04
작업 곤란도	-.14	-.14
타당도 척도	-.12	-.14
정서안정(부족)	-.09	-.07
외향성	.08	.12
호감성	-.12	.04
성실성	-.03	.06
지적개방	.07	.01
허위성	-.05	-.01
도덕성	.00	.06
동기왜곡	-.03	.00

AI나 PI검사 모두 능력검사들이므로 **표 5**에 나타난 무시할만한 상관관계들은 AI와 PI검사가 성격측면이 아닌 능력측면을 목적한 바대로 측정하고 있음을 나타낸다. 특히 그동안 PI검사가 성격측면의 하나가 아닌가 하는 의견들도 제시되었었는데(Jensen, 1993), 이 상관지표들은 PI검사가 성격검사와는 구별되는 독특한 선발도구가 될 수 있음을 밝혀 준다. 이 표를 자세히 살펴보면, PI와 반사회성, 작업곤란도, 외향성 척도는 약하지만 의미있는 상관을 보이고 있어 PI가 높은 사람은 반사회적인 면이 덜하며 작업

곤란도를 적게 느끼고 외향적인 경향이 있다고 볼 수 있다. 이 결과는 PI가 자신, 타인들, 업무와 관련 묵시적 지식을 원활히 사용하는 능력인 점을 생각하면 기대할 수 있는 것이다. 한편, AI가 높은 응시자는 작업곤란도를 덜 느끼지만 호감성이 부족한 것으로 해석될 수 있다. 즉 이들은 타인에 대하여 관대하고 협조적이기 보다는 자기중심적인 특성을 보이는 경향이 있다고 할 수 있다.

면접과의 관계분석

면접은 AI, PI, 사회상식검사 및 성격검사를 모두 포함한 종합 적성검사와 TOEIC점수를 가지고 1차선발을 한 후 실시되어 최종합격을 결정하는 자료이다. 1차 합격자들 약 1000여명에 대한 AI와 PI검사가 면접점수와 가지는 상관을 구해본 결과, 3번의 공채채용 걸쳐 PI 및 AI검사 모두 .10미만의 주목할 정도가 아닌 관계를 보이고 있다. 물론 전체응시자들을 대상으로 분석할 때보다는 범위의 제한이 있어 어느 정도 상관계수가 낮아질 것으로 예측했지만, 예상보다도 더 낮은 관계를 나타냈다. 또한 성격검사들 중에서는 외향성이나 정신분열증 척도가 면접점수와 약하지만 의미있는 상관관계(각각 .13, .10)를 나타내고 있어 외향적인 응시자와 정신분열증을 보이지 않는 응시자일수록 면접점수를 높게 받는 경향이 있었다. 그 외의 척도들과는 거의 상관이 없었다.

이와 같은 상관관계로 미루어 보아 본 PI검사는 성격이나 면접특성과는 확실히 구별되는 선발변인으로 유용한 선발도구가 될 수 있음을 알 수 있었다.

학력과의 관계분석

다음 **표 6**은 학력별 PI검사 평균(표준편차) 및 PI검사와 학력간의 상관관계를 나타내고 있다.

표 6. 학력별 PI검사의 평균(표준편차) 및 PI검사와 학력간의 상관계수

	중학교	고등학교	전문대	대학교	대학원	상관계수
95 하반기	42.97 (16.11)	50.04 (13.03)	52.32 (13.03)	56.27 (11.12)	57.63 (10.67)	.28
96 하반기	66.24 (8.46)	69.80 (7.37)	72.34 (6.52)	74.85 (6.01)	73.60 (6.57)	.26

능력검사들이 학력과 상관이 있을 것은 기대되는 사실이지만, 예상대로 PI검사는 학력과 AI검사의 관련성(평균 .53) 보다 학력과 훨씬 약한 관계(평균 .27)를 보이고 있다. 즉, 학력이 올라갈수록 PI도 증가되고 있지만 고졸이상에서는 큰 차이가 보이지 않아, AI가 학력에 영향을 받는 정도에 비하면 학력차이에 의해 그 점수가 크게 좌우되지는 않는 편이다. AI검사들 중에는 학력과의 관련성이 언어력(.48), 수리력(.38), 추리력(.35), 정보검색력(.27), 공간지각력(.10)의 순으로 나타나, 언어력, 수리력, 추리력이 비교적 학력을 많이 반영하고 정보검색력이나 공간지각력은 학력의 영향을 덜 받는 것으로 보여진다. 이와같은 결과는 앞의 3가지 능력이 학교에서의 성공적인 수행을 위해 필수적인능력들을 대표하므로 예측가능한 것이다.

학교군, 전공, 성, 및 TOEIC과의 관계분석

학교 군을 서울·연·고대, 이화여대, 서울 기타대, 지방국립대, 지방기타대의 5군으로 나누었을 때 AI검사(특히 언어력)는 기대되는 바와 같이 유의미한 상관을 보였지만(.28), PI 검사는 5개 학교 군별(서울·연·고대, 이화여대, 서울기타대, 지방국립대, 지방기타대) 차이가 거의 없었다(.09). 즉 일류대학출신 응시자들의 경우 AI는 높은 경향이 있었지만 PI검사에서는 그 외의 대학출신 응시자들과 큰 차이를 보이지 않았고, 이 결과는 다시 한 번 AI와 PI가 비교적 독립적인 특성임을 나타낸다. 그외 전공별(경상계열, 인문사회계열, 어문계열, 공학계열, 자연계열) 및 성별 차이를 분석해 본 결과 AI나 PI검사 모두 이들 변인들과는 의미 있는 관련이 없음이 드러났다. 또한 PI검사는 외국어능력 검사인 TOEIC점수와의 관련성은 무시할만한 정도로 나타났다. TOEIC점수와 AI검사점수와의 상관은 .21로 아주 높은 편은 아니었으나 예상대로 언어력 검사와 상관은 .45로 높았다.

1차 합격과의 관계분석

마지막으로 PI검사와 1차 합격/불합격과의 관계를 분석해 보았다. 95년과 96년 하반기 공채때 1차 합격자와 불합격자의 AI 및 PI검사 평균과 표준편차

는 다음과 같다(표 7).

표 7. 95년, 96년 하반기 1차 합격자와 불합격자의 평균(표준편차) 비교

95년 하반기	1차 합격자 N = 2,844	불합격자 N = 19,635
AI (총점 300)	179.29 (19.12)	135.60 (26.01)
PI (총점 100)	62.66 (8.09)	52.88 (12.01)

96년 하반기	1차 합격자 N = 1,325	불합격자 N = 17,872
AI (총점 300)	196.02 (19.01)	148.18 (20.94)
PI (총점 100)	77.61 (4.84)	70.30 (6.58)

또한 1차 합격여부와의 상관관계를 5개 직군별로 분석실시한 결과 95년 하반기 공채시, AI검사와 마찬가지로 PI검사도 합격자와 불합격자를 잘 변별해 주고 있는데 PI검사와 1차 합격여부와의 상관은 .18(경영지원직)~ .36(기술연구직)으로 기술연구직에서 상관이 제일 높았다. 그외에도 3차례에 걸친 공채에서 PI검사와 전체적성검사 총점과의 상관관계가 .56~.67로 나타났고 역시 기술연구직에서 높은 상관이 도출되었다. 앞에서 언급되었듯이 이 직군에서는 다른 직군과 비교해 볼 때 PI검사가 AI검사와 상관도 가장 낮아, PI검사가 선발장면에서 전체 적성검사와 1차합격을 설명하는 부분이 크며 또 독자적으로 기여하는 부분이 있다는 증거를 제시하고 있는 점이 흥미롭다. 만일 앞으로의 채용시에도 PI검사가 기술연구직 선발에 높은 타당도를 계속 보인다면, 그 요인을 철저히 분석해 보고 아울러 이 직군에서는 PI검사점수의 비중을 높이는 방법도 신중히 검토되어야 할 것이다.

96년 하반기 공채시에는 1차합격과 AI검사와 .39의 관계성을 (언어력 .23, 수리력 .33, 추리력 .32, 공간지각력 .19, 정보검색력 .20), PI검사와는 .26의 상관계수를 나타냈다. 이 계수들을 그대로 보면 AI가 PI보다 1차 합격과 더 높은 관련성이 있는 것으로 결론 내릴 수 있지만, AI검사는 총 300점이고 PI검사는 총 100점인 것을 고려해보면 점수범위의 제한 혹은 분산정도가 영향을 미치고 있다는 것을 쉽게 알 수 있다.

논 의

최근 대기업을 중심으로 직무적성검사에 대한 관심과 수요가 급격히 늘고 있는 시점에서 수 십년전부터 논의되어왔던 '선발검사의 예측력'이란 주제가 새롭게 부각되고 있다. 그동안 전통적인 AI 검사에 대한 직무수행 예측력에 대해서는 McClelland로부터 Schmidt와 Hunter에 이르기까지 매우 다른 입장을 나타내는 연구자들이 많았지만 선발검사의 예측력향상은 모든 관련 연구자들의 공동목표가 되어 왔음이 틀림없다. 특히 인간의 정신능력에 관한 정의가 변화하고 전통적인 지능영역이 확장되고 있는 지금, '직무성공을 제대로 예측하기 위해서는 선발검사에서 무엇을 측정해야 하는가?' 라는 질문을 구체적으로 던진 McClelland(1973)의 진지한 제안들을 재조명해 보고 AI외의 잠재력 있는 선발변인을 밝혀내는 작업은 매우 의의가 있는 일이다.

본 연구는 McClelland의 애매모호하고 광범위한 역능·능력개념을 자신, 타인, 업무와 관련된 묵시적 지식이라는 구성개념으로 체계적으로 정리, 발전시킨 Sternberg와 그의 동료들의 연구결과들을 바탕으로 PI라는 잠재적 선발변인에 대해 개관하고, 선행연구에서 개발된 신입사원용 PI검사에 대해 국내 모기업을 대상으로 구성개념 타당화 작업을 수행해 보았다. 어떤 심리검사든지 본격적으로 활용되기 위해서는 다수의 해당집단을 대상으로 조사·분석하는 다단계 연구가 필요한데 선행연구는 예비조사를 거쳐 현직자 대상으로 집단차이 및 상사지도평가와의 상관관계 분석을 통해 구성개념타당도 및 준거관련(공인) 타당도를 탐색했었다. 본 연구에서는 위 검사를 실제 모기업 채용(95년 하반기부터 96년 하반기에 걸친 3차례 공채)에 적용하여 기존 선발변인들(AI검사, 성격검사, 면접, 학력, 학교군, TOEIC)과의 변별타당도를 중심으로 구성개념 타당도를 재확인하였다.

연구결과 전통적인 AI검사와는 5개 직군에 걸쳐 평균 .26의 약한 상관을 보여, AI검사와 중복되는 면이 크지 않은 반면, 검사내용이 전통적 지능검사에서 벗어난 AI검사와는 .31~ .40의 상관을 나타냈다. 이

러한 상관관계 분석결과는 Sternberg 및 그의 동료들의 연구결과와는 차이가 있는데 이는 앞에서도 논의되었듯이 AI검사내용의 차이뿐 아니라 조사대상자의 특성에서 큰 차이가 나기 때문인 것 같다. 즉, Sternberg와 동료들의 연구에 참가한 표본은 그 수가 매우 적어 제한된 능력범위의 집단을 대표했고 본 연구에서는 각 채용시마다 약 2만명 정도를 대상으로 검사를 실시했으므로 이전 연구와 비교할 때 상관계수는 증가할 수밖에 없다.

또한 특정문화에서 축적된 지식을 반영하는 결정지능을 측정하는 언어력 검사와는 .22~.37까지의 상관을 보이고, 유동지능을 측정한다고 가정되는 수리력, 추리력, 공간지각능력과 .08~.22까지의 상대적으로 낮은 상관관계를 보이고 있어, PI가 유동지능 보다는 결정지능과 더 많은 관련이 있다는 기존의 연구결과와 일치된다. 그외에도 본 PI검사는 이 그룹 종합적성검사 총점과는 직군별로 .56~.67의 높은 상관을 보이고 1차 합격여부와도 .18~.36의 상관을 나타내고 있어 전체검사점수 및 1차 합격에 관한 설명력이 만족할만하다고 할 수 있다.

반면에, 성격검사와의 상관은 기대된 바와 같이 무시할만했는데, 그 중 외향성, 반사회성, 작업곤란도 척도와는 -.14~.12의 상관을 보이고 있어, 약하지만 PI점수가 높은 사람일수록 외향적이고 사회성이 높으며 작업곤란도를 덜 느낀다고 할 수 있다. 또한 면접점수와는 .06의 낮은 상관을 보이고 있어 거의 관계가 없었다. 그 외에도 학력과는 .26의 상관을 보였는데, AI하위검사들(예: 언어력, 수리력, 추리력)에 비해서는 학력의 영향을 덜 받는 것으로 나타났다. 또 PI검사는 TOEIC점수와도 아주 낮은 상관을 나타냈고 출신학교(수준)의 등급, 전공계열별 및 성별로도 차이가 없었다.

위 결과들을 볼 때 PI검사가 종래의 사회적 지능 측정치와는 다르게 일반적으로 선발상황에 많이 사용되는 예측변인들(즉, AI검사, 성격검사, 면접)과는 구별되는 고유한 특성을 측정하고 있다고 볼 수 있어 새로운 선발변인인 PI와 AI 혹은 성격간의 상대적 독립성을 간접적으로 밝히는 중요한 경험적 증거를 제시하였다. 이는 선행연구 결과들과 함께 본 PI검사의 변별타당도를 지지할 뿐만 아니라 Sternberg와 그의 동료들의 PI연구결과들을 대체적으로 지지하고 있어 그 의미가 크며 앞으로 PI의 연구분야를 촉진시키는데 기여했다.

특히 미국의 경우 인사심리학에 관련된 논문가운데 선발검사와 관련된 것이 많은데, '한국심리학회지: 산업 및 조직'에 10년간 실린 논문들의 내용을 분석한 결과 국내에서 10년동안 「한국심리학회지: 산업 및 조직」 논문지에 발표된 심리검사의 활용연구는 단 5편에 그쳤고 그중 채용시 예측변인으로 사용됐던 검사결과에 대한 보고는 3편이었던 점을 감안해 보면(유태용, 탁진국, 1997) 본 연구에서 국내기업들을 대상으로 여러가지 선발변인들에 관한 분석을 실시했다는 점에서 본 연구에 의의가 있다.

그뿐 아니라 그동안 AI를 위주로 선발검사를 실시해 왔던 국내의 인사선발 분야에서 처음으로 실용적인 능력검사가 실제채용에 적용되어 선발도구로써 PI 검사의 사용가능성을 제시한 계기가 되었다. 결론적으로 본 PI검사는 실용적인 기업조직상황에 대처하는 능력이나 의사결정능력에 관한 측정을 포함하고 있어 점차 다양화되어가는 세계화 시대에 기업이 필요로 하는 능력과 적성을 가진 인재를 선발하는데 중요한 역할을 할 수 있다는 가능성을 보여주었다. 만약 위와 같은 모기업의 인사채용을 계기로 다른 회사들도 선발도구 안에 PI 검사를 포함시킨다면 우리나라와 같이 선발시험에 매우 민감한 반응을 보이는 문화에서 그동안 AI만을 강조하고 보상, 강화해 왔던 우리 가정과 학교도 이러한 능력을 소중히 여기고 보상하며 적극적으로 장려하게 될 것이 틀림없다. 그렇다면 응시자의 능력이나 잠재력에 관해 좀 더 종합적인 측정을 할 수 있을뿐 아니라 자연스럽게 학생들의 PI를 개발시키고 성장시키는 계기가 되어 일석이조의 효과를 얻게 된다.

그러나 본 PI 검사는 다음과 같은 제한점을 가지고 있다. 어느 검사든지 인사선발장면에서 채용도구로 자리를 확고히 하기 위해서는 (1) 검사가 바탕을 두고 있는 이론에 대한 확고한 지지와 (2) 경험적 타당성 자료들이 계속해서 모아져야 한다.

첫째, 이론적 측면에서 본 PI 검사는 PI를 포괄적으로 측정하고 있다기 보다는 묵시적 지식이라는 구성개념을 측정하고 있다는 제한점을 가지고 있다. 물론 지능과 지식은 서로 관련이 있을 수밖에 없고 묵시적 지식의 사용능력은 PI의 중요한 구성요소임에는 틀림이 없다. Schmidt와 Hunter는 직무지식이 묵시적 지식과 같은 개념이라고 주장했는데, 직무지

식중 절차적 직무지식이 업무관련 묵시적 지식에 포함되는 하위개념으로 생각할 수 있다. 즉, 직무지식이란 직무수행을 위해 알아야 될 직무에 대한 사실, 원칙, 방법 등에 관한 지식(Schmidt & Hunter, 1993)이므로 업무에 관련된 묵시적 지식을 반영한다. 따라서 이 묵시적 지식이 직무지식이라고 일축해 버리는 Schmidt와 Hunter의 주장은 지나치지만, 앞으로 PI가 유사변인들과는 구별되는 개념으로 자리잡으려면, 여러 직업에 걸쳐 적용할 수 있는 PI구성개념에 대해 좀 더 포괄적이고 체계적인 연구와 검증이 요망된다.

그외에도 본 PI검사는 묵시적 지식을 측정하는데 중점을 두고 있어, 응시자가 그 지식을 가지고 업무에서 실제로 행동할 지에 대해서는 확신할 수 없다. 따라서, 앞으로는 선발도구에 모의실험(simulation)과 같은 행동평가도 함께 포함시키는 것이 이상적이다. 하지만 모의실험에서 보인 행동을 실제 직무에서 수행하는지에 대한 보장도 역시 할 수 없는 것이 사실이다.

둘째, 타당성 자료수집에 관해서 본 연구에서는 변별타당도를 조사하기 위해 PI검사와 기존의 선발변인들과의 단순상관관계 분석이 실시되었다. 그러나 새로운 예측 혹은 선발변인이 다른 예측변인들과 구별되는 변별타당도를 지니고 있다고 완벽하게 주장하기 위해서는 다른 예측변인들과의 상관이 낮은 것 외에도 준거와의 관련성(상관)을 제시하고 또 동일한 준거를 대상으로 다른 예측변인들이 설명하는 분산이외에 추가적으로 준거를 설명하는 부분이 있다는 것을 보여야 한다. 이미 선행연구에서 본 PI검사가 현직자들의 직무수행과 관련이 있다는 것을 밝혔기 때문에, 후속연구에서(가능하다면 향후 2~3년 후) 상사평가나 실제업무성적과 같은 준거변인을 대상으로 다중회귀분석을 통해 AI검사, 성격검사, 면접 및 기타 선발변인에 대한 PI검사의 상대적 예측력을 검토하여 PI검사가 이들 기존의 선발변인들에 추가되어 유의한 설명력을 나타낸다는 점을 제시해야 할 것이다. 이에 따라서 앞으로 선발장면에서 PI검사의 필요성이 결정되며 상대적 가중치가 재조정될 수 있을 것이다. 각 선발검사의 상대적 예측력을 검토하여 그 비중을 적절히 반영한다면 전체적 선발시스템의 예측력을 향상시키는데 기여할 수 있을 것이 틀림없다.

또한 앞으로 본 검사유형의 PI검사가 채용현장에서 활발히 사용되기 위해서는 직무수행능력에 관한 준거변인(예: 인사고과, 상사평가, 부하직원수, 승진정도 등)뿐만 아니라 그밖의 직무관련 변인인 직무만족·직무스트레스 등과 PI와의 연관성을 연구하거나, 또 유사개념으로 알려진 정서적 지능, 직무지식 등과의 관계성(독립성)을 조사하는 연구도 요망된다. 왜냐하면 이들과 차별화되는 PI 고유성에 관한 연구결과에 PI검사의 중요성이 달려있기 때문이다. 또한, 향후 채용인력 대상으로 계속해서 교차타당도 분석을 하여 본 검사가 측정하고 있는 구성개념과 측정방법에 대해 검토해 보고 이와 병행하여 PI외의 새로운 선발변인의 개발에 관한 연구도 계속 진행되어야 하겠다. 이 때 새로운 변인의 선택기준은 '기존의 선발변인들과는 구별되는 비교적 독립적인 특성을 대표해야 하며 또한 중요한 준거들과 높은 상관을 보이는 특성들을 포함시키는 것'이 되어야 할 것이다.

참고문헌

김명소, 김명언, 이도형(1996). 산업장면에서의 실용적 지능/능력. **한국심리학회지: 산업 및 조직, 9**(1), 117-137.

김명소(1997). 인사선발의 개선방향. **호서대 사회과학연구소 산업인력개발분과 학회지**, pp. 5-39.

유태용, 탁진국(1997). 창간10주년을 맞는 '한국심리학회지: 산업 및 조직'의 내용분석 및 앞으로의 제언. **1997년 한국 산업 및 조직심리학회 동계학술대회**, pp. 107-134.

중앙일보(1997). 구직한랭전선...대졸자. **가을시름**, 10월 17일자.

Barrnet, G. V., & Depinet. R. L(1991). A Reconsideration for Testing for Competence rather than for Intelligence. *American Psychologist, 46*, 1012-1024.

Boyatiz, R. E.(1982). *The Competent Manager*, New York: Wiley.

Cantor, N. & Harlow, R.(1994). Social Intelligence and Personality : Flexible Life Task Pursuit. In R.J. Sternberg & P. Ruzgis(Eds.), *Personality and Intelligence.* Cambridge : University Press.

Cantor, N. & Kihlstrom, J. F.(1987). *Personality and Social Intelligence (1st ed.).* Englewood Cliffs, NJ: Prentice-Hall.

Cantor, N., & Kihlstrom, J. F.(1994). *Personality and Social Intelligence(2nd ed.).* Englewood Cliffs, NJ: Prentice-Hall.

Cronbach, L. J.(1960). *Essential of Psychological Testing.* New York: Happer & Row.

Day, D. V., & Silverman, S. B.(1989). Personality and Job Performance: Evidence of Increamental Validity. *Personnel Psychology, 42,* 25-26.

Gardner, H.(1983). *Frames of Mind.* New York: Basic Books.

Gardner, H.(1987). Developing the Spectrum of Human Intelligence. *Harvard Education Review, 57,* 187-193.

Horn, J. L.(1982). The Theory of Fluid and Cristallized Intelligence in Relation to Concepts of Cognitive Psychology and Adulthood. In F. I. M. Craik & A. Trehub(Eds.), *Aging and Cognitive Processes*(pp. 237-278), New York: Plenum.

Horn, J. L., & Cattell, R. B.(1966). Refinement and Test of Theory of Fluid and Cristallized Intelligence. *Journal of Educational Psychology, 57,* 253-270.

Hunter, J.E., & Hunter, R. F.(1984). Validity and Utility of Alternative Predictors of Job Performance. *Psychological Bulletin, 96,* 72-98.

Hunter, J. E.(1986). Cognitive Ability, Cognitive Aptitudes, Job Knowledge, and Job Performance. *Journal of Vocational Behavior, 29,* 340-362.

Jensen, A. R. (1993). Test Validity: g versus "tacit knowledge." *Current Directions of Psychological Science, 2,* 9-10.

Klemp, G. O., Jr., & McClelland, D. C.(1986). What Characterizes Intelligent Functioning among Senior Managers? In R. J. Sternberg & R. K. Wagner (Eds.), Practical Intelligence: Nature and Origins of Competence in the Everyday World (pp. 31-50). Cambridge, England: Cambridge University Press.

Matney, D. C., & MacMains, D.(1977). Repression Sensitization Status and Social Intelligence Characteristics. *Psychological Report, 41,* 837-838.

McClleland, D. C.(1973). Testing for Competence rather than for "Intelligence." *American Psychologist, 28,* 1-14.

McClleland, D. C.(1993). Intelligence is not the Best Predictor of Job Performance. *Current Directions in Psychological Science, 2,* 5-6.

Neisser, U.(1976). General, Academic, and Artificial Intelligence. In L. Resnick(Ed.), *The Nature of Intelligence* (pp. 135-144). Hillsdale, NY: Erlbaum.

Osipow, S., & Walsh, W. B.(1973). Social Intelligence and the Selection of Counselors. *Journal of Counseling Psychology, 4,* 866-869.

Polanyi. M.(1976). Tacit Knowledge. In M. Marx & F. Goodson (Eds.), *Theories in Contemporary Psychology* (pp. 330-344). New York: Macmillan.

Ree, M. J. & Earls, J. A.(1993). g is to Psychology What Carbon is to Chemistry: A Reply to Sternberg and Wagner, McCelland, and Calfee. *Current Directions in Psychological Science, 2,* 11-12.

Schaie, K. W.(1978). Toward a Stage Theory of Adut Cognitive Development. *International Journal of Aging and Human Development,* 8, 129-138.

Schmidt, F. L., & Hunter J. E.(1981). Employment Testing: Old Theories and New Research Findings. *American Psychologist, 36.* 1128-1137.

Schmidt, F. L., & Hunter J. E.(1992). Development of a Causal Model of Processes Determinig Job Performance. *Current Directions in Psychological Science, 1*(3), 89-92.

Schmidt, F. L., & Hunter J. E.(1993). Tacit Knowledge, Practical Intelligence, General Mental Ability, and Job Knowledge. *Current Directions in Psychological Science, 2*, 8-9.

Sternberg, R. J.(1985). *Beyond IQ : A Triarchic Theory of Human Intelligence*. Cambrige : Cambridge University Press.

Sternberg, R. J.(1995). Theory and Measurement of Tacit Knowledge as a Part of Practical Intelligence. *Zeitschrift für Psychologie, 203*, 319-334.

Sternberg, R. J.(1996). *Successful Intelligence*. New York: Simon & Schuster.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M.(1981). People's Conceptions of Intelligence. *Journal of Personality and Social Psychology : Attitude and Social Cognition. 41*, 37-55.

Sternberg, R. J., & Smith, C.(1985). Social Intelligence and Decoding Skills in Nonverbal Communication. *Social Cognition*, 2, 1168-1192.

Sternberg, R. J. & Wagner R. K.(1993). The g-ocentric View of Intelligence and Job Performance is Wrong. *Current Directions in Psychological Science, 2*, 1-5.

Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath. J.A.(1995). Testing Common Sense. *American Psychologist, 50*(11), 912-927.

Wagner, R. K. & Sternberg, R. J.(1985). Practical Intelligence in Real-world Pursuits: The Role of Tacit Knowledge. *Journal of Personality and Social Psychology, 49*, 436-458.

Wagner, R. K. & Sternberg. R. J.(1990). Street Smarts. In K. Clark & P. Clark (Eds.), *Measures of Leadership*(pp. 493-504). Greensboro, NC: Center for Creative Leadership.

Wagner, R. K. & Sternberg, R. J.(1993). *Tacit Knowledge Inventory for Manager*. The Psychological Corporation.

Wigdor, A. K. & Garner, W. R. (Eds.) (1982). *Ability Testing : Uses, Consequences, and Controversies*. Washington, DC : National Academy Press.

Williams, S. A., Denny, N. W., & Schadler, M. (1993). Elderly Adults' Perception of Their Own Cognitive Development during the Adult Years. *International Journal of Aging and Human Development, 16*, 147-158.

1차 원고 접수 : 1998년 3월 12일
최종 수정본 접수: 1998년 5월 4일

A Study on the Construct Validation of Practical Intelligence Inventory for Entry-level Job Applicants

Myoung-So Kim

Hoseo University

Researchers have begun to explore new selection variables in search of measures to complement a long industry tradition of IQ-type intelligence or ability tests as predictors of job performance. Among the most promising variables is practical intelligence(PI), the construct supposed and measured by Sternberg. A PI inventory for entry-level job applicants was developed and known to have predictive validity with the criterion measure of supervisor's ratings in Kim, Kim, Lee (1996)'s study. The purpose of the present study was to examine the construct (particularly discriminant) validity of this inventory in a large corporation. More specifically, the relationships between the PI inventory and other common selection measures such as general or academic intelligence(AI) test, personality test, interview, TOEIC, education level, etc. to examine whether the PI inventory has the potential to make a unique contribution to the field of selection testing. The inventory was which comprised of 25 job-relevant situations or scenarios with 4 or 5 response alternatives was administered to approximately 20,000 entry-level job applicants each time in May, 1995, December, 1995, and May, 1996. It was found that the PI inventory showed weak correlations with AI inventory ranging from .22 to .29 across five job groups. It also correlated trivially with personality test(about -.14~ .14), interview(about .06), and TOEIC(about .09). In addition, the correlation between the PI test and education level was .26 which was much smaller than that between AI inventory and education level. Consequently, PI is not a proxy for measures of AI, personality or interview, and the test of PI could increase the variance not accounted for in the real-world criterion of job performance by existing selection measures. Lastly, the implications of these findings and limitation of this study were discussed. Also additional research tasks on PI were suggested.