

## 성격검사문항에 대한 이상점 반응 모형의 적용\*

이 선 희†

충남대학교 심리학과

본 연구는 문항반응과정에 대한 이상점 반응(ideal point response) 모형과 우세과정(dominance process) 모형을 5점 척도로 측정된 성실성 척도에 적용해 봄으로써 다양한 성격 측정 도구들의 문항개발과 채점방식이 암묵적으로 가정하고 있는 문항반응과정에 대한 근본적인 가정을 검증하고자 하였다. 국제 성격문항집(International Personality Item Pool)의 문항들을 기초로 본 연구에서 개발한 성실성 척도에 대한 응답 자료를 문항반응이론 방법으로 분석한 결과, 대체로 모든 문항에서 우세과정 모형과 이상점 반응 모형의 모형적합도가 유사하게 좋게 나타났지만, 우세과정 모형보다는 이상점 반응모형을 지지하는 옵션반응곡선을 나타내는 문항들도 관찰되었다. 또한 이상점 과정 모형과 우세 과정 모형을 각각 채점에 적용하여 산출된 점수와 학업성취관련 준거변인들과의 상관관계의 정도를 비교한 결과 두 모형에서 계산된 성실성 점수가 거의 유사한 수준의 예측타당도를 보였다. 이러한 결과가 성격 문항의 개발과 채점, 그리고 성격점수의 이용 목적과 관련하여 어떠한 함의를 가지는 지 논의하였다.

주요어 : 문항반응과정, 이상점 반응 모형, 우세과정 모형, 성격측정, 성실성

\* 이 논문은 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (KRF-2006-H00016)

\* 문항번역을 도와준 오인수, 논문항상을 위해 유익한 조언을 해준 Wen-Hung Chen, 김청택, 반재천 선생님께 감사드립니다.

† 교신저자: 이선희, 충남대학교 심리학과

Tel: 042-821-6370, E-mail: sunhee\_lee@cnu.ac.kr

다양한 심리학 분야의 연구자들은 성격을 측정하여 개인의 행동을 이해하고, 진단하고, 예측하려는 노력을 계속해 왔다. 특히, 5요인 성격모형의 등장은 성격 연구자들에게 성격의 구조에 대한 공유된 틀을 제공해 주었고(Costa & McCrae, 1992; Goldberg, 1990), 이러한 발전은 최근 성격에 대한 연구들이 보다 활발하게 진행되는 데 상당한 기여를 하고 있다. 예를 들어, 성격 5요인은 적응 과정(Connor-Smith & Flachsbart, 2007)과 정신병리(Derefinko & Lynam, 2007), 흡연행동(Hampson, Andrews, Barckley, Lichtenstein, & Lee, 2000; Terracciano & Costa, 2004), 직무수행도(Barrick & Mount, 1991; Hurtz & Donovan, 2000; Salgado, 1997) 등을 예측하고 이해하는 데 유용한 변인으로 인정받고 있다.

그런데 심리학 연구 및 진단과 평가에 이용되는 대부분의 성격변인들은 자기 혹은 타인 보고를 바탕으로 한 질문지식 도구를 이용하여 측정된다. 따라서 신뢰롭고 타당한 성격검사의 개발은 성격의 이론적 연구나 실제적 적용에서 매우 중요한 기초 작업이라고 할 수 있다. 일반적으로 심리검사를 개발하는 과정을 살펴보면, 우선 측정하고자 하는 구성개념에 대한 명확한 정의를 바탕으로 이를 반영하는 다수의 문항을 개발하고 그 중에서 측정하고자 하는 구성개념을 가장 적절하게 짚 수 있는 문항을 선별하여 측정도구를 구성한다(Reckase, 1996). 이러한 척도 구성 작업에는 전통적으로 고전검사이론(classical test theory)을 바탕으로 한 요인분석 방법이 주로 많이 이용되어 왔다(Ford, MacCallum, & Tait, 1986). 그러나 최근 들어 고전검사이론에 대한 대안으로 문항반응 이론(item response theory: IRT)을 성격 척도의 개발 및 채점에 적용하려는 시도들이 있어왔다. 문항반응 이론은 문항에 대한 응답

과 측정하고자 하는 심리적 구인 간의 관계를 기술하는 일단의 통계적 모형들을 말하는데, 이러한 모형들은 전체 측정도구의 특성뿐만 아니라 개별 문항의 특성에 대한 정보를 준다는 점에서 고전검사이론과 차별성을 가진다. Reise와 Waller(1990)가 성격 측정도구에 이모수 문항반응 모형(two parameter IRT model)<sup>1)</sup>을 적용한 이래, Reise와 Henson(2000)은 문항반응 이론을 이용하여 5요인 성격모형을 컴퓨터 적응 검사(computerized adaptive test)로 개발하였고, Brown과 Harvey(2003)는 문항반응 이론을 이용하여 거짓응답(faking)을 찾아내기 위한 기법을 개발하는 등 다양한 시도가 계속되고 있다.

그런데 성격 문항에 대한 기존의 분석방법들이 ‘우세 과정(dominance process)’이라는 특정한 문항 반응 방식을 가정하고 있다는 지적이 최근에 제기되었다(Stark, Chernyshenko, Drasgow, & Williams, 2006). Likert(1932) 측정이론의 기본 가정인 우세 과정 가정은 각 개인들과 문항들은 한 특질의 연속선상의 특정한 위치를 가지며, 한 개인이 그 특질을 더 많이 가지고 있을수록 그 문항에 동의할 확률이 커진다고 본다. 예를 들어, ‘나는 친구들과 만나 이야기를 나누는 것을 좋아한다.’라는 문항이 외향성을 측정하는 문항이라면, 외향성이 더 높은 사람일수록 이 문항에 그렇다고 동의할 확률이 높아진다는 것이다.

이에 반해 Thurstone(1928)은 ‘이상점 반응 과정(ideal point response process)’ 가정을 주장하였다. 이 가정은 원래 태도 측정의 맥락에서 소개된 것으로, 우세 과정 가정과 마찬가지로 각 개인들과 문항들은 한 특질의 연속선상에

1) 이모수 로지스틱 모형의 기울기 모수는 문항과 구인과의 관계의 정도를 반영하고 역치 모수는 그 문항에 동의하기가 얼마나 쉬운가를 나타낸다.

서 특정한 위치를 가진다고 본다. 그러나 우세 과정 가정과는 달리, 이상점 반응 가정은 사람들은 특질상에서의 자신의 위치와 문항의 위치가 유사할수록 그 문항에 동의할 확률이 높아지며, 반대로 자신의 위치와 문항의 위치가 멀리 떨어져 있을수록 그 문항에 동의할 확률이 낮아진다고 주장한다. 위에서 예를 든 문항은 외향성이라는 특질 상에서 중상(中上) 정도의 외향성을 나타내는 문항이라고 가정해 보자. 이상점 반응 과정 가정에 따르면, 이 문항에 가장 동의할 확률이 높은 사람은 외향성 상에서의 이 문항의 위치인 중상 정도의 외향성을 가진 사람들이다. 특정 개인의 외향성의 정도가 이 위치에서 멀어질수록 이 문항에 동의할 가능성이 낮아지게 되는데, 이는 양방향에서 가능하다. 즉, 중상보다 낮은 정도의 외향성을 가진 사람들은 이 문항이 나타내는 위치보다 덜 외향적이기 때문에 이 문항에 동의하지 않을 것이며, 반대로 이 문항보다 더 높은 외향 성향을 가진 사람 역시 자신의 위치와 문항의 위치가 일치하지 않기 때문에 동의할 가능성이 떨어질 수 있다. 바로 후자의 경우에서 이상점 과정 모형과 우세 과정 모형의 차이가 발생하게 된다. 즉, 우세 과정 모형에서는 외향성이 매우 높은 사람도 이 문항에 동의할 가능성이 높다고 보는 반면, 이상점 과정 모형은 그렇지 않을 것이라고 예상하는 것이다.

사람들이 문항에 응답할 때 어떠한 과정을 따르는가에 대한 이러한 가정은 어떤 문항을 좋은 문항으로 선택하여 최종 검사문항으로 포함할지, 어떻게 개개인의 점수를 산출할지, 그리고 더 나아가 그 검사에서 측정된 특질과 다른 변인과의 관계를 추론하고 이를 이용해 미래의 행동을 예측하는 데 중요한 영향을 미

칠 수 있다. 따라서 본 연구는 문항반응이론 모형을 이용하여 성격 측정 문항에서 이상점 반응 과정 가정과 우세 과정 가정 중 어떠한 가정이 응답 자료를 설명하는데 더 적절한지를 모형 적합도를 바탕으로 비교해 보고, 더 나아가 서로 다른 이 두 가지 가정을 바탕으로 산출된 성격점수를 이용하여 사람들의 행동을 예측할 때 어떤 방식이 더 예측력이 좋은지를 비교해 보고자 하였다.

#### 우세과정 이론과 이상점 반응 과정 가정

심리검사 문항의 선별, 평가, 채점에 일반적으로 사용되는 모형에는 고전검사이론, 요인분석 모형, 로지스틱 문항반응모형(Logistic Item Response Model) 등이 있다. 이들은 공통적으로 특정 속성에서 높은 위치를 가지고 있을수록, 그 특질을 재는 문항에 동의할 가능성이 높아지며 따라서 더 높은 측정점수를 얻게 될 것이라는 우세과정을 가정한다.

이러한 반응 양식은 일반적인 로지스틱 문항반응모형으로 모형화할 수 있으며, 그림 1의 a와 b와 같은 문항 특성 곡선(Item Characteristic Curve)으로 표현될 수 있다. 문항 특성 곡선은  $\theta$ 라고 흔히 표기되는 특질과 문항에 동의할 확률의 관계를 나타낸 곡선이다. 예를 들어, 특정 문항이 재고자 하는 특질( $\theta$ )이 외향성이라면, 그림 1의 a와 b에서 보는 바와 같이 외향성에서 높은 위치를 가진 사람일수록, 즉, 보다 외향적일수록, 그 문항에 동의할 확률이 증가하게 되는 것이다. 그림 1에서 a와 b 문항의 차이는 특질상에서의 문항의 위치가 어디에 있는가에 있다. 문항 a는 중간정도에 위치한 반면, 문항 b는 외향성의 상단에 위치하고 있다. 이 두 문항은 비록

문항의 위치는 다르더라도 개인의 외향성이 증가함에 따라 문항에 동의할 확률 또한 증가한다는 점에서 유사한 양상을 보이고 있다.

이에 반해 이상점 반응 과정의 문항반응 양식은 기존의 로지스틱 문항반응 모형이 아닌, 언폴딩 문항반응모형(Unfolding Item Response Model)으로 가장 적합하게 모형화할 수 있다 (Roberts, Donoghue, & Laughlin, 1999). 이상점 반응 과정의 문항반응 양식의 문항특성곡선은 그림 2의 a와 b처럼 표현될 수 있다. 그림 2의

문항 a는 그림 1의 문항 a와 마찬가지로 중간 정도의 외향성을 나타내는 문항으로 이상점 반응 과정 모형에서는 중간 정도의 외향성을 지닌 사람이 이 문항에 동의할 가능성이 높은 반면 문항의 위치보다 외향성이 낮거나 혹은 높을수록 동의할 가능성은 감소하는 것으로 가정한다. 특히, 외향성이 일정 수준 이상으로 높은 경우 이 문항에 동의할 가능성이 오히려 감소한다는 점에 주목해야 할 것이다. 그림 2의 문항 b는 높은 외향성을 나타내는 문항으로

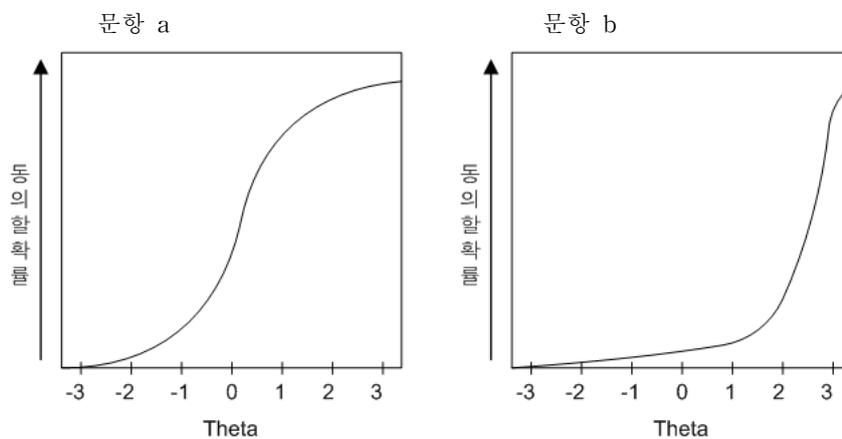


그림 1. 우세과정 모형

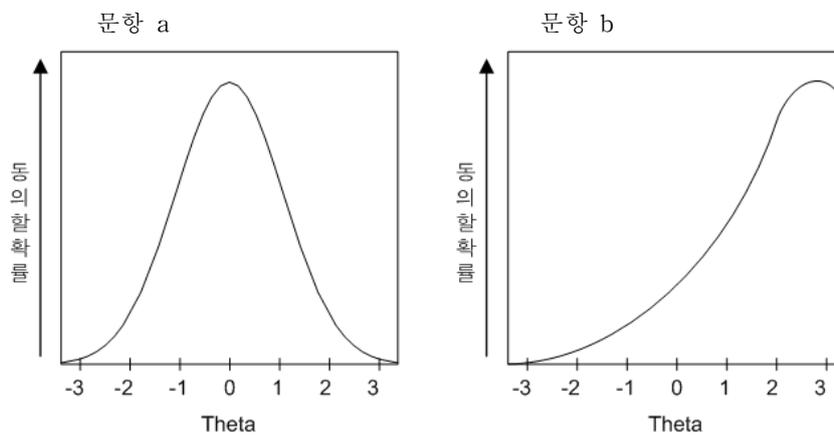


그림 2. 이상점 과정 모형

로 그림 1의 문항 b와 매우 유사한 문항반응 특성을 보이다가 매우 극단적인 외향성에 도달해서야 동의할 확률이 감소하는 양상을 보인다.

이와 같이 우세과정 가정에서는 개인의 외향성이 증가할수록 외향성을 재는 문항에 동의할 확률이 단조적(monotonic)으로 증가하는 반면, 이상점 반응 과정 모형의 문항특성곡선은 단조적으로 증가하다가 어느 순간에 이르러서는 다시 감소한다는 점에서 우세 과정 모형과 대조된다. 이처럼 응답확률이 증가하다가 감소하는 현상을 접힘(folding)이라고 부르는 데, 이러한 접힘 현상은 이상점 모형이 보다 적절하다는 것을 시각적으로 보여주는 결과이다.

그런데 문항의 위치에 따라 이러한 두 가정간의 차이가 더욱 두드러질 수도 있고 그렇지 않을 수도 있다는 점에 주의해야 할 것이다. 즉, 중립적인 문항에서는 두 모형이 가정하는 문항반응 양상이 매우 다르지만(문항 1a와 2a), 극단적인 문항에서는 이 두 모형이 가정하는 반응양상이 별로 다르지 않을 수 있다(문항 1b와 2b). 특히 극단적인 문항에서는 중간 정도의 성향을 가진 사람들은 어느 모형을 적용하던 그 타당도에 별 문제가 없는 반면, 그 특질에서 극단적인 위치에 있는 사람들일수록 그 사람들의 특질에서의 위치와 문항에 대한 반응의 관계가 왜곡되게 나타날 수 있다는 점에도 주목해야 한다.

사람들이 과연 어떤 방식에 따라 문항에 응답하는가에 대한 가정은 어떤 특질을 측정하느냐에 따라 그 적절도가 결정될 수 있다(Roberts et al., 1999; Stark et al., 2006). 예를 들어, 수학능력 검사의 경우 수학능력이 뛰어난 수록 수학능력 검사 문항에 정답을 맞출 가능

성이 단조적으로 증가한다는 가정은 매우 타당하다고 판단된다. 이는 중간 정도의 난이도를 가진 문항에서 수학실력이 매우 뛰어난 사람이 정답을 맞힐 확률이 오히려 떨어진다 것은 상식적으로 가능성이 없기 때문이다. 따라서 이러한 검사 문항의 제작, 채점, 평가에는 우세과정 모형을 이용하는 것이 적합할 것이며, 실제로 많은 경험적인 자료들에서 기존의 로지스틱 문항반응이론 모형이 능력검사 문항 응답 자료에 매우 적합한 것으로 나타났다.

그러나 최근 들어 기존의 우세과정 가정을 바탕으로 하는 로지스틱 문항반응이론 모형이 태도나 성격을 재는 문항에 있어서는 적합하지 않을 수 있다는 인식을 바탕으로 과연 어떠한 문항반응과정 가정이 태도나 성격 측정에 더 적합한지에 대한 경험적인 검증의 필요성이 제기되었다(Chernyshenko, Stark, & Chan, Drasgow, & Williams, 2001; Roberts et al., 1999). 언폴딩 문항반응모형을 이용한 최근의 몇몇 연구들은 실제로 특정한 태도나 성격에 관한 문항에 대한 반응결과가 기존의 로지스틱 문항반응이론을 응용한 모형보다 언폴딩 문항반응모형으로 분석하는 것에 더 적합하다는 결과를 보여주고 있다(Roberts et al., 1999; Stark et al., 2006).

구체적으로 살펴보면, Roberts 등(1999)은 낙태에 대한 태도를 재는 문항들에 대한 781명의 응답결과를 분석한 결과 특히 중립적인 문항들(예, “낙태는 기본적으로 비도덕적이지만 여성의 건강이 위협받을 때는 허용해야 한다.”)에서 문항특성곡선이 중간에 꺾이는 현상이 확연히 나타남을 보여주었다. Stark 등(2006)은 이상점 모형과 우세과정 모형을 16요인 성격검사에 대한 응답에 적용한 결과 이상점 모

형이 우세과정 모형보다 더 적합하거나 우세과정 모형과 비슷한 정도의 적합도를 보이는 것을 확인하였다.

이와 같이 이상점 모형을 성격검사에 적용하려는 시도는 매우 최근의 경향이라서 다양한 성격검사에 이를 적용하여 과연 이상점 모형이 기존의 우세 모형보다 유용한지, 그리고 어떤 점에서 유용한지에 대한 지속적인 검토가 필요한 시점이다. 특히 성격문항에 이상점 모형을 적용한 기존의 연구(Stark et al., 2006)는 이분화된 응답 자료를 사용한 반면, 본 연구에서는 성격 5요인 중의 하나인 성실성(conscientiousness)을 측정하는 측정문항들에 대한 5점 척도형의 응답 자료를 기존의 로지스틱 문항반응모형과 언폴딩 문항반응모형으로 분석하여 어느 모형이 성격 문항에 대한 응답에 더 적절한지를 검증해 보고자 하였다.

성격적도 문항에 어떤 문항반응모형을 적용하는 것이 적절한가에 대한 다양한 기준이 있을 수 있으나, 그 중 중요한 하나의 방법은 특정 응답자료에 어떤 모형이 더 나은 통계적 적합도를 지니는가를 비교하는 것이라고 할 수 있다. 따라서 본 연구에서는 두 모형간의 상대적인 모형 적합도(model fit)를 비교하기 위해 적합도 곡선(fit plot)과  $\chi^2/df$  비율 통계치를 비교하였다.

문항반응모형의 상대적 적절성을 평가할 수 있는 또 하나의 기준은 관련 변인들에 대한 예측타당도를 비교하는 것이다. 기존의 연구들은 어떤 문항반응모형을 이용하여 점수를 산출했는가에 따라 그 특질에서 개인들의 순서가 뒤바뀌는 결과를 낳을 수 있음을 보여주었다(Roberts et al, 1999; Stark et al, 2006). 즉, 특정 검사 문항들에 대한 동일한 응답을 각각의 모형을 바탕으로 개개인의 점수를 산출했

을 때, 우세과정을 가정한 로지스틱 문항반응 모형에서 나온 점수들에서의 사람들의 순서와 이상점과정을 가정한 언폴딩 문항반응 모형에서 나온 점수들에서의 사람들의 순서가 다르게 나타났다. 특히 이러한 현상은 점수가 낮거나 높은 사람들에게서 더 많이 나타났다. 이러한 경우 만약 잘못된 모형을 이용한 점수 순서를 바탕으로 인사선발 등의 의사결정을 하게 된다면 성격검사의 유용성(utility)을 격감시키는 결과를 초래할 수 있다. 따라서 본 연구에서는 기존의 연구들에서 한 걸음 더 나아가 우세과정 모형과 이상점 모형에서 각각 산출된 성실성 점수들의 예측타당도를 비교해 봄으로써 어느 모형이 관련된 행동을 예측하는 데 더 유용하게 이용될 수 있는지를 비교해 보았다.

성실성은 여러 중요한 개인의 행동을 예측하는 것으로 알려져 있다. 그 중에서도 특히 성실성이 학업 성적을 의미 있게 예측하며, 이러한 학업 성적에 대한 성실성의 효과는 학업성적에 대한 자기효능감에 의해 매개된다는 연구결과들이 보고되었다(Blickle, 1996; Colquitt & Simmering, 1998; Lee & Klein, 2002). 따라서 본 연구에서는 이상점 모형과 우세과정 모형을 이용해 성실성 점수를 각각 산출하여 어느 성실성 점수가 자기효능감과 학업 성적을 더 잘 예측하는지를 비교하였다.

## 방 법

### 참가자 및 자료수집 방법

본 연구의 참가자는 한 대학교의 심리학 관련 교양 과목 수강생 681명이었다. 이중 남학

생이 359명(52.7%)이었으며, 여학생이 322명(47.28%)이었다. 학년별로는 1학년 26.1%, 2학년 37.0%, 3학년 15.3%, 4학년 21.2%였다.

연구참가자를 모집하기 위해 연구자가 개강 첫 주에 해당 과목의 총 15개 반(총수강생 수=1,135명)에 들어가 본 연구를 소개하고 참가희망자에게 연구참여 동의서를 받았다. 이 중 동일한 교수가 담당하는 3개 반(전체 수강생 수=196명)에게는 성격설문에 추가하여 타당도 자료를 수집하기 위해 2차례의 추가 설문문에 대한 참여와 해당 수업에서의 중간시험과 기말시험의 점수를 연구에 이용하는 것에 대한 동의서를 받았다. 나머지 12개 반에서는 성격설문에 대한 참여희망만을 물었다. 본 연구의 참여자들에게는 소정의 보너스 점수가 제공되었으며, 비참여자들은 보고서 작성 등의 다른 방법으로 동일한 정도의 보너스 점수를 획득할 수 있었다. 최종적으로 681명(전체 수강생의 60%)이 연구에 참여하였으며, 타당도 자료를 수집한 3개 반의 196명 중 성격설문과 1차 설문문에 참여한 참가자는 127명(수강생의 65%), 2차설문까지 참여한 참가자는 119명(수강생의 61%), 세 가지 설문 모두와 성적자료까지 제공한 참가자는 117명(수강생의 60%)이었다.

모든 연구참가자들은 웹설문을 통해 개인적으로 성격문항에 응답하였다. 타당도 연구에 참여한 참여자들에게는 학기 초에 연구자가 수업시간에 방문하여 중간시험 점수에 대한 자기효능감(1차)에 관한 지필식 설문을 집단적으로 실시하였으며, 중간시험 이후 중간시험 점수가 학생들에게 고지된 후 동일한 방식으로 기말고사에 대한 자기효능감에 대한 설문(2차)을 집단적으로 실시하였다. 학기종료 후 담당교수로부터 중간시험과 기말시험의 점수

를 수집하였다.

## 측정변인

### 성실성

성격 5요인 모형의 한 요인인 성실성을 측정하는 측정도구를 구성하기 위해 국제 성격 문항집(International Personality Item Pool, 이후부터는 IPIP라고 칭함, 2001)의 문항을 이용하였다. IPIP는 다양한 성격 특질을 측정할 수 있는 2,000개 이상의 문항들을 연구자들이 무료로 이용할 수 있도록 웹사이트에 공개된 문항은행이다. Goldberg 등(2006)에 의하면, IPIP의 문항들은 25개의 언어로 번역되었으며, 이 문항을 이용한 논문이 80여 개 이상 발표될 만큼 많은 성격연구자들의 관심을 받고 있다.

본 연구에서는 IPIP 문항 중에서 기존에 사용되고 있는 다른 5요인 성격 검사도구인 NEO\_PI\_R(Costa & McCrae, 1992)의 성실성 요인 정의를 측정할 수 있는 문항이라고 기존 연구자들에 의해 제시된 60개의 문항을 우리나라 말로 번역하였다. NEO\_PI\_R의 성실성 요인은 자기효능감(self-efficacy), 질서성(orderliness), 의무감(dutifulness), 성취지향성(achievement-striving), 자기통제성(self-discipline), 조심성(cautiousness) 등의 6개의 하위 요인으로 구성되어 있다. 문항의 내용타당도를 확보하기 위해 5요인 성격이론에 대해 잘 알고 있는 5명의 심리학 박사 및 대학원생에게 부탁하여 번역된 문항들을 성격 5요인으로 재분류하고 문항의 의미가 모호한 문항들을 표시하도록 하였다. 이러한 과정을 통해 1) 성실성이 아닌 다른 요인으로 분류되는 문항, 2) 한국어로 번역했을 때 다른 문항과 의미가 매우 유사한 문항, 3) 의미가 명확하지 않은 문항, 4) 평가

자간의 불일치가 높은 문항 등을 제외하고, 각 하위 요인에 해당되는 문항 수를 공평하게 배분할 수 있도록 고려하여 최종적으로 30개의 문항을 선정하였다.

응답자들은 행동이나 성향을 묘사하는 각 문장에 대해 동의 정도를 5점 척도('1=전혀 그렇지 않다'에서 '5=정말 그렇다') 중 하나를 선택하였다.

### 자기효능감

각 참여자들은 해당 과목에 대해 6개의 시험점수(95(A+), 90(A), 85(B+), 80(B), 75(C+), 70(C)) 각각에 대해, 적어도 해당점수를 맞을 수 있다고 생각하는 정도를 0('전혀 가능성이 없다')에서 100('100% 확신한다')까지의 숫자 중 하나로 표시하도록 하였다. 여섯 개 점수에 대한 평가들의 합 점수(최대 가능 점수=600, 최소 가능 점수=0)를 해당 과목의 학업 성적에 대한 자기효능감으로 이용하였다. 이러한 자기효능감 측정방법은 Locke, Frederick, Lee, 그리고 Bobko(1984)가 이용한 절차를 따른 것이다.

### 학업성적

해당 심리학 교양 수업에서의 중간시험과 기말시험 성적의 평균을 학업성적으로 이용하였다. 학업성적은 100점 만점으로 계산하였다.

### 분석

단일척도 문항반응이론은 문항에 대한 반응이 하나의 연속적인 잠재변인의 함수로 결정된다고 가정한다. 따라서 문항반응이론을 적용하기 위해서는 우선 척도의 단일 차원성

(unidimensionality)을 확립하는 단계가 선행되어야 한다. 이를 위해 본 연구에서는 SAS 9.0을 이용하여 탐색적 요인분석을 실시하였다.

일단 척도의 단일 차원성을 확인한 후, 우세과정을 가정하는 로지스틱 문항반응모형과 이상점 과정을 가정하는 언폴딩 문항반응모형을 자료에 적용하였다. 본 연구에서 사용된 문항들이 다중 반응 범주형(multiple response category) 문항이기 때문에 로지스틱 문항반응모형 중에서도 일반화 부분 점수 모형(Generalized Partial Credit Model, Muraki, 1992)을 이용하였으며, 이 모형의 문항모수치(item parameter)를 추정하기 위한 컴퓨터 프로그램으로는 MULTILOG(Thissen, 1991)을 사용하였다. MULTILOG는 다중 반응 범주형 문항 자료에 대한 다양한 모형들의 모수치를 추정해주는 프로그램으로 널리 사용되고 있다.

언폴딩 문항반응모형으로는 언폴딩 문항반응모형 중에서도 가장 일반적인 모형으로 평가받고 있는 Roberts와 그의 동료들(Roberts & Laughlin, 1996; Roberts et al., 1999)이 개발한 일반화된 그레이드 언폴딩 모형(Generalized Graded Unfolding Model, GGUM)과 문항모수치를 추정하기 위한 통계프로그램으로는 GGUM 2004를 이용하였다.

성실성 척도에 대한 응답자료에 어느 문항반응 모형이 더 적합한지를 평가하기 위해 적합도 곡선과  $\chi^2/df$  비율 통계치를 사용하였는데 이들은 MODFIT(Stark, 2001)를 이용하여 계산하였다. 또한 로지스틱 문항반응모형과 언폴딩 문항반응 모형 각각을 기초로 산출한 성실성 점수가 응답자들의 자기효능감과 학업성적을 예측하는 타당도를 비교하기 위해 각각의 상관계수를 비교하였다.

## 결 과

30개의 성실성 문항의 단일 차원성을 점검하기 위해 최대우도방법을 사용하여 탐색적 요인분석을 실시하여 스크리 플랏(scree plot)을 살펴본 결과, 제 1요인과 제 2요인에서 급격한 변화가 관찰되었고, 그 이후 제 5요인까지도 꺾임이 관찰되었다. 이와 같이 다요인구조의 가능성을 보여주는 결과는 성실성 척도를 NEO\_PI\_R(Costa & McCrae, 1992)의 하부요인들을 고려하여 구성한 데서 오는 결과라고 추측된다. 그러나 고유치(eigenvalue)에 의한 제 1요인의 설명량이 61%로 나타났고, 5요인구조의 요인 패턴도 단순구조가 나타나지 않았기 때문에 1요인으로 보는 것이 적절하다고 판단하였다.

표 1은 1요인 구조의 요인부하량을 보여준다. ‘진실만을 말하는 편이다’ ‘내키는 대로 행동하는 것을 좋아한다’ ‘성급한 결정을 내린다’를 제외한 모든 문항의 요인 부하량이 .30을 넘었다. 비록 위 세 문항은 요인 부하량이 낮았지만, 이상점 반응모형에 적합한 문항이 우세반응모형을 기초로 하는 요인분석모형에 적합하지 않게 나타날 가능성이 있기 때문에 본 분석에 포함시키고 그 결과를 관찰하기로 하였다.

### 모형의 적합도 비교

우세과정을 가정하는 일반화 부분 점수 모형(Muraki, 1992)과 이상점 모형인 GGUM을 각각 30개의 성실성 문항 응답에 적용하였다. 표 2는 MODFIT(Stark, 2001)을 사용하여 계산한 조정된  $\chi^2/df$  비율의 평균, 표준편차, 그리고 빈도를 제시하고 있다<sup>2)</sup>. MODFIT에서는 각

개별 문항(singlet)에 대한  $\chi^2$  통계치와 아울러 두 문항의 쌍(doublet), 세 문항의 셋(triplets)에 대한  $\chi^2$  통계치를 계산해 준다. Drasgow과 동료들(Drasgow, Levin, Tsien, Williams, & Mead, 1995)은 조정된  $\chi^2/df$  비율이 1 보다 작으면 ‘매우 작다’, 1-2 사이이면 ‘작다’, 2-3 사이이면 ‘약간 크다’, 3 이상이면 ‘크다’로 해석할 수 있다고 제안하였다. 이 기준에 의하면, 두 모형에서 개별문항별로 보았을 경우 30개 문항 모두에서 조정된  $\chi^2/df$  비율이 1보다 작고 그 평균이 0.01( $SD=0.02$ )로 나타나, 두 모형이 모두 자료에 적합하다고 볼 수 있다. 두 개의 문항들의 짝들에 대해 계산한 경우에는 일반화 부분 점수 모형에서 큰 것으로 볼 수 있는 3 이상의 조정된  $\chi^2/df$  비율이 한 문항에서 나타났으며 GGUM에서는 30개의 문항 모두에서 3이하로 나타났다. 그러나 조정된  $\chi^2/df$ 의 평균은 일반화 부분 점수 모형에서 1.43 ( $SD=0.62$ )이고, GGUM에서는 1.44( $SD=0.58$ )로 나타나 두 모형에서 적합도의 차이가 거의 없는 것으로 판단된다. 세 문항의 짝들의 경우에서도 두 모형의 적합도 차이가 거의 나타나지 않았다.

본 연구에서 사용된 5개의 응답선택지 중의 마지막 선택지(5=“정말 그렇다”)에 대한 이상점 가정 모형의 문항 옵션반응함수(option response function: ORF) 곡선을 살펴보면, 총 30개의 문항 중에서 11개의 문항 (문항 2, 5, 9, 11, 12, 13, 17, 19, 20, 22, 24)에서 이상점 가정 모형을 지지하는 비단조적(non-monotonic) 반응곡선이 나타났다.

지면상 각 모형에 대한 모든 문항의 적합도 곡선을 제시할 수 없기 때문에, 그림 3에 문

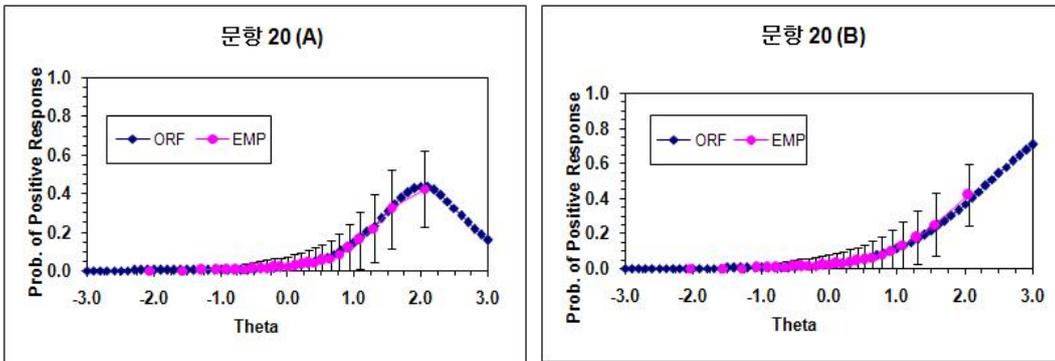
2) Stark(2001)에서 권장하는 바와 같이  $\chi^2$ 를 표본수가 3,000일 경우로 조정하였다.

표 1. 성실성 척도 요인부하량

번호	문항	성실성요인
1	좋은 해결책을 잘 생각해 낸다.	0.47
2	일을 어떻게 처리해야 되는 지를 안다.	0.49
3	일을 순조롭게 처리한다.	0.52
4	내가 하는 일에서 남보다 뛰어나다.	0.48
5	일을 성공적으로 완수한다.	0.60
6	정돈되지 않는 것들이 있어도 거슬리지 않는다.	0.40
7	계획에 따라 일을 한다.	0.57
8	깨끗이 치우는 것을 좋아한다.	0.48
9	질서 있고 규칙적인 것을 좋아한다.	0.49
10	쓴 물건을 제자리에 갖다 놓는 것을 자주 잊어버린다.	0.42
11	내가 한 약속을 지킨다.	0.44
12	규칙을 준수하고자 노력한다.	0.43
13	양심에 귀 기울인다.	0.47
14	진실만을 말하는 편이다.	0.23
15	내가 할 일을 남에게 미룬다.	0.44
16	성공하고자 하는 강한 의욕이 없다.	0.44
17	남들이 나에게 기대하는 것 보다 더 열심히 한다.	0.62
18	전심으로 맡은 바 일에 몰입한다.	0.57
19	열심히 일한다.	0.60
20	계획을 행동으로 옮긴다.	0.65
21	일을 시작하는데 누구가의 재촉이 필요하다.	0.38
22	해야 할 일을 미루지 않고 바로바로 처리한다.	0.61
23	결정을 나중으로 미룬다.	0.46
24	할 일이 생기면 곧바로 시작한다.	0.48
25	시작한 일은 끝을 낸다.	0.59
26	생각 없이 무언가를 덩석 시작한다.	0.30
27	내키는 대로 행동하는 것을 좋아한다.	0.27
28	성급한 결정을 내린다.	0.23
29	말을 신중히 골라한다.	0.37
30	미리 준비하지 않고 마지막 순간에 가서야 계획을 짜는 편이다.	0.53

표 2. 성실성 척도에 대한 조정된  $\chi^2/df$  비율 (adjusted  $\chi^2/df$  ratio)의 평균과 표준편차, 빈도

	Adjusted $\chi^2/df$ ratio의 빈도분포					평균	표준편차
	<1	1<2	2<3	3<4	>4		
<u>일반화 부분 점수 모형</u>							
개별문항(Singlets)	30	0	0	0	0	0.01	0.02
2문항쌍(Doublets)	7	19	3	1	0	1.43	0.62
3문항쌍(Triplets)	1	9	0	0	0	1.46	0.34
<u>GGUM</u>							
개별문항(Singlets)	30	0	0	0	0	0.07	0.07
2문항쌍(Doublets)	6	19	5	0	0	1.44	0.58
3문항쌍(Triplets)	1	9	0	0	0	1.51	0.35



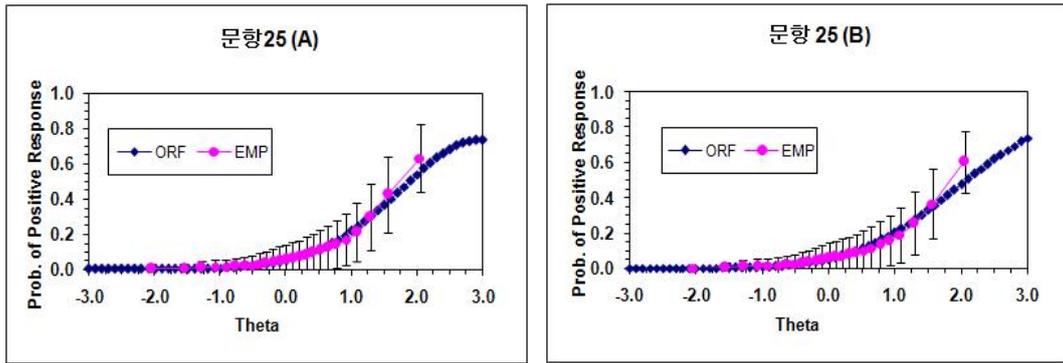
주. A=이상점 가정 모형, B=우세과정 모형  
 ORF=응답반응곡선(option response function)  
 EMP=경험적인 응답반응(empirical item response function)

그림 3. 문항 20의 대한 응답지 '5=정말 그렇다'에 대한 적합도 곡선

항 20 그리고 그림 4에 문항 25에 대한 이상점 가정 모형(A)과 우세과정 모형(B)의 적합도 곡선을 제시하였다. 이 적합도 곡선은 응답선택지 중의 마지막 선택지(5="정말 그렇다")에 대한 것이다. 각각의 문항에서 A 그림은 이상점 모형을 적용했을 때의 반응곡선을, B 그림은 우세 과정 모형을 적용했을 때의 반응곡

선을 보여주고 있다.

문항 20의 A 그림에서 보는 바와 같이, 이상점 모형을 적용했을 때 특질값(theta)이 0 이하인 경우에는 '정말 그렇다'를 선택할 확률이 거의 없다가 특질값이 0 이상인 영역에서 특질값이 증가함에 따라 선택 확률이 서서히 증가하였다. 그러나 특질값이 2 이상인 영역에



주. A=이상점 가정 모형, B=우세과정 모형  
 ORF=응답반응곡선(option response function)  
 EMP=경험적인 응답반응(empirical item response function)

그림 4. 문항 25의 대한 응답지 '5=정말 그렇다'에 대한 적합도 곡선

서는 오히려 그 확률이 점점 감소하는 접힘 현상이 나타났다. 이에 비해 문항 20(B) 그림은 우세과정 모형을 적용한 ORF 곡선을 보여주는데 여기서는 특질값이 증가함에 따라 '정말 그렇다'라고 응답한 확률이 단조적으로 증가하며 접힘 현상이 나타나지 않았다.

이 두 반응 곡선을 비교해 보면, 이상점 모형과 우세과정 모형이 예측하는 반응곡선의 차이는 특질값이 2보다 큰 구간에서만 명확하게 드러나는 것을 알 수 있다. 즉, 특질값이 2보다 작은 구간에서는 두 모형이 예측하는 곡선이 거의 차이가 없다가 특질값이 2인 영역에서 이상점 모형에서는 접힘 현상이 발생하며 특질값이 증가할수록 5번 응답지를 선택한 확률이 차차 감소하는 반면, 우세과정 모형에서는 지속적인 단조적 증가 패턴을 보여 주고 있다. 그런데 경험적인 응답반응(empirical item response function, EMP)을 보면, 이 구간에 속하는 경험적인 자료가 거의 없다는 것을 알 수 있다. 이와 같이 두 모형이 서로 다른 반응곡선을 예측함에도 불구하고 해당 구간에 경험

적인 자료가 없기 때문에 자료에 대한 모형적합도에서는 두 모형간의 차이가 거의 나타나지 않은 것으로 추측된다. 이상의 결과는 이상점 가정 모형의 적절성을 보여주는 접힘 현상이 나타난 다른 10개의 문항에서도 유사하게 나타났는데, 대부분의 문항에서 특질값이 2 근처에서 접힘 현상이 보였다.

한편, 그림 3에서 보는 바와 같이, 문항 25의 적합도 곡선은 이상점 가정 모형(A)과 우세과정 모형(B)에서 서로 거의 비슷한 형태의 반응함수를 보여주고 있으며, 이 두 모형 모두 매우 응답자료에 적합한 것으로 나타났다. 이상점 반응 모형을 적용한 경우에도 접힘 현상이 나타나지 않은 나머지 18개 문항에서도 이와 유사한 결과가 관찰되었다.

#### 두 모형의 예측타당도 비교

우세과정 모형과 이상점 가정 모형을 바탕으로 산출된 성실성 점수의 예측타당도를 평가하고자 각 모형에서 계산된 성실성 점수(특

표 3. 각 성실성 점수, 자기효능감, 학업성적의 상관 및 평균과 표준편차\*

	성실성 우세모형	성실성 이상점모형	자기효능감 (1차)	자기효능감 (2차)	학업 성적
성실성우세모형 <sup>1</sup>					
성실성이상점모형 <sup>2</sup>	0.96				
자기효능감1차 <sup>3</sup>	0.51	0.51			
자기효능감2차 <sup>4</sup>	0.33	0.35	0.61		
학업성적	0.37	0.37	0.34	0.38	
평균	-0.04	-0.02	475.91	425.08	79.00
표준편차	0.97	1.01	88.45	121.47	16.20

주. 성실성(N=129), 자기효능감 1차(N=127), 자기효능감 2차(N=119), 학업성적(N=117)

모든 상관계수는  $p < .01$ 로 유의함

1. 성실성 우세모형: 우세과정 모형에서 산출된 성실성 특질값
2. 성실성 이상점모형: 이상점 가정 모형에서 산출된 성실성 특질값
3. 자기효능감 1차: 학기초에 측정된 중간시험에 대한 자기효능감
4. 자기효능감 2차: 학기중반에 측정된 기말시험에 대한 자기효능감

질값)와 본 연구에서 준거변인으로 측정된 중간시험에 대한 자기효능감(1차), 기말시험에 대한 자기효능감(2차), 그리고 중간시험과 기말시험의 점수의 평균인 학업성적 간의 상관계수를 계산하였다. 표 3은 각 변인들의 상관계수 및 평균과 표준편차를 제시하고 있다.

우선 우세모형을 기초로 채점된 성실성 점수와 이상점모형을 기초로 채점된 성실성 점수간의 상관이 .96으로 나타나 어느 모형을 기초로 채점했는가에 따라 성실성 점수의 차이가 거의 없다는 것을 알 수 있다. 실제로 두 성실성 점수 모두 1차와 2차의 자기효능감 평정 그리고 학업성적과 중간정도의 상관관계를 가졌다. 1차 자기효능감과는 가장 상관이 높아, 우세과정 모형과 이상점 가정 모형에서 산출된 성실성 점수는 동일하게 1차 자기효능감과 .51의 상관계수를 보였다. 2차 자기효능

감 평정과는 이상점 가정 모형에서 산출된 성실성 점수와 .35의 상관계수로 보인 반면, 우세과정 성실성 점수와는 이보다 약간 낮은 .33의 상관계수를 보였다. 마지막으로 학업성적과는 우세과정 모형과 이상점 가정 모형에서 산출된 성실성 점수 모두 .37의 상관계수를 가졌다. 따라서 이상점 가정 모형과 우세과정 모형은 각각의 준거변인에 대한 예측타당도가 매우 유사하였다.

## 논 의

본 연구는 5점 척도로 측정된 성실성 척도에 이상점 반응 모형과 우세과정 모형을 적용시켜 봄으로써 다양한 성격 측정 도구들이 암묵적으로 가정하고 있는 문항반응과정에 대한

근본적인 가정을 검증해 볼 수 있는 기회를 제공하고자 하였다. IPIP 문항들을 기초로 본 연구에서 개발한 성실성 척도에 대한 응답 자료를 분석한 결과, 약 삼분의 일의 문항들에서 이상점 반응모형을 지지하는 종모양의 옵션반응곡선이 나타났다. 그러나 대체로 모든 문항에서 우세과정 모형과 이상점 반응 모형 모두가 모두 적절한 것으로 나타났다. 또한 이상점 과정 모형과 우세 과정 모형을 각각 채점에 적용하여 각 모형에서 산출된 점수와 학습관련 준거변인들과의 상관관계의 정도를 비교한 결과 두 모형에서 계산된 성실성 점수가 거의 유사한 수준의 예측타당도를 보였다.

이러한 본 연구의 결과는 16요인 성격검사에 대한 이분화된 응답자료를 분석한 Stark 등(2006)의 결과와 유사하다. Stark 등의 연구에서는 걱정(apprehension)척도와 같이 종 모양의 반응곡선을 보이는 문항이 있는 척도에서는 이상점 모형이 우세과정 모형 보다 더 적합하지만, 그러한 문항을 포함하지 않은 척도들에서는 이상점 모형과 우세과정 모형과 비슷한 정도의 적합도를 보이는 것을 확인하였다. Stark 등(2006)의 결과와 본 연구의 결과가 다른 것은 본 연구에서는 언폴딩 문항반응모형에서 종 모양의 반응 곡선을 보이는 문항들에서도 이상점 모형의 적합도가 로지스틱 문항 반응 모형의 적합도보다 우월하게 나타나지 않았다는 것 점이다. 이러한 차이는 본 연구의 참여자 중 성실성의 특질값이 2 이상인 사람은 채 3%가 되지 않은 반면, 언폴딩 문항반응모형에서 종 모양의 반응 곡선을 보이는 문항들에서 접힘이 나타나는 지점이 대부분 특질 값 2 이상이었던 점과 관련 있을 것으로 추측된다. 즉, 이상점 모형이 적용되지 않았을 때 특질값과 응답반응간의 관계의 왜곡이 나

타날 수 있는 극단적인 특질값 영역에 존재하는 관찰치들이 극히 소수였기 때문에 이상점 모형과 우세모형간의 적합도에서 차이가 나타나지 않을 것이라고 해석할 수 있다.

종합해 볼 때, 본 연구의 결과는 성실성 척도에 대한 응답 자료에서 이상점 모형이 우세 모형과 대등한 수준의 적합성을 가짐을 알 수 있다. 또한 예측타당도에서도 두 모형 간에 큰 차이가 없는 것으로 나타났다. 그러나 본 연구에서 사용된 문항이 대부분 극단적인 특질값을 가지는 문항들로 구성되어 있고 중간 수준의 특질값을 가지는 문항은 거의 없었던 점에 유의하여 이러한 결과를 해석할 필요가 있다.

고전검사 이론을 바탕으로 한 요인분석을 이용하여 척도를 구성할 때 문항선발의 중요한 선택 기준은 동일 척도 내의 다른 문항들과의 상관이 높아 요인분석시 공통요인에 의해 설명되는 변량의 양, 즉 요인부하량이 높은가의 여부이다. 그런데 만약 응답자들이 성격문항에 응답할 때 이상점 가정 모형을 따른다면, 중간 수준의 특질을 측정하는 문항에서는 그 특질에서 낮은 위치를 가진 사람들과 높은 위치를 가진 사람들 모두 응답할 확률이 낮다. 따라서 이러한 중간 수준의 문항들은 높은(혹은 낮은) 특질을 가질수록 응답확률이 높아지는 극단적인 위치에 있는 다른 문항들과의 상관이 낮을 수밖에 없다. 그러므로 고전검사 이론상의 문항선발 기준 하에서는 중간 정도의 특질을 나타내는 문항들이 척도구성에서 제거되기 쉽고, 이러한 이유 때문에 성격 문항 작성자들도 중간 수준의 특질값을 가지는 문항보다는 양극단의 문항을 작성하는 것이 일반적인 경향이다(Stark et al., 2006). 사실상 본 연구에서 사용한 IPIP의 문항들도 이

러한 경향에서 크게 벗어나지 않아 이상점 모형의 적합성을 평가하는 데 한계가 있을 수 있다. 향후 연구에서는 중간 수준의 특질값을 가지는 문항으로 이루어진 성격 척도에서 이상점 모형을 적용해 볼 필요가 있을 것이다.

이러한 견지에서 생각해 보면 중요한 질문은 과연 중간 정도의 특질을 나타내는 문항들을 성격 척도에 포함시키는 것이 필요한가이다. 만약 그런 문항들이 포함되는 것이 더 좋은 척도를 구성하는 데 보탬이 되지 않는다면, 중간 정도의 특질을 나타내는 문항들을 개발할 필요가 없을 것이며, 우세과정 모형에 대한 이상점 모형의 상대적 유용성도 높지 않을 것이기 때문이다. 좋은 척도의 기준의 하나는 원하는 특질 범위의 응답자들의 특질 값을 정확하게 측정해 내는 것인데, 성격검사는 이용 목적에 따라 어느 수준의 특질 범위에 관심이 있는가를 다룰 수 있다. 예를 들어, 일반적인 성격검사용이라면 전 범위의 특질값을 가진 응답자들을 골고루 정확하게 측정해 내기를 원할 것이다. 이와는 달리, 인사선발의 도구로 특정 성격검사를 사용한다면, 극단적인 범위의 특질 값들을 잘 측정해 내는 검사를 원할 것이다. 우리가 모든 범위의 특정 성격특질을 정확히 측정하고자 한다면, 즉, 검사정보곡선(test information curves)이 전 범위에 걸쳐 일정한 수준이 되는 검사를 구성하기를 원한다면, 과연 중간 수준의 특질값을 나타내는 문항들을 포함하는 것이 도움이 될 것인가 생각해 볼 수 있다. 그런데 일반적으로 성격검사들은 이분화 선택지를 사용하기 보다는 5점이나 7점의 선택지를 사용하기 때문에 기본적으로는 비록 양극단적인 특질을 나타내는 문항이라고 할지라도 중간 정도의 특질 값을 가진 응답자에 대해서도 특질 값을 추정할 수 있다. 문제

는 과연 양극단의 특질 값을 나타내는 문항들만을 사용하여 다중척도로 재는 것과 중간 수준의 특질 값을 나타내는 문항들도 검사에 포함하여 다중척도로 측정할 때 검사의 정보량(test information)이 증가하는 가일 것이다. 본 연구에서 수집된 자료로는 이에 대한 답을 경험적으로 검증할 수 없지만, 향후 이에 대한 연구가 이루어져야 할 것이다.

이상점 모형과 우세모형을 선택할 때 또 하나 고려해야 할 점은 우세모형은 이상점 모형의 특수한 경우라는 것이다. 본 연구의 결과에서 보는 바와 같이 우세모형이 적절할 때는 이상점 모형을 적용해도 우세모형과 동일한 반응곡선을 예측하였다. 따라서 특정 자료에 이상점 모형이 보다 적절한 지 혹은 우세 모형이 보다 적절한 지에 대한 확고한 사전 가정이 없을 경우에는 보다 일반적인 이상점 모형을 사용하는 것이 안전할 것이다. 이는 요인분석에서 요인간의 상관없이 없다고 가정하는 직교회전 보다는 이러한 사전 가정이 없는 사각회전을 이용하는 것이 보다 정확하고 현실적인 모형이라는 논리(Fabrigar, Wegener, MacCallum, & Strahan, 1999)와 유사하다고 볼 수 있다.

본 연구의 의의는 성격 5요인 중에서도 현재 직무수행도 및 학습수행도 등의 예측변인으로 가장 광범위하게 사용되고 있는 성격 변인인 성실성을 측정하는 문항에 대한 사람들의 응답방식에 대한 근본적인 가정을 경험적인 자료를 통해 검증해 보았다는 것이다. 특히 태도나 성격척도에 이상점 반응모형을 적용한 기존의 연구들은 이분화된 응답지를 이용한 반면 본 연구는 5점 척도로 측정된 자료를 분석하였다는 차별성을 지닌다. 또한 본 연구는 단지 기존의 연구들을 다른 분야에 적

용시키는 것을 넘어 이상점 모형과 우세 모형을 각각 채점에 적용하여, 각 모형에서 산출된 점수가 관련 행동 변인을 예측할 때 과연 어느 방식이 더 높은 예측타당도를 지니는지를 살펴보았다는 점에서도 의의를 가진다. 예측, 진단 등의 응용적인 용도를 사용하기 위해 성격검사 도구를 평가할 때, 예측타당도는 그 무엇보다도 중요한 기준을 제공하기 때문이다.

마지막으로, 본 연구의 일차적 목적은 아니었지만 본 연구에서 개발된 성실성 척도가 예측타당도를 비롯한 다른 척도 특성이 양호하였기 때문에 향후 연구와 예측을 위해 이용될 수 있는 유용한 성실성 측정도구를 개발했다는 점에서도 본 연구의 의의를 찾을 수 있을 것이다. 특히, 성실성 척도의 예측타당도를 평가하기 위한 수집된 학습 관련 준거변인들이 성실성과 서로 다른 시간에 상이한 방법으로 측정되었음에도 불구하고 적절한 예측타당도를 보였다는 점에 주목할 필요가 있다.

본 연구의 결과는 현재 사용하고 있는 다양한 성격 측정 도구들을 평가하고, 새로운 측정 도구들을 개발할 때 기존의 요인분석이나 로지스틱 문항반응 이론과 아울러 이상점 모형도 적합하다는 것을 보여 줌으로써 보다 다양한 성격검사의 구성 기준을 제시하였다. 올바른 반응 모형을 바탕으로 구성된 성격문항들과 채점방식에서 나온 점수들을 바탕으로 할 때 보다 높은 예측타당도를 가질 수 있을 것이며 이에 기초한 추론과 의사결정이 보다 적절할 수 있을 것이다. 타당도를 측정 점수에 기초한 특정한 해석이나 의사결정이 얼마나 적절한가에 대한 전반적인 평가(Messick, 1995)라고 할 때, 문항반응모형의 상대적 적절성을 평가하기 위한 이러한 노력은 결국 타당

화 작업의 일환이라고 할 수 있다. 또한 최근 들어, 우리나라의 많은 조직들에서 인성검사를 채용 기준으로 채택하고 있는데 이상점 문항반응 모형은 우세과정 모형과 마찬가지로 개인의 성격점수 산출과 컴퓨터 적응 검사(CAT)를 구성하는 데 이용될 수 있을 것이며, 이러한 시도들은 보다 정확한 측정을 보다 편리하게 하는데 도움이 될 것이라고 기대된다.

## 참고문헌

- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.
- Blickle, G. (1996). Personality traits, learning strategies, and performances. *European Journal of Personality, 10*, 337-352.
- Brown, R. D., & Harvey, R. J. (2003, April). *Detecting personality test faking with appropriateness measurement: Fact or fantasy?* Paper presented at the 2003 Annual Conference of the Society for Industrial and Organization Psychology, Orlando, FL.
- Chernyshenko, O. S., Stark, S., & Chan, K. Y., Drasgow, F., & Williams, B. A. (2001). Fitting item response theory models to two personality inventories: Issues and Insights. *Multivariate Behavioral Research, 36*, 523-562.
- Colquitt, J. A., & Simmering, M. J. (1998). Conscientiousness, goal orientation, and motivation to learn during the learning process: A longitudinal study. *Journal of Applied Psychology, 83*, 654-665.

- Costa, P. T., & McCrae, R. R. (1992). *Professional manual for the NEO Personality Inventory NEO-PI-R and NEO Five Factor Inventory NEO-FFI*. Odessa, FL: Psychological Assessment Resources.
- Connor-Smith, J. K., & Flachsbart, C. (2007). Relations between personality and coping: A meta-analysis. *Journal of Personality and Social Psychology, 93*, 1080-1107.
- Derefinko, K. J., & Lynam, D. R. (2007). Using the FFM to conceptualize psychopathy: A test using a drug abusing sample. *Journal of Personality Disorders, 21*, 638-656.
- Dragow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple choice tests. *Applied Psychological Measurement, 19*, 143-165.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology, A critical review and analysis. *Personnel Psychology, 39*, 291-314.
- Goldberg, L. R. (1990). An alternative description of personality, The big-five factor structure. *Journal of Personality and Social Psychology, 59*, 1216-1229.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Robert, H., Michael, A. C., Robert, C. C., & Harrison, G. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.
- Hampson, S. E., Andrews, J. A., Barckley, M., Lichtenstein, E. & Lee, M. E. (2000). Conscientiousness, perceived risk, and risk-reduction behaviors, A preliminary study. *Health Psychology, 19*, 496-500.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance, The big five revisited. *Journal of Applied Psychology, 85*, 869-879.
- International Personality Item Pool (2001). *A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences..* <http://ipip.ori.org/>.
- Lee, S., & Klein, H. (2002). Relationships between conscientiousness, self-efficacy, self-deception, and learning over time. *Journal of Applied Psychology, 87*, 1175-1182.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 5-53.
- Locke, E. A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals and task performance. *Journal of Applied Psychology, 69*, 241-251.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50*, 741-749.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.
- Reckase, M. D. (1996). Test construction in the 1990s, Recent approaches every psychologist

- should know. *Psychological Assessment*, 8, 354-359.
- Reise, S. P., & Henson, J. M. (2000). Computerization and adaptive administration of the NEO-PI-R. *Assessment*, 7, 347-364.
- Reise, S. P., & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement*, 14, 45-58.
- Roberts, J. S., & Laughlin, J. E. (1996). A unidimensional item response model for unfolding responses from a graded disagree-agree response scale. *Applied Psychological Measurement*, 20, 231-255.
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Education and Psychological Measurement*, 59, 211-233.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- Stark, S. (2001). *MODFIT: A computer program for model-data fit*. Unpublished manuscript, University of Illinois at Urban-Champaign.
- Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment, Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91, 25-39.
- Terracciano, A., & Costa, P. T. Jr. (2004). Smoking and the Five-Factor Model of personality. *Addiction*, 99, 472-481.
- Thissen, D. (1991). *MULTILOG user's guide - Version 6*. Chicago, Scientific Software International.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554.
- 1 차원고집수 : 2008. 9. 30.  
수정원고집수 : 2008. 11. 17.  
최종게재결정 : 2008. 11. 24.

## Application of ideal point response model to personality scale

**Sunhee Lee**

Chungnam National University Department of Psychology

The present study examined whether ideal point response model is adequate for personality measurement scales by comparing the fits and predictive validities of ideal point response model and dominance response model. The analysis of data on a conscientiousness scale developed based on International Personality Item Pool showed that some of the items have bell-shape option response plot, which supports ideal point response model. However, both ideal point response and dominance response model fit the data similarly well. Predictive validity coefficients of conscientiousness scores based on each item response process model predicting academic achievement-related criteria were also similar to each other. The implications of the results were discussed in terms of development and scoring personality scales and the usage of personality scale scores.

*Key words* : *personality measurement, item response process, ideal point response model, dominance process model, conscientiousness*