

성격 측정에서 우세반응 모형과 이상점반응 모형의 비교*

박 동 건[†] 최 명 옥 이 원 선 이 혜 민 이 준 희

고려대학교 심리학과

본 연구는 성격 측정에 있어서 어떠한 측정모형이 가장 적합한지를 검증하기 위하여 고전검사이론과 우세(dominance) IRT 반응 모형, 이상점 모형(ideal point model)을 이용하여 성취지향, 공정성, 협조성, 진실성 등 네 차원의 문항들을 개발 및 선별하여 모형간 적합도를 비교하였다. 그 결과, 대체로 이상점 모형의 적합도가 우세 반응 모형보다 약간 더 높은 것으로 나오기는 하였지만 그 차이는 매우 미미하였다. 또한 진실성과 협조성 차원에 대해서는 두 모형으로 개발된 척도의 검사정보함수 간 차이는 거의 없었다. 준거 타당도를 살펴보기 위하여 세 가지 측정모형으로 개발된 네 차원 척도와 성적 간의 상관을 살펴본 결과 상관이 유의미하지는 않았지만 측정모형 간 차이는 거의 없었다. 논의에서는 본 연구의 결과 및 본 연구가 갖는 한계점, 산업 현장에서 갖는 의미에 대해서 살펴보았다.

주요어 : 성격 측정, 검사 구성, 문항반응이론, 고전검사이론, 우세과정, 이상점반응

* 본 논문은 2011년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 연구되었음(NRF-2011-327-B00974)

† 교신저자 : 박동건, 고려대학교 심리학과, sykhpark@korea.ac.kr

성격은 오랫동안 심리학에서 주요 연구주제 중 하나였으며, 성격에 대한 측정은 다양한 분야에 적용되어 활용되고 있다. 1990년 대 초 이래로 산업 및 조직심리 연구에서도 인사 선발을 위한 목적으로 성격검사가 많이 사용되고 있다(Hough & Scheider, 1996). 최근 기업들은 우수한 인재뿐 아니라 자사에 '적합한 인재(right people)'를 찾기 위해 자사와 개인 특성(trait) 간의 부합을 측정하기 위한 측면에서 성격검사를 많이 활용하고 있다(김명소, 이현주, 2006). 선발과정에서 성격검사의 사용은 직무수행 및 맥락수행 그리고 작업장에서의 폭력과 같은 다양한 준거를 예측하는 타당한 도구이며 생산성과도 깊은 관련을 맺고 있다는 연구 결과(안여명, 유태용, 2010; 유태용, 2007; 유태용, 민병모, 2001; Barrick & Mount, 1991; Hogan, Hogan, & Roberts, 1996; Hurtz & Donovan, 2000; Ones, Viswesvaran, & Reiss, 1996 등)들이 빈번히 제시되고 있다. 또한 성실성과 같은 일부 성격 차원들은 인지능력보다 수행을 더 잘 예측하는 것으로 나타났다(Schmidt & Hunter, 1998).

현재 행해지고 있는 성격측정은 대부분 자기보고의 질문지식 도구를 이용하여 측정되고 있다. 따라서 신뢰롭고 타당한 성격검사 도구의 개발은 성격의 이론적 연구나 실제적 적용에서 매우 중요한 기초 작업이라고 할 수 있다(이선희, 2008). 그 동안 성격 특성을 분류하는 방법을 개발하고, 중요한 성격특성이 무엇인지 확인하는 연구(예: Big Five factor model, Goldberg, 1997; 유태용, 김명언, 이도형, 1995)나, 성격측정의 유용한 적용의 향상을 위한 연구(예: Ellis, Becker, & Kimmel, 1993; Reise & Henson, 2000; Rouse, Finger, & Butcher, 1999; Simms & Clark, 2005), 그리고 문항 반응에 대

한 검사 상황의 영향(Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001) 등을 탐색한 연구는 많았다. 그러나 성격검사의 중요성과 높은 활용도에 비해 개인들이 어떠한 방식으로 성격문항에 응답하고, 이를 점수화(scoring)하는지와 관련되어 문항반응과 척도 구성절차의 본질을 검증하는 연구들은 거의 없었다(Stark, Chernyshenko & Drasgow, 2006).

일반적으로 성격 특성을 점수화하기 위해서는, 먼저 측정하고자 하는 특성의 구성개념을 명확히 정의하고, 이 정의를 반영하는 다수의 문항을 개발한다. 그리고 이 문항들 중에서 측정하고자 하는 구성개념을 가장 잘 측정할 수 있는 문항을 선별하여 성격검사 도구를 구성하게 된다. Stark 등(2006)은 이 과정에서 성격문항에 대한 응답이 어떤 과정을 거쳐 나타나는지에 대한 규명이 반드시 이루어져야 한다고 주장한다. 다시 말해, 응답과정 본질에 대한 가정이 반드시 만들어져야 한다는 것이다. 응답과정 본질에 대한 가정은 성격문항에 대한 응답이 척도 점수와 어떻게 연결되는지 그리고 척도를 구성할 때 어떤 문항이 제거되고 남는지에 직접적으로 영향을 미친다. 따라서 응답자의 반응 과정에 대한 옳지 않은 가정은 성격 검사점수의 정확성과 응답자 행동 예측에 부정적인 영향을 미칠 수 있다는 것이다.

오늘날 사용되는 대다수의 성격 척도는 고전검사이론과 요인 분석 기법에 근거한 Likert (1932)의 접근 방식에 기반하고 있고, Likert 접근은 문항반응의 기저로 우세과정(dominance process)을 가정하고 있다. 그러나 최근 일부 연구자들은 일련의 경험적 연구들(예: Chernyshenko, Stark, Drasgow & Roberts, 2007; Stark et al., 2006)을 통하여 성격측정에서 이러

한 우세반응 가정을 비판하며, 이상점반응 과정(ideal point response process)의 적용을 주장하였다. 이들은 문항반응에 대한 우세 접근은 많은 상황, 특히 인지능력 측정에는 적합하지만, 성격검사와 같은 비인지적 측정에 대해서는 문항 반응의 기저로 이상점반응 과정이 더 적합하다고 설명한다.

성격측정에서 오랫동안 사용되어왔던 Likert 척도에 기반을 둔 우세과정에 대한 비판과 이상점반응 가정의 적용에 대한 주장은 성격 연구자들에게 새로운 논쟁거리를 제공하였다. 이에 미국 산업 및 조직심리학회 SIOP에서 발간되는 「*Industrial and Organizational Psychology*」 2010년 3호에는 성격측정에서 이상점반응 가정의 적용에 대한 여러 연구자들의 찬반양론이 게재되기도 하였다. 이러한 관점에서 성격 측정에서 우세반응 과정과 이상점반응 가정에 대한 일부 후속 연구들(Tay, Drasgow, James, & Williams, 2009; Tay, Ali, Drasgow, & Williams, 2011)이 진행되었으며, 국내에서도 이선희(2008)가 성격측정에 대한 이상점 모형의 적용과 관련된 연구를 수행하였다.

그러나 이와 관련된 연구들은 아직 초기단계로서 제한된 장면에서 제한된 성격척도들에 대해서만 이루어진 상태이다. 더욱이, 지금까지 이루어진 대부분의 연구들도 Drasgow, Stark 및 그의 동료들에 의해 수행된 것으로서, 기본적으로 이상점반응 가정을 지지하는 연구자들의 입장에서 진행된 것이었다. 따라서 Stark 과 그의 동료들이 주장하는 것처럼 성격측정에서 이상점을 적용하는 것이 옳은지, 아니면 이상점 가정의 적용을 비판하는 다른 연구자들의 주장처럼 기존의 우세과정 적용이 옳은지에 대한 결론을 내릴 수 있을 만큼 충분한 연구가 이루어지지 않았다고 볼 수 있다.

75년 동안이나 많은 연구자들이 우세과정을 적용해왔을 때는 그것이 최선이었을 것이므로 몇 편의 연구를 통해 그 가정을 그렇게 쉽게 바꿀 수는 없을 것이다. 그렇지만 이상점반응 가정을 지지하는 몇몇 연구자들만의 주장이라고 치부해 버리기에는, 성격척도에서 이상점반응 가정을 적용해야 한다는 그들의 논리는 매우 설득력이 있다. Drasgow, Chernyshenko, 및 Stark(2010)은 “성격문항과 같이 내적성찰(introspection)을 요구하는 문항에 대한 응답은 비교과정을 포함한다.”고 주장한다. 이 주장에 대해서는 다른 연구자들(Brown & Mayduelivares; 2010)도 동의하지만, 여기서 비교하는 것이 무엇인지에 대해서는 연구자들 간에 차이가 있다. Drasgow 등은 응답자는 항상 자기 개념(self-perception)과 문항의 위치를 비교한다고 주장하지만, Brown과 Maydue-Olivares는 측정되는 구성개념과 구성개념을 측정하기 위해 사용되는 문항의 종류에 따라 개인이 비교하는 것이 달라진다고 주장한다. 이처럼 연구자마다 개인이 문항에 응답하는 과정에 대해 다른 생각을 가지고 있기 때문에 다양한 연구 장면에서 다양한 성격척도들에 대해 연구들이 지속적으로 수행되어 그 결과들이 축적됨으로써 종합적인 결론을 도출해야 할 필요성이 있다. Drasgow와 동료들(2010)의 주장처럼, 사람들이 문항에 응답할 때 어떠한 과정을 따르는가에 대한 가정은 어떤 문항을 좋은 문항으로 선택하여 최종 검사문항으로 포함할지, 어떻게 개개인의 점수를 산출할지, 그리고 더 나아가 그 검사에서 측정된 특성과 다른 변인과의 관계를 추론하고 이를 이용해 미래의 행동을 예측하는 데 중요한 영향을 미칠 수 있기 때문에 중요하다(이선희, 2008).

본 연구에서는 이러한 관점에서, Chernyshenko

등(2007)이 수행한 연구방법론을 토대로 기존의 연구에서 다루지 않은 다른 성격척도들에 대해 성격측정에서 우세 과정과 이상점반응 가정의 적용을 비교해 보고자 한다. 이를 위해 먼저 우세 과정과 이상점반응 가정에 대해 개략적으로 살펴보고, 각각의 가정에 대한 연구자들의 비판점을 알아볼 것이다. 다음으로 Chernyshenko 등(2007)의 연구와 동일한 방법론을 적용하여 고전검사이론(classical test theory; CTT), 우세 IRT(item response theory) 모형, 이상점 IRT 모형을 적용하여 문항을 개발하고, 각 모형을 통해 개발된 문항에 대한 적합도를 비교함으로써 어떠한 모형이 성격측정에 더 적합한지에 대해 논의할 것이다.

우세 과정과 이상점반응 과정 가정 비교

Thurstone과 Likert 척도는 모두 태도에 대한 연구 장면에서 척도를 구성하고 점수화하기 위한 방법으로 개발되었다. 그러나 Likert 방법으로 개발된 척도는 높은 문항-총점 상관, 높은 내적 일치 신뢰도 그리고 우세 단일 요인 구조를 갖는 경향이 있는 반면, Thurstone 접근은 낮은 문항-총점 상관과 낮은 내적 일치 신

뢰도와 2요인 구조를 갖는 경향이 있다 (Davison, 1977).

Thurstone과 Likert 접근 간 기본적인 차이는 문항 반응에 대한 기본 가정이다(Stark et al., 2006). 먼저 Likert는 우세 과정을 기본으로 가정한다. 즉, 개인과 문항들은 한 특성의 연속 선상에서 특정한 위치를 가지며, 한 개인이 문항에서 표현된 것보다 그 특성을 더 많이 가지고 있을수록 그 문항에 동의할 확률이 커진다고 본다. 그림 1(Stark et al., 2006)은 전형적인 우세반응 과정의 문항반응함수(item response function; IRF)를 보여준다. 이 그림에서 개인이 문항에 동의할 확률은 개인이 문항 위치보다 위에 있고, 문항과의 거리가 증가할수록 높아진다(부정문항은 역산으로 채점하여 동일하게 나타남). 그래서 우세 과정에 대한 문항반응함수는 일반적으로 단조적(monotonic)이며, S자 형태의 곡선을 가진다. 예를 들어, “내 방의 깔끔함은 평균수준이다.”와 같은 중간수준의 질서성(order) 문항에 대해 질서성이 낮거나 중간수준인 사람보다 극단적인 질서성을 가진 사람이 이 문항에 동의할 확률이 높다는 것을 의미한다.

이와 달리, Thurstone은 이상점반응 과정을

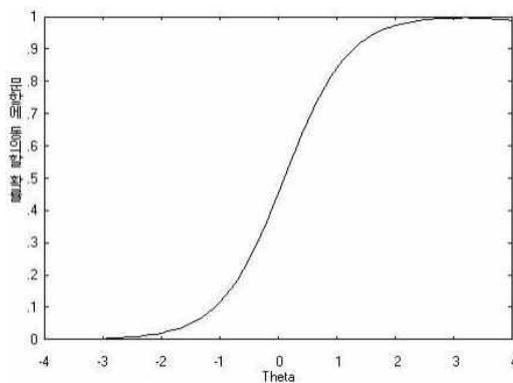


그림 1. 우세반응 과정

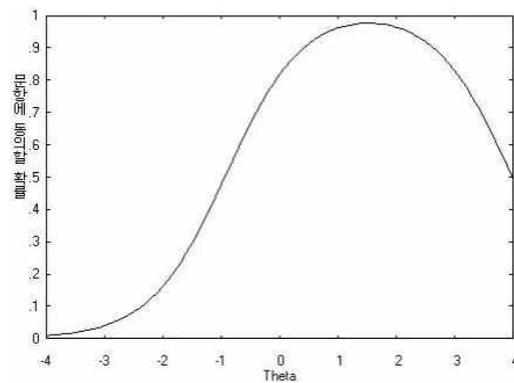


그림 2. 이상점반응 과정

기반으로 한다. Thurstone(1927, 1928)은 개인들이 어떤 문항에 응답할 때, “이 진술문이 나를 가깝게 묘사하고 있나?”를 생각하고, 특성연속선상에서 문항의 위치와 자신의 위치가 가까울 때, 사람들이 그 문항에 동의할 확률이 더 커진다고 주장한다. 문항의 태도 수준과 개인의 태도 수준이 동일할 때, 문항에 동의할 확률은 최대가 된다는 것이다. 즉, 사람들은 자신의 특성 수준이 문항의 특성 수준보다 더 높거나, 더 낮은 경우에는 문항에 동의할 확률이 낮아진다. 따라서 특성 연속선상(trait continuum)에서 개인의 위치와 문항 위치간 거리가 증가할수록 문항에 동의할 확률은 감소한다. 이 가정은 그림 2(Stark et al., 2006)와 같은 비단조적(nonmonotonic)이고 종모양의 문항반응함수를 가진다. 위의 우세 과정에서의와 같이 질서성의 예로 살펴보면, 이상점 과정은 “내 방의 깔끔함은 평균수준이다.”라는 문항에 대해 질서성이 높거나 낮은 사람들보다 특성연속선상에서 중간 수준의 질서성을 가진 사람들이 문항의 위치와 더 가깝기 때문에 문항에 동의할 확률이 더 높다는 것이다.

이와 같이, 우세 과정 가정에서는 개인의 질서성이 증가할수록 질서성을 측정하는 문항에 동의할 확률이 단조적으로 증가하는 반면, 이상점반응 가정 모형의 문항반응함수는 단조적으로 증가하다가 어느 순간 이르러서는 다시 감소한다는 점에서 우세 과정 모형과 대조된다(이선희, 2008). 이처럼 응답확률이 증가하다가 감소하는 현상을 Coombs(1964)는 언폴딩(unfolding)이라고 불렀는데, 이는 개인이 문항에 동의할 확률이 개인의 이상점으로부터 양쪽 방향으로 감소하기 때문이다. 성격측정에서 이상점 모형 적용을 주장하는 연구자들은 이 언폴딩 현상을 이상점 모형이 보다 적절하

다는 것을 시각적으로 보여주는 결과라고 본다.

성격측정에서 우세반응 가정의 한계와 이상점반응 적용에 대한 논의

전술한바와 같이, 현재 일반적으로 성격검사를 개발하기 위해 가장 많이 적용되는 방법은 고전검사이론, 요인분석모형, 로지스틱 문항반응모형 등이다. 이들은 공통적으로 특정 특성에서 높은 위치를 가지고 있을수록, 그 특성을 재는 문항에 동의할 가능성이 높아지며, 따라서 더 높은 측정점수를 얻게 될 것이라는 우세 과정을 가정한다. 그리고 우세 가정은 모든 문항에 대하여 단조성을 가정한다.

우세 가정에 근거한 척도 개발 절차에서는 모든 문항이 단조적이라는 가정 하에 낮은 문항-총점 상관을 보이는 문항을 제거한다. 전통적 기준에서 낮은 문항-총점 상관은 문항의 변별성이 낮다는 증거가 된다. 그러나 Chernyshenko 등(2007)은 만약 문항 풀(pool)에 비단조적인 문항이 있을 경우에는, 문항-총점 상관과 요인부하량(factor loading)이 문항 질을 판단하기 위해 적합한 통계치가 아니라고 주장한다. 왜냐하면 그것들은 본질적으로 문항점수와 특성 수준 간에 단조적인 관계를 가정하기 때문이다. 하지만 실제 단조적이지 않은 문항이 존재하며¹⁾ 이런 경우에 문항-총점 상

1) 태도 검사에서 중형태의 문항반응함수를 갖는 비단조적 문항이 변별력을 가지고 있는 경우가 존재함을 확인한 연구 결과가 있다(Andrich, 1996; Roberts, Laughlin, & Wedell, 1999). 문항이 중형의 문항반응함수를 갖는 경우, 서로 다른 특성 수준을 갖는 피험자가 동일한 응답 확률에 의해 동일한 점수를 받게 될 수 있는데, 이것은 낮은 문항-총점 상관을 초래한다(Chernyshenko 등, 2007).

관으로 문항의 질을 판단하는 것은 옳지 않다는 것이다. 이들은 비단조적인 문항은 문항이 갖는 변별 능력에도 불구하고 기존의 전통적인 평가 준거에 의해 ‘좋지 않은’ 문항으로 판단되어 제거될 수 있다고 설명한다. 따라서 척도 개발 단계에서 비단조적 문항을 제거하는 것은 특성 수준 연속선상의 특정 부분에 대한 측정의 정확성이나 문항 내용에 의도하지 않은 영향을 미칠 수 있다는 것이다.

현재 일반적으로 사용되는 성격 검사 문항들의 대부분은 중립적(예: “달성할 수 있을 정도의 목표만 세운다.”)이지 않으며, 긍정적(예: “어려운 목표에 도전하는 것이 좋다.”)이거나 부정적(예: “성취하기 어려운 목표를 세우지 않는다.”)이다. 그러나 Chernyshenko 등(2007)은 성격척도에 이렇게 상대적으로 극단성을 반영하는 문항만을 사용하는 것은 다음과 같은 이유로 바람직하지 않은 심리측정적 결과를 초래할 수 있다고 주장한다. 첫째, 이러한 문항만으로 구성된 검사는 전체 특성 수준에 걸쳐 고른 측정의 정확성을 제공하지 못한다는 것이다. 긍정 문항과 부정 문항은 매우 비슷한 정보 함수를 가진다. 따라서 척도에 이런 문항들만 포함하는 것은 특정 특성 수준에서만 높은 정보를 제공한다는 것이다. 특성 수준 전 범위에서 높은 측정 정확성을 요구하는 검사를 개발하기 위해서는 다양한 문항정보함수를 가지는 문항 풀이 필요하다. 그러나 문항의 질을 평가하기 위해 문항-총점 상관을 사용하게 되면 비단조성을 가지는 문항들이 제거되므로, 다양한 정보함수의 문항 풀 구성이 어렵다.

이와 같이, Likert와 Thurstone 척도 간 중요한 차이는 중간 수준을 나타내는 문항 진술문에 대한 생각이 다르다는 것이다. 예를 들어,

“모든 나라에서 강제적인 군대 훈련은 감소되어야 하지만 없어져야 하는 것은 아니다.”와 같은 중간 수준을 나타내는 문항에 대해 Likert는 사람들이 이 진술문에 대해서 동의 또는 비동의 하는 것이 ‘감소’ 측면인지, 아니면 ‘없어지는 것’의 측면인지를 말하는 것이 불가능하다고 주장한다. 따라서 이 진술문은 이중적 의미(double-barreled)를 갖고, 태도 측면에서 개인을 구별하는 데 가치가 거의 없으므로 이러한 중간 진술문은 제거되어야 한다고 주장한다. 반면, Thurstone은 이 진술문을 중간 위치(intermediate standings)에 있는 사람들의 태도를 정확하게 측정하기 위해 필요한 것으로 생각하기 때문에 중간 문항을 의도적으로 만들고 척도에 포함한다(Dragow et al., 2010).

둘째, 문항위치와 문항내용 간의 관계이다. 우세과정 모형 내에서는 문항 위치 모수 통계치가 문항 내용과 직접적 관련이 없다. 우세관점에 근거한 성격척도에서 긍정 문항과 부정 문항의 위치 모수 통계치(CTT 부정 문항의 역상수 p 값 또는 IRT의 문항 곤란도 b)는 매우 유사한 값을 갖는다. 따라서 새로운 검사를 개발하거나 기존 검사를 향상시키려 할 때 검사개발자들은 어떤 유형의 문항들이 필요한지에 대한 사전 통찰을 갖지 못한다. 즉, 인지능력검사에서는 곤란도 지수가 문항 내용과 직접적으로 관련되어 있으므로 개발자들은 원하는 측정적 특성을 갖는 문항을 어떻게 개발해야 하는지 알 수 있지만, 성격척도에서는 불가능하다는 것이다.

셋째, 비단조 문항을 제거하는 것은 척도 구성을 위해 가용한 문항 풀을 감소시킨다는 점이다. 다양한 범위의 곤란도 값을 가지는 큰 문항 풀은 평행(parallel form) 검사 구성이나 컴퓨터 적용검사(computerized adaptive testing)

개발을 위해 중요하다. 것이다.

결과적으로 우세 가정이 갖는 이러한 한계점들은 척도 구성 방법과 관련된 심리 측정 기법의 발전을 제한해 왔고, 따라서 문항 개발 및 선택에서 ‘단조성’이라는 제약조건을 적용하지 않고 단조적, 비단조적 문항을 모두 검사 내에 포함시킬 수 있는 심리 측정 모형의 가능성을 고려하는 것이 필요하다는 것이 Chernyshenko와 동료들(2007)의 주장이다.

Chernyshenko 등(2007)의 주장은 상식적으로나 논리적으로 보면 일견 타당한 듯 보인다. 그러나 이러한 주장에 대한 회의적인 시각도 존재한다. 먼저 앞서 제시했듯이, 응답자들은 항상 자기개념과 진술문의 위치를 비교하기 때문에 이상점 모형을 사용하는 것이 옳다는 이들의 주장과는 달리 Brown과 Maydue-Olivares (2010)는 측정되는 구성개념과 문항들에 따라 개인이 비교하는 것이 달라진다고 주장한다. 이분척도인 경우 응답자는 그 문항의 효용성이 역치이상이면 문항에 동의할 것이고, 그렇지 않으면 동의하지 않을 것이기 때문에, 이러한 경우 우세 과정 모형의 사용이 요구된다는 것이다. 또한 이들은 이상점반응 가정과 우세 과정의 가장 큰 차이점이라고 볼 수 있는 ‘중간수준’ 문항 개발의 어려움과 ‘중간수준’ 문항의 다차원성에 대한 문제점도 제기하였다. ‘중간수준’ 문항은 다차원적 특성을 가지므로 측정하고자 하는 구성개념과 관련 없는 다른 특성을 측정할 수도 있다. 따라서 중간문항을 넣는 것이, 실제 연구자가 측정하고자 하는 특성의 ‘중간수준’을 측정하지 못하고 오히려 응답자들을 혼란에 빠뜨릴 수도 있다.

Drasgow 등(2010)은 우세반응은 최대수행을 측정하는 인지능력 자료에 적합하지만, 일반적 특성을 측정하는 성격자료에는 적합하지

않다고 주장한다. 이에 대해 Reise(2010)는 Drasgow가 오랫동안 비인지적 구성개념 측정치에 대해 IRT 모형을 적용해온 연구결과들을 묵살하는 행위라고 비판한다. 즉, 매우 엄격하고 타당하게 연구된 많은 심리척도들이 우세반응 틀 내에서 개발된 척도라는 사실을 무시하고 있다는 것이다. 성격측정에서 이상점 반응 모형이 더 적합하다는 Drasgow 등(2010)의 주장이 경험적으로 충분히 지지되는지에 대한 의문이 제기되었다. Drasgow 등(2010)은 이상점 모형의 필요성을 지지하는 증거로 단일 카이제곱(single chi-square)의 부적합(badness-of-fit) 통계치를 제시하였지만, Reise(2010)는 이러한 통계치가 추정된 IRT 모수치가 얼마나 잘못된 것인지 또는 그러한 부적합이 야기하는 실용적 결과가 무엇인지에 대해 어떤 의미 있는 기준을 제공하는지가 전혀 명확하지 않다고 비판한다. 부적합이 이상점 반응 과정을 시사한다는 Drasgow 등(2010)의 믿음과 달리 카이제곱 부적합은 다차원성에 의해 초래된 것이거나 잘못된 문항 제작에 기인한 것일 수도 있다는 것이다.

더불어 성격측정에서 이상점 반응 모형을 적용하는 것은 우세 과정 모형을 적용하는 것보다 훨씬 더 복잡하고 많은 자료를 필요로 한다. 따라서 이러한 복잡성이 증가하는 것에 대한 명백한 이점이 있어야 한다. 즉, 중간수준 문항과 이상점 모형의 사용이 우세모형으로 채점된 더 단순한 문항과 비교했을 때, 신뢰도와 타당도 측면에서 실질적인 향상이 있어야 할 것이다(Oswald & Schell, 2010). 지금까지 살펴본 것처럼, 성격측정에서 이상점 적용에 대한 Drasgow 등(2010)의 주장이 받아들여지기 위해서는 더 많은 연구 장면에서 다양한 성격척도들을 대상으로 지속적인 검토가 필요

하다.

성격측정에서 우세 과정과 이상점반응 가정을 비교한 초기 연구(Stark et. al., 2006)는 기존에 우세 과정 가정을 기반으로 개발된 척도(예: 16PF)에 대해 이루어졌다. 문항반응에 대한 가정은 척도 개발시 문항 선택에 대해 영향을 미치기 때문에, 어떠한 가정을 기반으로 문항을 개발했는지에 따라 최종 구성되는 척도는 차이가 있을 수 있고, 따라서 각 가정을 기반으로 개발된 성격척도들을 대상으로 각 모형의 적합성을 비교해볼 필요가 있다. 이러한 관점에서 Chernyshenko 등(2007)이 처음으로 성실성의 하위척도 '질서성(order)'을 고전검사 이론과 우세가정 IRT, 이상점 가정 IRT의 3개 모형을 적용하여 척도를 개발하고 이들을 비교하는 연구를 실시하였다. 국내에서도 성실성 척도에 대한 우세 IRT 모형과 이상점 IRT 모형을 비교하는 연구가 수행되었다(이선희, 2008).

지금까지 성격척도에 대한 이상점 적용을 연구한 Chernyshenko 등(2007)의 연구나 이선희(2008) 연구는 모두 '성실성' 척도에 대한 우세 반응과 이상점반응을 비교하였다. 따라서 이들의 연구에서 나타난 결과들이 다른 성격척도에서도 동일하게 나타나는지를 검증해볼 필요가 있다. Chernyshenko 등(2007)이 연구에서 사용한 '질서(order)' 차원을 살펴보면, 대부분의 사람들이 매우 완벽하게 질서정연하거나, 믿기 힘들 정도로 비조직적이지 않다. 이와 같이 질서성이나 또는 어떤 성격 특성에서 극단적인 것은 제 기능을 못하는 삶이거나 정신병리적인 것을 반영할 수 있다(Markon, Krueger, & Watson, 2005; Oswald & Schell, 2010 개인용). 따라서 이러한 척도들에 응답할 때는 사람들이 문항의 위치가 적정 수준 이상일 때

는 문항에 동의하지만, 그 이상이거나 그 이하의 경우에는 동의하지 않기 때문에 언폴딩 모양이 나타날 가능성이 높다. 그러나 공정성이나 진실성처럼 그 특성이 높을수록 개인이 좋게 평가되는 경향이 있는 척도의 경우에는 개인의 응답이 우세반응을 가질 수도 있다.

또한 이들의 연구에서 나타난 결과들이 다른 연구 장면(예: 문화권에 따른 반응과정 차이)에서도 동일하게 나타나는지를 검증해볼 필요가 있다. 물론 이선희(2008)의 연구가 다른 연구 장면을 제공하기는 하지만, 이 연구에서는 이상점 모형과 우세 모형 각각의 가정을 기반으로 척도를 구성하고 이들 간의 적합도를 비교한 것이 아니라, 연구자가 자체적으로 개발한 동일한 30문항에 대해 두 모형을 비교하는 연구를 실시하였다. 또한 이선희(2008)의 연구에서는 우세반응과 이상점반응 가정의 가장 큰 차이점 중 하나인 '중간 문항'을 포함하지 않았다. 따라서 중간 문항을 포함하여, 각 방법으로 개발된 척도 간에 어떠한 차이가 나타나는지를 검증해 볼 필요성이 있다.

따라서 본 연구에서는 Chernyshenko 등(2007)이 수행한 연구 설계를 다른 성격척도 개발에 적용하여, 이상점반응과 우세반응 가정 중 어떠한 가정이 응답 자료를 설명하는 데 적절한지를 모형 적합도를 바탕으로 비교해보고, 나아가 서로 다른 가정과 방법으로 산출된 3가지 성격점수를 이용하여 사람들의 행동을 예측할 때, 어떤 방식이 관련된 행동을 예측하는 데 더 유용하게 이용될 수 있는지를 비교해 보았다. 이를 위해 본 연구에서는 네 가지 척도를 개발하였다. 먼저 준거타당도 검증을 위해 본 연구에서 준거로 사용되는 학업 성적과 관련성이 높은 '성취지향성'(황지희, 2007;

Elliot & Church, 1997)을 개발하였고, 앞서 살펴본 것처럼, 성격특성 자체가 갖는 특징에 따라 반응 가정이 다르게 나타나는지를 검증하기 위해 일반적으로 사회적 바람직성 경향이 높게 나타날 것이라고 예상되는 공정성, 진실성, 협조성 차원을 포함하였다.

연구 방법

전술한 바와 같이, 본 연구에서는 Chernyshenko 등(2007)의 연구에서 ‘질서’ 척도를 개발하고 모형을 비교한 방법론을 적용하여 공정성, 성취지향성, 진실성, 협조성을 측정하는 성격척도를 3가지 방법(CTT, 우세 IRT, 이상점 IRT)으로 개발하고, 각 방법으로 개발된 결과를 비교하고자 한다. 먼저 문항 개발을 위해 기초적으로 국제 성격 문항집(International Personality Item Pool, IPIP, 2001)을 참고하였다. IPIP는 여러 성격 특성을 측정할 수 있는 문항들이 공개되어 있는 웹사이트의 문항 풀이다. 본 연구에서 관심을 가지는 차원들에 알맞은 문항을 개발하기 위해 이러한 문항 풀을 참고하여 새로운 문항 풀을 개발하였다.

문항 개발

공정성, 성취지향성, 진실성, 협조성 4개 차원에 대한 문항들을 개발하였다. 문항들은 본 연구의 목적을 위해 각 차원의 모든 특성 수준의 범위(긍정 문항, 부정 문항, 중간/중립 문항)를 포괄할 수 있도록 개발되었다.

앞서 설명한 것과 같이 일반적으로 검사 개발 시 중간 범위의 문항들은 문항-총점 상관이 낮게 나타나는 특성 때문에 문항 개발 시

배제되는 경하는 경향이 있었다. 그러나 본 연구에서는 연구의 목적에 따라 문항을 개발하는데 있어 중간 범위의 문항들을 포함시켰다. 중간 범위의 문항들은 해당 차원에서 중간 수준의 특성을 가진 사람들의 행동 및 잠재 특성을 잘 묘사할 수 있도록 개발되었다. 예를 들어, 성취지향의 중간수준 문항은 “성적이거나 성과가 중간 수준은 되어야 만족한다.”와 같이 개발되었다. 본 연구에서는 이러한 문항들을 중간문항이라고 지칭한다.

개발된 문항들이 각 차원의 특성을 잘 나타내는지를 확인하고, 각 문항의 위치(location)를 정확하게 파악하기 위해 산업 및 조직 심리학을 전공하고 있는 7명의 석사 과정 학생들을 대상으로 문항검토 작업을 실시하였다. 먼저 이들에게 각 문항을 제시하고, 개발된 문항들 중 해당 차원에 적절하지 않거나 내용이 중복된다고 생각되는 문항들을 확인하도록 하였다. 그리고 각 문항에 대해 본인이 생각하는 문항의 위치를 5점으로 평가하게 하였다. 문항의 위치가 부정 문항이라고 생각되면 1점에, 긍정 문항이라고 생각되면 5점에, 중간문항이라고 생각되면 3점에 응답하게 하였다. 이를 통해 각 차원에 대해 적절하지 않다고 판단된 문항들과 문항위치에 대한 평가가 학생들 간에 불일치하는 문항들을 제외하였다. 남은 문항들 중 문항위치와 문항 비율을 고려하여 최종 문항을 선택하였다.

그 결과 각 차원 당 부정 문항 10개, 중간 문항 10개, 긍정 문항 10개로 총 30개 문항씩 선택되었으며, 4개 차원 120개 문항을 척도 개발에 사용하였다. 표 1은 최종 문항개발을 위해 선택된 성취지향 척도의 문항 풀이다(공정성, 진실성, 협조성에 대한 개발문항은 부록에 제시).

참가자 및 자료 수집

본 연구의 참가자는 한 대학교의 심리학 전공 및 교양 과목 수강생 310명과 일반인 383명 총 693명이었다. 참가자들의 응답은 온라인과 오프라인을 통하여 수집되었으며, 불성실한 응답은 표본에서 제거되었다. 응답자 분포는 남성 379명(54.7%), 여성 314명(45.3%)이었고, 20대가 571명(82.3%)로 가장 많았으며, 30대 101명(14.6%), 40대 이후가 21명(3.1%)이었다.

응답자들은 각 차원 당 30개 문항 총 120개 문항에 대해, 각 문항이 자신의 행동이나 성격 특성을 얼마나 잘 나타내는지에 대해 4점 척도(1=전혀 그렇지 않다에서 4=매우 그렇다)로 응답하였다. 전통적 CTT에서는 다분문항 반응자료가 이분문항 반응자료에 비해 정규성 가정을 덜 위반하고, 더 정확한 문항-총점 상관(item-total correlation)을 산출해 주기 때문에(Chernyshenko 등, 2007), 본 연구에서도 CTT에서는 다분 응답 자료를 사용하였다. IRT 측면에서도 다분 문항 반응의 결과가 제공해주는 장점이 많지만 다분 문항 분석을 위해 필요한 표본의 크기가 너무 크기 때문에, Chernyshenko 등(2007)의 연구에서와 같이 본 연구에서도 응답 자료를 이분화 하여 사용하였다.

척도 개발

본 연구의 목적은 성격측정에서 우세과정과 이상점반응 과정을 비교하기 위한 것으로, 구체적으로는 각 가정을 바탕으로 하는 고전검사이론(CTT), 우세 IRT, 이상점 IRT 척도화(scaling) 방법에 따라 구성되는 척도의 결과를

직접적으로 비교하고자 하는 것이다. 이를 위해 각 차원 별로 앞서 개발한 동일한 30개 문항 풀에 3가지의 방법을 적용하여 척도 개발을 실시하였다. 우세 가정을 통해 개발되는 척도는 'CTT' 척도와 '우세 IRT' 척도로 칭하고, 이상점 가정을 통해 개발되는 척도는 '이상점 IRT' 척도로 칭하였다.

CTT 척도 구성

전통적 CTT 척도를 개발하기 위하여 성격 척도 개발 시 사용되는 표준적인 Likert 접근 방식에 따라 개발하였다. 먼저 부정적으로 작성된 부정 문항들은 역산하여 정적인 문항 총점 상관이 나오도록 하였다. 중간 문항들은 긍정 문항들의 총점과의 상관 분석을 통해 역산 여부를 결정하였다. 긍정 문항과 정(+)의 상관이 나오는 문항들은 긍정 문항으로, 부(-)의 상관이 나오는 문항들은 부정 문항으로 간주하였다. 그 다음 모든 문항들에 대해서 문항-총점 상관을 산출하고, 이를 바탕으로 상위 15개 문항을 선정하였다. 그 후, 긍정문항과 부정문항의 균형을 맞추기 위해 일부 문항을 교체하였다. 예를 들어, 상위 15개 문항 중에서 9개 문항이 긍정문항이고, 6개 문항이 부정문항이라면, 9개 긍정문항 중 가장 낮은 문항-총점 상관을 가진 1개 문항을 제외하고, 대신 부정문항 중 6개 문항다음으로 문항-총점 상관이 높은 문항 1개를 포함하였다. 마지막으로 최종 15개 문항 척도의 일차원성을 확인하기 위해 주축 요인 분석을 실시하였다.

우세 IRT 척도 구성

우세 IRT 척도를 구성하기 위하여 2PLM(2 parameter logistic model)에 근거하여 척도를 구성하였다. 2PLM에서, 문항 i 에 동의할 가능성은

표 1. 성취지향 30개 문항

No.	문항내용	ITC*	요인부하량
성취01 ^a	성취하기 어려운 목표를 세우지 않는다.	.34	.36
성취02 ^a	시련과 고난을 겪으면서까지 성공하고 싶지는 않다.	.55	.59
성취03 ^a	어려운 문제에 직면하면 쉽게 포기한다.	.59	.64
성취04 ^a	계획을 세울 때마다 작심삼일이다.	.48	.53
성취05 ^a	과제나 업무에 지나칠 정도의 시간과 노력을 투자하지 않는다.	.50	.52
성취06 ^a	무언가를 희생하면서까지 성공하고 싶지는 않다.	.50	.55
성취07 ^a	완벽할 때까지 일을 하는 사람을 보면 이해가 가지 않는다.	.45	.47
성취08 ^a	해내기 어려운 일은 일찍 포기하는 것이 현명하다	.48	.50
성취09 ^a	목표를 달성하기 위해 스트레스를 받으면서까지 노력하지 않는다.	.46	.50
성취10 ^a	성공할 가능성이 낮은 일에 매달리지 않는다.	.33	.34
성취11 ^a	달성할 수 있을 정도의 목표만 세운다.	.42	.43
성취12 ^a	커트라인이 있는 시험은 통과할 수 있는 정도로 공부한다.	.33	.33
성취13 ^a	성적이나 성과가 중간 수준은 되어야 만족한다.	.14	.11
성취14 ^a	나의 성취지향은 평균 수준이다.	.44	.46
성취15 ^a	목표가 있긴 하지만 그 목표만을 위해 살지는 않는다.	.28	.28
성취16 ^a	다른 사람들과 비슷한 정도의 평가는 받아야 한다.	.11	.08
성취17 ^a	노력해서 중간정도만 가면 만족한다.	.54	.58
성취18 ^a	실패하지 않을 정도의 노력만 한다.	.59	.63
성취19 ^a	무엇이든 남들만큼만 하면 된다.	.49	.52
성취20 ^a	성공을 위해 어느 정도 노력하지만, 그것이 내 삶의 전부는 아니다.	.32	.32
성취21	어려운 목표에 도전하는 것이 좋다.	.59	.64
성취22	원하는 것은 꼭 이뤄야 직성이 풀린다.	.30	.32
성취23	더 높은 위치에 오르기 위해 노력한다.	.56	.62
성취24	목표는 반드시 이루고 만다.	.59	.65
성취25	할 수 있는 한 모든 것을 성취하기 위해 노력한다.	.51	.56
성취26	목표한 것 이상을 해낸다.	.52	.57
성취27	다른 사람의 업적을 뛰어넘기 위해 노력한다.	.52	.57
성취28	어느 시험이든 만점을 목표로 공부한다.	.48	.53
성취29	불가능해 보이는 목표라도 끊임없이 도전한다.	.58	.62
성취30	목표한 것을 달성하기 위해서라면 며칠쯤 밤을 셀 수 있다.	.54	.57

N=693. * ITC=item-total correlation

^a역문항(11번에서 20번은 중간수준 문항)

$$P(u_i = 1|\theta = t) = \frac{1}{1 + \exp[-1.7a_i(t - b_i)]}$$

로 나타나며, a_i 는 문항 i 의 변별모수(즉, 문항 변별도), b_i 는 문항의 i 의 위치모수(즉, 문항곤란도)이다. 다분(4점)으로 된 본 연구의 응답 자료는 2PLM에 적용하기 위해 이분으로 변경되었다. CTT와 같이 부정 문항과 중간문항 중 부정 문항으로 나타난 문항들은 역산하였고, 2PLM 문항 모수는 BILOG 프로그램(Mislevy & Bock, 1991)을 활용하여 산출하였다.

추정된 모형들의 적합도는 Drasgow, Levine, Thissen, Williams 및 Mead(1995)가 제안한 카이 자승(chi-square; χ^2) 통계량을 사용하여 살펴보았다(모형-자료 적합도 검증을 위해 산출되는 개별 문항, 두 문항 쌍, 세 문항 쌍 적합도에 대한 자세한 설명은 Tay et. al., 2011의 연구 참고). 카이자승 적합도 통계치는 MODFIT (Stark, 2008) 프로그램을 활용하여 산출되었고, 일차적으로 개별 문항에서 적합도가 좋지 않은 문항(χ^2/df 3 이상; Zickar & Drasgow, 1996)과 변별도가 .4 미만인 문항들은 제거되었다. 다음으로 남은 문항들 중 변별도가 높은 문항들을 기준으로, 각 차원에서 문항들이 측정하고자 하는 특성의 정보를 골고루 반영할 수 있도록 문항을 선택하였다. 그리고 지역 독립성을 검증하기 위하여 두 문항 쌍(doublets)과 세 문항 쌍(triplets)의 카이 자승 적합도 통

계치를 확인하였고, 지역 독립성을 위배하는 문항들을 교체하였다. 또한 두 문항 쌍 및 세 문항 쌍의 낮은 카이 자승 값은 단일차원성(unidimensionality)의 근거로 해석되기(Chernyshenko 등, 2007) 때문에 높은 카이 자승 값을 가지는 문항들은 교체되는 것이 타당하다. 마지막으로 긍정 문항과 부정 문항의 비율을 고려하여 최종 15개 문항을 선택하였다.

이상점 IRT 척도 구성

이상점 IRT 척도는 GGUM(generalized graded unfolding model)에 근거하여 구성되었다. 이분 자료에 대한 GGUM의 공식(세부 내용은 Roberts, Donoghue, & Laughlin, 2000 참고)은 다음과 같다.

$$P(U_i = 1|\theta_j) = \frac{\exp(\alpha_i[(\theta_j - \delta_i) - \tau_{i1}]) + \exp(\alpha_i[(2\theta_j - \delta_i) - \tau_{i1}])}{1 + \exp(\alpha_i[3(\theta_j - \delta_i)]) + \exp(\alpha_i[(\theta_j - \delta_i) - \tau_{i1}]) + \exp(\alpha_i[2(\theta_j - \delta_i) - \tau_{i1}])_i}$$

이 공식에서 θ_j 는 연속선상 기저 반응에 대한 응답자 j 의 위치, α_i 는 문항 i 의 변별모수, δ_i 는 연속선상 기저 반응에 대한 문항 i 의 위치모수, 그리고 τ_{i1} 는 잠재 연속선에 대한 내면적(subjective) 반응 범주의 역치(threshold)이다.

본 연구에서는 다분(4점)으로 되어있는 응답 자료를 이분으로 변환하고, GGUM2004(Roberts, 2004) 프로그램을 사용하여 GGUM 모수를 산출하였다. 우세 IRT 척도 구성과 마찬가지로 모형 적합도 지수를 산출하기 위하여 MODFIT을 사용하였다. MODFIT 프로그램을 통해 일차적으로 단일문항에 대해 적합도가 좋지 않은 문항(χ^2/df 3 이상)과 변별도가 .4 미만인 문항들은 제거되었다. 다음으로 두 문항 쌍과

2) Zickar & Drasgow(1996)는 조정된 χ^2/df 비율이 1보다 작으면 '매우 작다', 1-2 사이이면 '작다', 2-3 사이이면 '약간 크다', 3이상이면 '크다'로 해석할 수 있다고 제안하였고, Chernyshenko 등(2007)은 이를 기준으로 척도 개발 시 개별 문항에서 χ^2/df 비율이 3이상으로 나타나는 문항들을 우선적으로 제거하였다

세 문항 쌍의 카이 자승 적합도 통계치를 통해 지역 독립성을 위배하는 문항들을 제거하였다. 남은 문항들 중, 변별도가 높은 문항들을 기준으로 문항을 선택하였다. 또한 부정 문항, 중간 문항, 긍정 문항의 비율을 맞추기 위해 일부 문항을 교체하였다. 이러한 과정을 통해 최종적으로 각 차원 당 15개 문항을 선택하였다.

준거타당도 분석

세 척도화 방법에 따라 구성된 성격척도들의 준거타당도를 비교해보기 위해 본 연구에 참여한 대학교의 심리학 전공과목 1개 반 109명의 성적 자료를 수집하였다. 성적자료는 중간고사, 기말고사, 과제 그리고 출석을 포함한 것으로 100점 만점을 기준으로 채점되었다.

준거 상관 분석

각 척도 구성 방법들의 예측타당도를 비교하기 위해 각 모형으로 구성된 척도와 학업 성적과의 상관계수를 비교하였다. CTT 방법은 최종 선택된 15개 문항의 개인 총점과 학업 성적 간의 상관을 산출하였고, 우세 및 이상점 IRT는 각각 15문항으로 산출된 개인의 능력모수와 학업성적 간의 상관을 산출하였다.

순위 비교

Dalal, Withrow, Gibby, 및 Zickar(2010)는 각 모형을 통해 구성된 성격척도들이 고용 결정에 미치는 영향을 검증할 때 오직 상관 계수에만 초점을 맞추면 안 된다고 주장했다. Ross, Stecher, Miller, & Levin(1998)은 상관 계수는 변인들 간의 선형적 관계에 대한 가정에 영향을 덜 받는 추정치라는 것을 보인 Drasgow와

Kang(1984)의 연구결과에 근거하여, 상관계수의 이러한 경향성은 이변량 분포의 특정 범위에서의 서열 순위 변화에 민감하지 못한 결과를 낳는다고 했다(이은정, 박동건, 2003). 즉 상관 분석은 모수 추정 방식에 의해 생길 수 있는 지원자들의 서열 순위 변동에 민감하게 반응하지 못할 수 있다는 것이다. 따라서 본 연구에서는 각 척도 구성 방법에 따라 선발 및 평가 상황에 어떤 영향을 주는지 알아보기 위해 세 가지 방법을 통해 추출된 문항들로 나타난 개인들의 점수에 순위를 매겼다. 그 후 각 방법들 간의 순위 상관을 살펴보고, 나아가 각 척도화 방법에서의 개인 특성 순위를 비교해 보고자 하였다.

연구 결과

CTT 척도 개발 결과

CTT 척도 개발을 위해 먼저 4개 차원의 각 30개 문항 풀에서 Nunnally와 Bernstein(1994)이 척도 내 문항 유지의 기준치로 제시한 .3 이하의 문항-총점 상관을 나타낸 문항들을 제거하였다. 각 차원 별로 제거된 문항의 수는 성취지향 3개, 공정성 4개, 진실성 4개, 협조성 4개이다. 제거된 문항들은 진실성의 1개 문항을 제외하고, 모두 문항위치 평가에서 특성 수준 연속선상의 중간범위에 위치한다고 평가된 문항들이었다. 제거된 문항 예로는 성취지향을 측정하는 “성적이나 성과가 중간 수준은 되어야 만족한다.”와 같은 문항이 있었다. 이는 이상점 반응 척도화 방법에서는 포함될 수도 있는 중립적인 문항들이 CTT 기반 척도 개발방법에서 낮은 문항-총점 상관에 의해 제

거될 수 있다는 Stark 등(2006)의 주장을 어느 정도 지지할 수 있는 결과라고도 볼 수 있다. 그러나 각 차원에서 개발된 중간문항은 10개 씩으로 각 차원에서 모두 절반 이상의 문항들이 .3이상의 문항-총점을 보였다. Chernyshenko 등(2007)의 연구에서도 이와 비슷한 결과를 볼 수 있는데, 그들이 중간문항이라고 개발한 13개 문항 중 7개 문항만이 .3미만의 문항-총점 상관을 보였다. 따라서 이러한 결과들을 보면 CTT 척도화 기법이 중간문항에 대한 낮은 문항-총점 상관을 도출하도록 한다는 Stark 등(2006)의 주장이 명백하게 지지된다고 말하기는 어렵다.

문항-총점 상관을 통한 문항 제거 후, 높은

표 2. CTT 방법으로 개발된 성취지향 통계치

문항 No.	평균	표준 편차	ITC*	요인 부하량
성취02	3.00	.75	.54	.60
성취03	3.13	.66	.58	.61
성취05	2.76	.68	.47	.50
성취06	2.78	.70	.49	.54
성취17	3.00	.66	.50	.55
성취18	3.02	.68	.58	.62
성취19	2.87	.62	.45	.49
성취21	2.93	.68	.63	.67
성취23	3.08	.58	.62	.66
성취24	2.85	.67	.63	.68
성취25	3.13	.68	.54	.58
성취26	2.89	.62	.56	.60
성취27	2.76	.72	.53	.57
성취29	2.96	.79	.55	.62
성취30	2.83	.75	.58	.58

*ITC=item-total correlation

문항-총점 상관을 기준으로 긍정문항과 부정문항의 비율을 고려하여 각 15개의 문항을 선별하였다. 각 차원 별 척도의 평균(표준편차)은 성취지향성 43.99 (6.36), 공정성 43.62(6.97), 진실성 41.41(6.11), 협조성이 44.65(7.01)였다. 내적일관성 신뢰도는 성취지향이 .89, 공정성은 .90이고, 진실성은 .87, 그리고 협조성은 .92였다.

표 2는 CTT 척도화를 통해 최종 개발된 성취지향 척도의 산술통계와 문항-총점 상관, 요인 부하량을 보여준다. 성취지향 척도의 문항-총점 상관은 .47에서 .63사이(평균 .55)로 높은 것으로 나타났다. 공정성은 .49에서 .65(평균 .58), 진실성은 .28에서 .66(평균 .51), 그리고 협조성은 .47에서 .69(평균 .62)로 모두 높은 수준으로 나타났다.

우세 IRT 척도 개발 결과

부정 문항을 역산한 30개 문항 자료를 이분화하고, BILOG 3.0으로 2PLM 문항 모수 추정을 실시하였다. 우선적으로 낮은 변별도 모수(a값이 .4 이하)를 나타낸 문항을 제거하였다. 낮은 변별도 모수를 갖는 문항은 정보를 적게 제공하기 때문에, 측정오차를 줄이는데 큰 기여를 하지 못한다(Chernyshenko et, al., 2007). 낮은 변별도로 인해 제거된 문항은 성취지향 3개, 공정성 3개, 진실성 6개, 협조성 4개이다. 제거된 문항 중 중간문항의 개수는 성취지향 2개, 공정성 3개, 진실성 2개, 협조성 4개였다.

다음으로 MODFIT으로 산출한 교정된 개별 문항 χ^2/df 값으로 적합성이 낮은 문항(3이상)을 식별하여 제거하려고 하였으나, 모든 차원의 모두 문항이 1이하의 좋은 적합도를 갖는 것으로 나타났다. 두 문항 쌍과 세 문항 쌍에

대한 교정된 χ^2/df 값을 통해 지역독립성 가정을 위배하는 문항들의 쌍을 확인하여 제거하였다. 예를 들어, 성취지향 18번 문항 “실패하지 않을 정도의 노력만 한다.”와 19번 문항 “무엇이든지 남들만큼만 하면 된다.”는 모두 높은 변별도를 가지지만, 두 문항 쌍에서 카이제곱 적합도 비율 값이 23.17로 나타났다. 이는 두 문항의 내용이 서로 비슷하기 때문에 지역독립성을 위배하는 것으로 나타나 18번 문항을 제거하였다.

마지막으로 각 차원에서 남은 문항들 중 변별도가 높은 순으로 최종 15문항을 선택하였고, 긍정문항과 부정문항의 비율을 맞추기 위해 일부 문항을 교체하였다. 선정된 15개 문항에 대한 모형 적합도 결과에서는 모든 차원에서 개별 문항의 경우에는 15개 문항이 모두 1이하로 나타났다. 두 문항 쌍은 성취지향 .87, 공정성 .12, 진실성 1.07, 협조성 .25로 나타나 진실성을 제외하고 모두 1이하로 나타났다. 세 문항 쌍의 경우에도 성취지향 1.44, 공정성 .22, 진실성 1.77, 협조성 .38로 공정성과 진실성은 1이하로 매우 낮은 수준으로 나타났고, 성취지향과 진실성도 2이하로 만족할만한 수준이었다.

성격과 같은 비인지적인 영역을 측정하는 검사의 척도를 구성하는데 있어서 지역독립성을 고려하는 것은 CTT와는 다른 IRT의 가장 중요한 특징이다. CTT는 내적 일관성 신뢰도가 높은 척도를 만들기 때문에, 문항 내용이 비슷한 문항들이 선택되기 쉽다. 인지적 영역에서는 이러한 것들이 심리측정적으로 많은 문제가 되지 않지만, 비인지적 영역에서는 비슷한 문항이 반복되는 것은 응답자로 하여금 서로 비슷한 응답을 하도록 유도하거나 기억 효과(memory effect)를 유발하여 지역독립성을

해칠 수 있다. IRT에서는 문항개발 단계에서 지역독립성에 문제가 되는 문항들을 식별할 수 있고, 이들을 제거하거나 통합할 수 있다. 이러한 이유로, CTT와 우세 IRT 척도가 동일한 우세가정을 기반으로 함에도 불구하고 공통으로 선택된 문항이 서로 달라진다 (Chernyshenko 등, 2007). 본 연구에서도 이러한 과정에 따라 CTT와 우세 IRT로 개발된 척도 간에 약간의 차이가 있었다. 각 차원별로 CTT와 우세 IRT 간에 공통된 문항 개수는 성취지향 10개, 공정성 8개, 진실성 13개, 협조성 9개였다.

표 3은 성취지향 차원에서 우세 IRT 방법으로 선정된 문항내용과 문항모수이다. 각 차원별 변별도 모수의 평균은 성실성 .90, 공정성

표 3. 우세 IRT로 개발된 성취지향 문항모수

문항 No.	2PLM 문항 모수	
	a(변별도)	b(곤란도)
성취02	.99	-1.07
성취03	1.25	-1.44
성취05	.88	-.70
성취07	.70	-1.92
성취08	.70	-1.50
성취09	.91	-1.29
성취12	.97	-1.24
성취19	.63	-1.23
성취21	1.00	-1.04
성취23	1.10	-1.71
성취24	.98	-.83
성취25	.88	-1.53
성취28	.61	-.15
성취29	1.02	-.66
성취30	.95	-1.06



그림 3A. 부정문항의 2PLM 문항반응함수

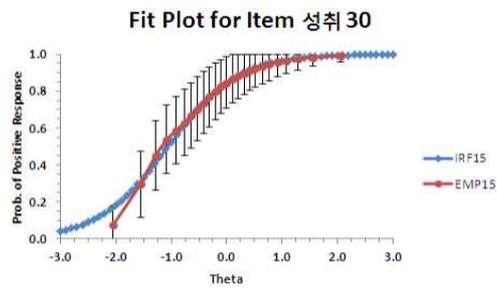


그림 3B. 긍정문항의 2PLM 문항반응함수

.82, 진실성 .97, 협조성 1.63으로 나타났다. 이는 선행연구(Chernyshenko et. al., 2007)에서 개발했던 질서성(.82)과 비슷한 수준이다. 그러나 본 연구의 참여자가 사회적 바람직성을 나타낼 필요가 없는 표본이었음에도 성취지향의 모든 문항의 곤란도 모수가 (-) 값을 가지는 것으로 나타났다. 이는 성취지향 차원의 특성을 적게 가지고 있는 사람들도 각 문항에 긍정적으로 응답할 가능성이 높은 것을 의미한다. 즉 성취지향 차원에서는 전체적인 문항에 대한 동의도가 높다고 할 수 있다. 협조성, 공정성, 진실성 차원에서도 곤란도 모수가 대부분 (-) 값을 가지는 것으로 나타났다.

Dragow 등(2010)은 우세 IRT 모형은 위치모수와 문항내용이 직접적인 관련이 없다고 비판하는 데, 본 연구결과에서도 이러한 점을 볼 수 있다. 예를 들어, 그림 3A와 3B는 부정문항(성취02)과 긍정문항(성취30)의 IRF를 나타내는데, 그림을 보면 두 문항 모두 단조 증가하는 형태를 보이고 있는 것을 알 수 있다. 하나는 부정 문항이고 다른 하나는 긍정 문항이지만, 두 문항의 문항모수가 매우 유사하기 때문에 IRF 또한 매우 유사한 형태를 띠고 있다. 부정문항은 역산되기 때문에 IRF는 감소하는 모양이 아닌 증가하는 모양을 나타내게 된다. 위치모수는 문항에 동의하거나 동의하지

않을 확률이 .5인 지점이기 때문에, 역산은 IRF의 위치에는 영향을 미치지 않는다. 따라서 유사한 모수를 가진 역산 문항과 그렇지 않은 문항의 2PLM IRF는 문항 위치를 중심으로 대칭모양을 보이기 때문에, IRF의 모양이 매우 유사해지게 된다(Chernyshenko 등, 2007).

그림 4는 우세 IRT로 개발된 성취지향 차원의 검사정보함수(Test Information Function; TIF)이다. 그림을 보면 이 척도는 특성 연속선상에서 낮은 범위(-1.5에서 0.5)에 대한 정보가 많고, 극단부분에 대한 정보는 매우 적은 것을 볼 수 있다. 본 연구에서 개발된 성취지향에서 정적인 위치모수($b > 0$)를 가진 문항이 없었기 때문에, 높은 특성 수준에서 검사정보가 낮게 나타났다. Chernyshenko 등(2007)의 연구에서는 본 연구보다 약간 높은 -1에서 0 사이의 특성 수준에서 가장 정보가 높은 것으로 나타

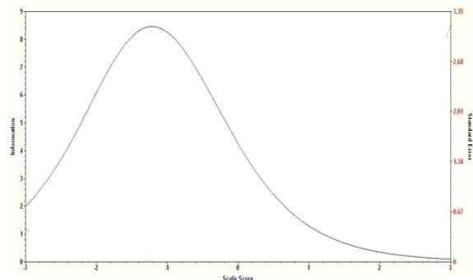


그림 4. 성취지향 우세 IRT 검사정보함수

났다. 특성연속선상에서 특정 범위에서의 정보가 높다는 것은 그 범위에서 가장 낮은 측정오차를 가진다는 것을 의미하고, 이것은 그 범위의 특성 수준을 가진 사람들을 가장 잘 변별해준다는 것을 의미한다.

Chernyshenko 등(2007)은 -1과 +1사이에서 검사정보가 높은(즉 낮은 측정오차를 가지는) 척도들이 선발장면의 스크리닝(screening)에 적합하고, 양 극단의 측정에 관심을 갖는 임상 및 상담 장면의 검사로서는 적절하지 못하다고 제안하였다. 이러한 측면에서 본다면, 산업 및 조직심리학에서는 주로 인사선발을 위한 성격검사를 사용하고 있기 때문에 Chernyshenko 등(2007)이 주장하는 것과 같이 성격검사에서 이상점 반응을 적용하는 것이 꼭 필요한 것인지에 대한 추가적인 의문이 발생한다. 더욱이 성격검사를 최종 합격 결정 단계가 아닌 면접대상자를 선별하기 위한 초기 스크리닝 단계에서 많이 사용하는 국내 선발 장면에서는 극단수준 보다 중간수준이나 그 이하에서 변별성이 높은 검사가 더 활용가치가 있을 것이다.

이상점 IRT 척도 개발 결과

각 차원 별로 초기 개발된 30개 문항에 대한 응답 자료를 이분화하고, GGUM 2004를 사용하여 모수추정을 하였다. 먼저 추정된 문항 모수에서 낮은 변별도(.4 이하)를 갖는 것으로 확인된 문항을 제거하였다. 성취지향, 공정성, 협조성에서는 .4이하의 낮은 변별도를 갖는 문항이 없었고, 진실성에서는 2개 문항이 제거되었다. MODFIT으로 산출한 개별 문항의 교정된 χ^2/df 값으로 적합성이 낮은 문항을 제거하려고 하였으나, 4개 차원의 모든 문항이 1 이하의 좋은 적합도를 갖는 것으로 나타났다.

두 문항 쌍과 세 문항 쌍에 대한 교정된 χ^2/df 값을 통해 일차원성 및 지역독립성 가정을 위배하는 문항들의 쌍을 확인하여 제거하였다. 마지막으로 긍정 문항, 중간 문항, 부정 문항의 비율을 맞추기 위해 일부 문항을 교체하여 각 차원별로 최종 15문항이 선정되었다. 각 차원에서 개발된 척도들에 대한 모형-자료 적합도 값을 살펴보면, 개별 문항은 4차원 모두에서 1이하로 매우 좋은 것으로 나타났다. 두 문항 쌍도 성취지향 .60, 공정성 .17, 진실성 .95, 협조성 .16으로 모두 1이하로 나타났다. 세 문항 쌍은 성취지향 .86, 공정성 .27, 진실성 1.47, 협조성 .25로 진실성을 제외하고 모두 1이하로 나타났으며, 진실성도 2이하로 만족할만한 수준을 보였다.

표 4는 이상점 IRT 척도의 GGUM 모수를 보여준다. 선정된 15개 문항 중에서 8개의 문항은 문항반응곡선에서 비단조성을 나타내었다. 예를 들어, 비단조성을 보인 20번 문항(“성공을 위해 어느 정도 노력하지만, 그것이 내 삶의 전부는 아니다.”)의 GGUM 모수는 $\alpha=1.75$, $\delta=-.68$, $\tau=-2.21$ 로 나타났으며, 응답 반응곡선은 그림 5와 같다.

그림 5를 보면, 평균수준의 성취지향 특성을 가진 응답자는 이 문항에 동의할 가능성이 높고, 성취지향이 매우 높거나 낮은 응답자는 이 문항에 동의하지 않을 가능성이 높다. 따라서 Chernyshenko 등(2007)은 이러한 문항들의 문항-총점 간 상관은 거의 0에 가까워져 CTT와 우세 IRT 척도에서는 제거되는 문항이 되며, 비단조성을 띄지 않는 문항은 상대적으로 극단적인 위치모수를 가진다고 주장하였다. 하지만 본 연구에서는 Chernyshenko 등(2007)의 주장과는 다소 일치하지 않는 결과가 산출되었다. 예를 들어, 성취지향 차원에서는 비단조

표 4. 이상점 IRT로 개발된 성취지향 문항모수

문항	GGUM 문항 모수		
	α (변별도)	δ (곤란도)	τ
성취03	2.49	-4.08	-2.72
성취05	1.51	-3.52	-2.82
성취07	1.20	-5.03	-3.11
성취08	1.14	-4.37	-2.81
성취09	1.41	-4.16	-2.79
성취12*	.74	-2.30	-1.87
성취13*	.64	-.60	-2.48
성취15*	1.38	-.88	-2.11
성취18	1.67	-3.95	-2.72
성취20*	1.75	-.68	-2.21
성취21*	1.62	.93	-2.00
성취23	1.75	2.40	-4.16
성취26*	1.37	.97	-2.14
성취28*	1.14	1.06	-1.16
성취30*	1.51	1.22	-2.29

*언폴딩(unfolding) 문항반응곡선을 가지는 문항

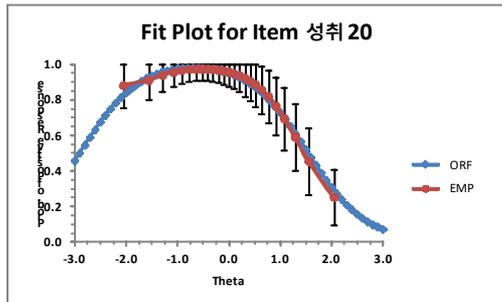


그림 5. 성취20번 문항의 GGUM 문항반응곡선

성을 나타내는 8개 문항 중에서 2개 문항(13번, 15번)만 .3 이하의 문항-총점 상관을 보였고, 나머지 6개 문항은 모두 .3이상의 문항-총점 상관을 나타냈다. 또한 비단조성을 보이는

문항들 중 CTT에서 3개 문항이, 우세 IRT에서는 4개 문항이 선정되었다.

더불어, 이상점 IRT 척도화는 문항내용과 위치모수(δ)가 직접적으로 관련된다는 Drasgow 등(2010)의 주장도 본 연구에서는 다르게 나타났다. 표 4를 살펴보면, 이상점 IRT 방법을 통해 구성된 문항내용과 위치모수(δ)가 모두 일치하지는 않는 것을 볼 수 있다. 예를 들어, 18번 문항(“커트라인이 있는 시험은 통과할 수 있는 정도로 공부한다.”)은 중간문항으로 개발되었지만 -2.30의 높은 부적 위치모수를 갖는 것으로 나타났다. 또한 21번 문항(“목표한 것 이상을 해낸다.”)은 매우 높은 성취지향성을 나타내는 문항으로 개발되었으나 위치모수가 .93정도로 중간보다 약간 높은 정도의 성취지향을 나타내었다. 이는 문항개발상의 문제로 인해 나타난 결과일 수도 있다. 그러나 본 연구에서는 각 특성범위 수준에 적합한 문항 개발을 위해 여러 사람들의 의견을 수렴하여 특성수준을 결정하였다. 따라서 이러한 결과는 전 범위의 특성수준을 나타내는 문항(특히 중간수준의 문항)을 개발하는 것이 현실적으로 매우 어려운 일임을 보여준다.

또한 본 연구에서 개발한 4개 차원 중 성취지향 차원은 비단조성을 보이는 문항이 절반 이상으로 나타났지만, 다른 3개 차원에서는 실제로 비단조성을 보이는 문항이 공정성 5개, 진실성 5개, 협조성 3개로 다소 적었다. 이는 앞서 서론에서 설명한 것과 같이 척도 특성에 따라 비단조성이 더 잘 나타나는 성격특성이 있는지에 대한 논의거리를 제공하는 결과라고 볼 수 있다.

그림 6은 성취지향의 이상점 IRT의 TIF를 나타낸다. 앞서 제시한 우세 IRT의 TIF와 비교해보면, 부정적으로 극단에서부터 중간수준 위

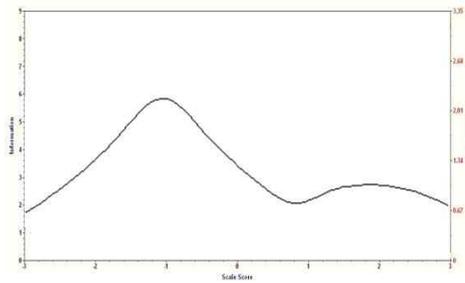


그림 6. 성취지향 이상점 IRT 검사정보함수

치까지는 우세 가정의 TIF가 이상점 가정의 TIF보다 높은 것으로 나타났으며, 정적으로 극단적인 부분에서 이상점 가정이 우세 가정에 비해 더 높게 나타났다. 이는 공정성 차원에서 동일하게 나타났다. 반면에 진실성과 협조성 차원에서는 우세 가정 및 이상점 가정의 TIF 간 차이가 거의 없는 것으로 나타났다. 본 연구에서 나타난 이러한 TIF의 경향은 우세반응 가정 TIF는 중간수준(-1과 +1사이)에서 가장 높은 정보함수를 제공하고, 이상점반응 가정에서는 전 범위에서 높은 정보함수를 제공할 것이라는 Chernyshenko 등(2007)의 주장과는 다르게 나타났다. 실제 이들의 연구에서도 이상점 TIF 0이하의 낮은 특성 수준에서만 높은 검사정보를 제공하고 있고 0이상의 특성 수준에서는 우세 가정과 차이가 없었다.

지금까지 4개 차원을 각 3가지 척도화 방법으로 구성된 척도 결과를 살펴보았다. 본 연구의 궁극적 목적은 성격측정에서 우세반응 가정과 이상점 반응 가정을 비교하는 것이었다. 기본적으로 모형-자료 적합도 검증 비교를 통해 어떤 모형이 성격자료에 더 적합한지를 검증하게 되므로, 적합도 통계치를 제공할 수 없는 CIT를 제외하고, 우세 IRT 모형과 이상점 IRT 모형의 적합도를 비교해보고자 한다

표 5는 각 차원별 우세 및 이상점 IRT의 카

이자승 적합도 지수의 평균과 표준편차를 나타낸다. 결과를 보면, 성취지향, 진실성, 협조성에서는 이상점 IRT의 적합도가 우세 IRT의 적합도 보다 약간 더 좋은 것으로 나타났다. 반면 공정성에서는 우세 IRT의 적합도가 약간 더 좋은 것으로 나타났다. 그러나 적합도의 차이를 비교하기 위한 명확한 기준이 없고, 개별 문항에서는 우세와 이상점 IRT 모두 1이하의 매우 좋은 적합도를 갖는 것으로 나타났기 때문에, 이러한 결과를 통해 어떤 척도화 방법의 적합도가 더 좋다고 명확한 결론을 내리기는 어렵다. 또한 Drasgow 등(2010)이 주장한 것과 같이 중간문항과 같은 비단조성 문항을 포함하는 것이 반드시 모형 적합도를 향상시키지도 않았다. 협조성과 공정성 척도는 모든 특성 수준(공정, 중간, 부정)의 비율을 맞추기 위해 중간문항을 의도적으로 포함시켰을

표 5. 각 차원별 카이자승 적합도 평균(표준편차)

	성취지향	공정성	진실성	협조성
<u>우세 IRT</u>				
개별 문항	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
두 문항 쌍	.87 (3.67)	.12 (.84)	1.07 (2.50)	.25 (1.05)
세 문항 쌍	1.44 (3.31)	.22 (.83)	1.77 (2.59)	.38 (1.12)
<u>이상점 IRT</u>				
개별 문항	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
두 문항 쌍	.60 (3.76)	.17 (1.01)	.95 (2.39)	.16 (.61)
세 문항 쌍	.86 (2.70)	.27 (1.05)	1.47 (2.40)	.25 (.75)

때 오히려 적합도가 낮아지는 경향이 있었다. 따라서 본 연구결과만을 통해 성격측정에서 어떤 모형의 사용이 옳은지에 대한 명확한 결론을 도출할 수 없다.

준거 타당도 검증

준거상관 분석

각 척도화 방법의 예측타당도를 비교하기 위하여 3가지 방법으로 구성된 척도에 대한 개인점수(CTT 총점, 우세 IRT의 개인능력모수, 이상점 IRT의 개인능력모수)와 학업성적(준거) 간의 상관분석을 실시하였다. 그 결과, 성취지향을 포함한 모든 차원들에서 학업성적과의 상관은 유의미하지 않았다. 성취지향의 경우 유의미하지는 않았지만 CTT와 이상점 IRT는 비슷한 상관을 나타냈고, 상대적으로 우세 IRT와의 상관이 다소 낮았다. 또한 모든 차원에서 우세가정을 기반으로 한 CTT 총점과 우세 IRT, 이상점 가정을 기반으로 한 이상점 IRT로 산출된 개인들의 능력 값들의 상관은 높은 것으로 나타났다.

순위 비교

선발장면에서는 지원자들의 순위가 선발에

표 6. 성취지향 척도와 학업성적 간 상관계수

	N	1	2	3
1. CTT 총점	693			
2. 우세 IRT ^a	693	.85***		
3. 이상점 IRT ^b	693	.77***	.85***	
4. 학업성적	109	.10	.06	.10

*** $p < .001$

^a우세 모형에서 산출된 성취지향 개인특성 값

^b이상점 모형에서 산출된 성취지향 개인특성 값

영향을 미치므로 세 척도화 방법에서 산출된 개인들의 순위를 비교해 보았다. 먼저 순위 상관을 살펴보았는데, CTT와 우세 IRT는 .84 ($p < .001$), CTT와 이상점은 .77($p < .001$), 우세 IRT와 이상점 IRT는 .85($p < .001$)의 높은 상관을 보였다. 그러나 앞서 설명한 것과 같이 상관은 경향성으로 실제 순위 차이에 대한 세부 차이를 반영하지 못한다. 따라서 각 척도화 방법으로 산출된 순위를 직접적으로 비교해 보고자 하였으나, 세 척도화 방법 모두 동점자가 많아 사실상 순위 비교가 불가능했다. 특히, 성취지향의 우세 IRT에서는 1위가 127명이나 발견되었다. 이는 앞서 검사정보함수에서 본 것과 같이 극단적으로 정적인 특성을 가지는 개인들을 변별하지 못한다는 것을 구체적으로 보여준다.

논 의

본 연구의 목적은 고전검사이론(CTT), 우세 IRT 모형 그리고 이상점 IRT 모형이 성격 문항 개발 및 선별에 있어 어떠한 차이를 갖는지를 살펴보고, 우세반응 가정을 기반으로 구성된 우세 IRT 척도와 이상점 반응 가정을 기반으로 구성된 이상점 IRT 척도의 적합도를 비교해보는 데 있었다. 지금까지 성격측정과 관련된 대부분의 검사들은 우세 가정을 기반으로 개발되어왔으나, 최근 일부 연구(예: Chernyshenko et al., 2007; Stark et al., 2006)들은 우세 모형이 인지능력검사에는 적합하지만 개인의 특성 수준을 나타내는 성격 검사에는 우세 모형보다 이상점 모형이 더 적합하다고 주장하였다. 이에 따라 본 연구에서는 성취지향, 공정성, 협조성, 진실성이라는 네 개의 척도

개발 과정에 CTT, 우세 반응 IRT 모형, 이상점 반응 IRT 모형을 적용해 봄으로써, 세 모형을 비교해보고자 하였다. 또한 각 척도와 성적 간의 준거 타당도 검증은 통해 세 척도화 방법이 타당도 측면에서도 차이가 있는지를 알아보려고 하였다.

본 연구의 결과는 다음과 같이 요약해 볼 수 있다. 첫째, Stark 등(2006)의 주장과 같이, 고전검사이론에 기반을 둔 척도 개발 과정에서 .3 이하의 낮은 문항-총점 상관을 나타내 제거된 문항들은 대부분 중간문항이었다. 그러나 반대로 많은 중간문항들이 .3 이상의 높은 문항-총점 상관을 갖는 것으로 나타났다. 따라서 중간문항과 같은 비단조적인 문항들이 대부분의 낮은 문항-총점 상관을 갖기 때문에 우세반응 가정 하에서는 이러한 문항들이 모두 제거된다는 Stark 등(2006)의 주장이 명백히 지지된다고는 볼 수 없다.

둘째, MODFIT을 통하여 두 문항 반응 모형의 적합도를 비교해 본 결과, Chernyshenko 등(2007)의 연구와 비슷하게 전반적으로 이상점 모형의 적합도가 우세모형보다 약간 더 우수했다. 그러나 두 모형 모두 개별 문항(singlets)에서는 조정된 χ^2/df 통계치가 1 이상인 문항이 없었고, 두 문항 쌍(doublets)과 세 문항 쌍(triplets) 간의 평균값 차이도 모두 1 이하로 그 차이가 매우 미미했다. 또한 공정성 척도의 경우 0.05라는 아주 미세한 차이이기도 하나 우세 모형의 적합도가 이상점 모형의 적합도보다 약간 더 나왔다. 그러나 앞서도 언급했듯이 적합도의 차이를 비교하기 위한 기준이 아직 명확하게 존재하지 않으므로, 주어진 적합도 차이만을 가지고 어떤 모형이 더 낫다는 결론을 내리기는 어렵다.

셋째, 검사정보함수를 살펴 본 결과, 성취지

향과 공정성 차원의 경우 우세 모형의 검사정보함수가 이상점 모형의 함수보다 중간이하의 특성수준에서 더 높았으며, 정적으로 특성수준이 높은부분(+2이상)부분에서는 이상점 모형이 우세 반응 모형보다 높게 나타났다. 그러나 진실성과 협조성 차원의 경우 두 모형의 검사정보함수 간 차이는 거의 없었다. 이러한 결과는 거의 모든 특성수준에서 이상점 모형으로 개발된 문항들의 검사정보함수가 우세 모형으로 개발된 문항들에 비해 더 높다고 주장한 Chernyshenko 등(2007)의 의견과 일치하지 않는다.

넷째, 비단조성을 나타내는 문항들의 문항-총점 간 상관은 0에 가까워져 고전검사이론과 우세 반응 모형 척도에서는 제거된다던 Chernyshenko 등(2007)의 주장과는 달리, 본 연구에서는 비단조성을 나타내에도 불구하고 문항-총점 상관이 .3 이상인 문항이 여러 개 발견되었다. 이는 비단조성을 나타낸다고 해서 반드시 문항-총점 상관이 0에 가까워지는 것이 아니며, 이러한 문항이 고전검사이론이나 우세 모형에 기반을 둔 척도 개발 과정에서도 포함될 수 있음을 의미한다. 이는 첫 번째 결과와 같은 맥락이다.

마지막으로, 세 척도화 방법으로 개발된 성취지향, 공정성, 협조성, 진실성 척도와 학업 성적 간의 준거 타당도를 살펴 본 결과 성취지향을 포함한 모든 차원과 성적 간의 상관은 유의미하지 않았다. 유의미하지는 않았지만 상관 정도에서도 큰 차이가 없었다. 또한 고전검사이론과 우세 모형, 이상점 모형을 통해 산출된 개인들의 순위를 비교해보고자 하였으나, 각 방법으로 산출된 방법에서 모두 동점자가 많아 세부적으로 순위를 비교해 볼 수는 없었다.

본 연구는 Chernyshenko 등(2007)의 연구 방법을 참고함으로써 그들과 유사한 몇 가지 제한점을 갖는다. 첫째, 다분 척도를 이분으로 바꾸어서 사용했다는 점에 있다. 다분 척도를 이분 척도로 변환하여 사용할 경우 어느 정도의 정보를 손실하게 된다(Chernyshenko et al., 2007). 실제 본 연구에서는 고전검사이론을 통한 문항 선별을 위해 다분 4점 척도로 설문을 실시하였다. 그러나 다분 문항으로 모수 추정을 하기에는 표본이 충분히 획득되지 않아 IRT를 활용한 문항 선별 부분에서는 다분 반응을 이분 반응으로 전환하여 사용하였다. 그러나 이러한 변환은 Chernyshenko 등(2007)의 연구 방법에 따른 것으로 결과에는 큰 영향을 미치지 않는 것으로 판단된다. 그렇지만 추후 연구에서는 충분한 표본 수집을 통해 다분 문항에 대한 우세가정과 이상점 가정을 비교해 볼 필요가 있다.

둘째, 중간수준 문항 제작과 관련된 문제이다. 이는 Brown과 Maydeu-Olivares(2010) 등을 비롯한 여러 연구자들(Oswald & Schell, 2010; Reise, 2010)에게 비판점으로 제기되어 온 것으로 가장 현실적인 문제점은 중간 문항을 개발하는 것이 쉽지 않다는 것이다. Brown과 Maydeu-Olivares(2010)는 중간 문항 개발이 어려운 이유를 세 가지로 설명하고 있다. 첫 번째는 조건 절을 사용할 때 발생하는 문제이다. 예를 들어, “나는 매일 해야 할 일에 대한 계획을 세움에도 불구하고, 제 시간에 일을 끝내지 못한다.”라는 문항을 살펴보자. 이 문항에서 일부 사람들은 뒤의 문장인 “제 시간에 일을 끝내지 못함”에는 해당하지만, 앞 문장의 “매일 계획을 세움”에는 해당하지 않을 수 있다. 또는 반대의 경우도 있을 수 있다. 이런 문항은 응답자들에게 혼란을 주거나, 임의로

두 문장 중 한 문장에만 초점을 맞추므로써 결과적으로 무선적인 응답의 가능성을 증가시킨다. 이처럼 문항이 이중 의미로 귀결될 때 응답 선택에 문제가 발생할 수 있다는 것이다. 두 번째 문제는 문항이 응답자와 준거집단의 비교 판단을 요구하는 경우에 발생한다. 예를 들어 “내 방의 청결함은 평균수준이다.”라는 문항은 다른 사람들의 청결함에 대한 비교 판단을 요구한다. 예를 들어, 응답자 A와 B가 동일하게 청결한 수준을 가진 경우, 그들의 준거 집단에 따라 그들의 응답이 달라질 수 있다. 응답자 A의 주변 사람들이 매우 청결할 경우 A는 그 사람들과 자신을 비교하여 자신의 청결함 수준을 낮게 평가할 것이다. 반면 준거 집단의 청결성 수준이 매우 낮은 응답자 B는 그 집단과 자신을 비교하여 자신의 청결성 수준을 높게 평가할 것이다. 즉 이러한 판단은 두 응답자가 동일한 수준의 청결성을 가질지라도 응답자의 준거 틀에 의해 응답편향을 나타내기가 쉽다. 세 번째 문제는 중간 문항의 작성을 위해 구성개념과 관련이 없는 맥락이나 특정 세부사항이 문항 내용에 포함될 때 발생한다. 예를 들어, 중간 정도의 외향성을 측정하기 위해 만들어진 “나는 카페에서 친구와 조용히 수다 떠는 것을 즐긴다.”라는 문항의 경우, 이 문항의 구체적인 맥락은 다차원성을 유발할 것이고, 외향성 척도에서 낮은 변별도를 나타낼 것이다. 즉 이 문항은 이상점 모형에서는 외향성 척도이지만, 우세 모형에서는 “조용히 수다를 떠다.”라는 점에서 내향성을 의미하는 척도가 될 수 있으며 “카페에서 친구와 수다 떠는 것을 즐긴다.”라는 부분으로 인하여 온정성(warmth) 척도도 될 수 있다. 이처럼 중간 수준을 나타내는 응답자들의 행동을 묘사하려 할 경우, 그 묘사는 매우

모호해지거나 어떤 특정한 상황이나 시간이 적용됨으로써 응답자들에게 혼란을 줄 수 있다.

본 연구에서는 중간문항 제작 시 이러한 문제점을 가능한 고려하여 중간문항을 개발하려고 시도하였다. 그러나 결과에서 볼 수 있듯이 제작된 일부 중간 문항들의 문항내용과 위치모수(δ)가 일치하지 않고 극단에 가까운 문항 위치를 나타내었다. 이는 Chernyshenko 등 (2007)의 연구결과에서도 비슷하게 나타났다. 이러한 결과들은 중간 문항 개발의 현실적 어려움을 다시 한 번 나타내는 것으로, 추후 연구에서는 이러한 혼란과 개발의 어려움을 줄이기 위하여 중간 문항의 내용과 응답 절차에 대한 더 많은 탐구가 필요할 것이다.

셋째, 우세 IRT에서 문항들의 곤란도(difficulty)가 대부분 음수(-)로 나타났다는 점이다. 따라서 본 연구에서 척도 구성을 위해 문항을 선택할 때, 다양한 특성 수준을 포함하지 못하였고, 이러한 결과는 검사정보함수 및 적합도 결과 해석에 있어 주의를 요한다. 미래 연구에서는 다양한 곤란도가 나타나는 문항개발을 통해 편파된 곤란도가 모형의 적합도에 영향을 미치는지를 검증해보는 것도 의미가 있을 것이다.

마지막으로 본 연구에서는 척도화 방법들 간의 준거타당도를 비교하여 측정 정확성 뿐만 아니라 타당도 측면에서 어떤 방법이 더 의미 있는지를 알아보려고 하였다. 그러나 학업성적과 관련이 있을 것이라고 예측했던 성취지향 척도를 포함하여 모든 척도에서 유의미한 상관관계가 나타나지 않았고, 동점자가 많은 관계로 순위 비교도 불가능하였다. 따라서 세 척도화 방법들 간의 타당도에 대한 결론을 도출할 수 없는 아쉬움이 있다. 또한 본 연구 장면이 실제 조직 장면이 학교 장면이라는 한계

가 있다. 산업 및 조직 장면에서 성격은 성과나 개인의 행동을 예측하기 위한 목적으로 사용되기 때문에 이러한 타당도를 검증해 보는 것은 중요하다. 따라서 미래 연구에서는 실제 업무 수행에 대한 자료 수집 등이나 시뮬레이션을 통해 세 척도화 방법들 간 준거 타당도를 비교해보는 것이 필요하다.

이러한 제한점에도 불구하고, 본 연구는 성격측정에서 최근 새롭게 부상하고 있는 쟁점을 경험적으로 검증해보았다는 데 그 의미가 있다. 서두에 진술했듯이, 성격측정에서 우세 가정이 75년 동안이나 많은 연구자들에 의해 적용되어 왔고, 최근 일부 연구자들(Stark et al., 2006, Chernyshenko et al., 2007)에 의해 성격측정에서 이상점 반응 적용에 대한 쟁점이 제기되었다. 성격 측정에서 개인 응답에 대한 가정의 가정이 무엇인가는 척도 개발 시 어떤 문항을 선택하는지와 어떻게 개인 검사 점수를 산출할지와 직접적으로 관련되기 때문에 매우 중요한 문제이다. 이러한 이슈에 따라 본 연구에서는 아직까지 많은 연구가 이루어지지 않은 문항개발 과정에 더 적합한 모형이 무엇인가에 대한 또 하나의 경험적 증거를 제공하였다는 데 의의가 있다. 물론 본 연구를 통해 어떠한 모형이 더 적합한지에 대한 명확한 결론을 제시하지는 못하였다. 그러나 본 연구와 같은 연구들이 지속적으로 다양한 장면에서 다양한 척도들로 수행되어, 그 결과들이 축적된다면, 그리고 그러한 연구들이 어떠한 모형이 더 적합한지, 아니면 어떠한 모형을 사용해도 적합도에 차이가 없는지 등에 대한 일관된 결과를 나타낸다면, 향후에는 이러한 논쟁에 대한 결론을 도출할 수 있을 것이다.

본 연구는 성격측정에서 처음으로 이상점

가정을 적용하여 척도를 개발한 Chernyshenko 등(2007)의 연구를 바탕으로, 이들이 다루지 않은 척도들(성취지향, 공정성, 진실성, 협조성)을 대상으로 모형들의 차이를 비교해 보았다. Chernyshenko 등(2007)이 사용한 질서성 차원과 달리, 공정성, 협조성, 진실성 차원은 사회적으로 바람직한 응답의 영향을 많이 받을 수 있는 차원들이다. 즉 이러한 성격 척도들의 경우 중간 수준의 응답자들 역시 솔직하게 응답을 하지 않을 가능성이 많다. 실제 공정성 척도의 경우 매우 미미한 차이이기도 하나 우세 모형의 적합도가 아주 조금 더 나왔다. 이와 비슷하게, Carter와 Dalal(2010)의 연구에서는 JDI(job descriptive index)의 척도들 중 업무 만족 척도에서만 언폴딩 문항들이 나타났다. Carter, Lake 및 Zicka(2010)는 이러한 결과는 구성개념의 본질 자체가 문항의 언폴딩에 영향을 줄 수 있다는 것을 의미한다고 주장하였다. 즉 이상점 모형은 특정 성격차원에서만 적합한 것일 수 있다는 것이다. 본 연구의 결과는 Carter 등(2010)의 이러한 주장을 어느 정도 뒷받침 할 수 있는 것으로, 이상점반응 모형에 더 적합한 성격 척도와 우세반응 모형에 더 적합한 척도가 다를 수 있음을 시사한다. 따라서 향후 연구에서는 선행연구나 본 연구에서 다루지 않았던 또 다른 성격차원들에 대해서도 이와 비슷한 연구를 수행해 보아야 할 것이다.

지금까지 살펴본 것과 같이, 본 연구결과로는 성격측정에서 어떠한 모형을 적용하는 것이 옳은가에 대한 명확한 결론을 제시할 수 없었다. 이는 이선희(2010)의 연구나 Chernyshenko 등(2007)의 연구결과도 마찬가지다. 세 연구 모두에서 미세하게 이상점 모형의 적합도가 더 높긴 했지만, 실용적인 측면에서 그 정도

의 차이가 검사개발 과정에서의 어려움과 복잡성을 추가하는 것만큼의 실질적 이점을 제공해주는지는 명확하지 않다. 이상점 반응 모형은 중간수준의 문항과 함께 사용될 때 그 의미가 있다. 그러나 앞의 제한점에서 살펴보았듯이 중간문항을 개발하는 것은 그리 쉬운 일이 아니다.

이상점 반응 모형을 적용하는데 존재하는 또 다른 실용적 문제는 척도 길이와 관련된다. 다른 모형들과 달리 현 시점에서 이상점 문항 제작과 관련된 척도 길이에 대한 정보는 거의 존재하지 않는다. Dalal 등(2010)에 따르면 이상점 측정치들은 특성분포의 전 범위에 대한 정보를 담아낼 수 있어야 한다. 따라서 우세 모형보다 척도의 길이가 더 길어야 한다. 즉 우세 모형에 따라 제작된 기존의 검사들에 중간 특성을 나타내는 문항들이 추가되어야 한다. 그러나 보다 간단한 수준의 검사를 선호하는 최근 트렌드를 고려하면, 이처럼 척도의 길이가 더 긴 선발 도구는 문제가 될 수 있다. 이처럼 이상점 모형이 기존의 고전검사이론이나 우세반응 모형보다 훨씬 두드러진 적합도를 나타내지 않는 이상 더 많은 시간과 노력을 들여 이상점 모형을 사용하는 것이 맞는지 실무자들과 연구자들은 신중히 고려해 보아야 한다.

이상점 모형은 우세 모형보다 더 복잡한 모형이다. 또한 문항개발과 적용에 있어서도 아직은 우세 모형보다 더 복잡한 과정과 많은 시간을 요한다. 과학은 간명성을 선호한다. 따라서 이상점 모형이 그러한 복잡성과 노력을 들여도 충분히 좋을 만큼 더 높은 측정정확성과 타당성을 제시할 때까지는 척도 개발에 있어서 더 단순한 모형을 선호하는 것이 바람직할 것으로 보인다. 또한 더 정확한 모형 선택

을 위하여 우세 반응 모형처럼 이상점 반응 모형의 적합성을 검증하기 위한 더 많은 연구들이 필요하다.

참고문헌

- 김명소, 이현주 (2006). 성격검사의 형식이 응답왜곡(faking)에 미치는 효과: Normative 형식과 ipsative 형식의 비교. *한국심리학회지: 산업 및 조직*, 19(3), 371-393.
- 안여명, 유태용 (2010). 개인 및 팀 수준에서 성격과 적응수행 간의 관계: 일반적 자기 효능감의 매개효과와 변혁적 리더십의 조절효과. *한국심리학회지: 산업 및 조직*, 23(1), 155-179.
- 유태용 (2007). 성격의 6요인(HEXACO) 모델에 의한 성격특성과 조직구성원 직무수행 간의 관계. *한국심리학회지: 산업 및 조직*, 20(3), 283-314.
- 유태용, 김명현, 이도형 (1995). 5요인 성격검사의 개발 및 타당화 연구. *한국심리학회지: 산업 및 조직*, 10(1), 85-102.
- 유태용, 민병모 (2001). 다양한 장면에서 수행을 예측하기 위한 5요인 성격모델의 사용 가능성과 한계: 성격요인과 수행 간의 관계에 대한 통합분석. *한국심리학회지: 산업 및 조직*, 14(2), 115-134.
- 이선희 (2008). 성격검사문항에 대한 이상점 반응 모형의 적용. *한국심리학회지: 일반*, 27(3), 839-857.
- 이은정, 박동건 (2003). 성격 검사의 응답왜곡 탐지: 직무 바람직성의 개념화 및 선발 결정에의 영향. *한국심리학회지: 산업 및 조직*, 16(2), 121-151.
- 황지희 (2007). 대학생의 성취목표지향성과 학습 전략 및 학업성취의 관계. 연세대학교 대학원 석사학위 논문.
- Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology*, 49, 347-365.
- Barclay, J. E., & Weaver, H. B. (1962). Comparative reliabilities and ease of construction of Thurstone and Likert attitude scales. *The Journal of Social Psychology*, 68, 109-120.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Brown, A. & Maydeu-Olivares, A. (2010). Issues That Should Not Be Overlooked in the Dominance Versus Ideal Point Controversy. *Industrial and Organizational Psychology*, 3, 489-493.
- Carter, N. T., & Dalal, D. K. (2010). An ideal point account of the JDI Work satisfaction scale. *Personality and Individual Differences*, 49, 743-748.
- Carter, N. T., Lake, C. J., & Zickar, M. J. (2010). Toward Understanding the Psychology of Unfolding. *Industrial and Organizational Psychology*, 3, 511-514.
- 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.
- Chernyshenko, O. S., Stark, S., Drasgow, F., &

- Roberts, B. W. (2007). Constructing Personality Scales Under the Assumptions of an Ideal Point Response Process: Toward Increasing the Flexibility of Personality Measures. *Psychological Assessment, 19*(1), 88-06.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Dalal, D. K., Withrow, S., Gibby, R. E., & Zickar, M. J. (2010). Six Questions That Practitioners (Might) Have About Ideal Point Response Process Items. *Industrial and Organizational Psychology, 3*, 498-501.
- Davison, M. L. (1977). On a metric, unidimensional unfolding model for attitudinal and developmental data. *Psychometric, 42*, 523-548.
- Dragow, F., & Kang, T. (1984). Statistical power of differential validity and differential prediction analyses for detecting measurement nonequivalence. *Journal of Applied Psychology, 69*, 498-508.
- Dragow, F., Levine, M. V., Tsien, S., Williams, B. A., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple choice tests. *Applied Psychological Measurement, 19*, 143-165.
- Dragow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 Years After Likert: Thurstone Was Right! *Industrial and Organizational Psychology, 3*, 515-520.
- Edwards, A. L., & Kenney, K. C. (1946). A comparison of the Thurstone and Likert techniques of attitude scale construction. *Journal of Applied Psychology, 30*, 72-83.
- Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods, 4*, 265-287.
- Ellis, B. B., Becker, P., & Kimmel, H. D. (1993). An item response theory evaluation of an English version of the Trier Personality Inventory. *Journal of Cross-Cultural Psychology, 2*, 133-148.
- Goldberg, L. R. (1997). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe, 7*(pp.7-28). Tilburg, The Netherlands: Tilburg University Press.
- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Erlbaum.
- Hogan, R. Hogan, J., & Roberts, B. W. (1996). Personality Measurement and employment decision: Questions and answers. *American Psychologist, 51*(5), 469-477.
- Hough, L. M., & Scheider, R. J. (1996). *The frontiers of I/O Personality Research*. In K. R. Murphy(Ed.), *Individual differences & behavior in organizations*(pp.31-88). San Francisco: Jossey-Bass.
- 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). <http://ipip.ori.org/>
- Lake, C. J., & Zichar, M. J. (2010). *Respondent and test delivery characteristics that induce item unfolding*. Unpublished manuscript.
- Likert, R. (1932). A technique for the

- measurement of attitudes. *Journal of Applied Psychology*, 91, 25-39.
- Marcus Crede. (2010). Two Caveats for the Use of Ideal Point Items: Discrepancies and Bivariate Constructs. *Industrial and Organizational Psychology*, 3, 494-497.
- Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, 88, 139-157.
- McCrae, R. R., & Costa, P. T., Jr. (1989). The structure of interpersonal traits: Wiggins' circumflex and the five-factor model. *Journal of Personality and Social Psychology*, 56, 586-595.
- Mislevy, R. J., & Bock, R. D. (1991). *BILOG user's guide*. Chicago, IL: Scientific Software
- Nunnally, J. C., & Berstein, L. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660-679.
- Oswald, D., & Schell, K. L. (2010). Developing and Scaling Personality Measures: Thurstone Was Right-But So Far, Likert Was Not Wrong. *Industrial and Organizational Psychology*, 3, 481-484.
- O'Reilly, C. A., & Pfeffer, J. (2000). *Hidden value: How great companies achieve extraordinary results with ordinary people*. Boston: Harvard Business School Press.
- Reise, S. P. (2010). Thurstone Might Have Been Right About Attitudes, but Drasgow, Chernyshenko, and Stark Fail to Make the Case for Personality. *Industrial and Organizational Psychology*, 3, 485-488.
- 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. (2004). *A computer program for GGUM 2004: Technical reference manual*.
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, 24, 3-32.
- Roberts, J. S., Laughlin, J. E., & Wedell, D. H. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement*, 59, 211-233.
- Ross, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644.
- Rouse, S. V., Finger, M. S., & Butcher, J. M. (1999). Advances in clinical personality measurement: An item response theory analysis of the MMPI-2 PSY-5 scales. *Journal of Personality Assessment*, 72, 282-307.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings.

- Psychological Bulletin*, 124, 262-274.
- Simms, L. J., & Clark, L. A. (2005). Validation of computerized adaptive version of the Schedule for Nonadaptive and Adaptive Personality (SNAP). *Psychological assessment*, 17, 28-43.
- Stark, S., Chernyshenko, O. S., Chan, Lee, & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, 86, 943-953.
- 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. *Journal of Applied Psychology*, 91, 25-39.
- Stark, S. (2008). MODFIT: *A computer program for model-data fit*. Unpublished manuscript, University of Illinois at Urban-Champaign.
- Tay, L., Drasgow, F., James, R., & Williams, B. (2009). Fitting Measurement Models to Vocational Interest Data: Are Dominance Models Ideal? *Journal of Applied Psychology*, 94(5), 1287-1394.
- Tay, L., Ali, U. S., Drasgow, F., & Williams, B. (2011). Fitting IRT Models to Dichotomous and Polytomous Data: Assessing the Relative Model-Data Fit of Ideal Point and Dominance Models. *Applied Psychological Measurement*, 35(4), 280-295.
- Thurstone, L. L. (1927). A law of comparative judgment. *Personnel Psychology*, 44, 703-742.
- Thurstone, L. L. (1928). Attitude can be measured. *American Journal of Sociology*, 33, 529-554.
- Waples, C. J., Weyhrauch, W. S., Connell, A. R., & Culbertson, S. S. (2010). Questionable Defeats and Discounted Victories for Likert Rating Scales. *Industrial and Organizational Psychology*, 3, 477-480.
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, 20, 71-87.
- 1차 원고접수 : 2012. 3. 8
2차 원고접수 : 2012. 5. 10
최종게재결정 : 2012. 5. 14

Comparison between dominance process model and ideal point response model for personality assesment

Dong-Gun Park Myung-Ok Choi Won-Sun Lee Hye-Min Lee Jun-Hee Lee

Korea University Department of Psychology

The present study investigated the utilities of two types of item response process models(dominance model and ideal point model) for personality item parameter estimation and scoring. The authors developed scales for four personality traits(achievement, fairness, cooperation and honesty) using classical test theory, dominance item response theory(IRT) method, and ideal point IRT method and compared the methods in terms of model-data fit, information and criterion validity. Results show that the fit of ideal point IRT model was better than that of dominance IRT model, but the difference between the fit of two models was very slight. The test information functions of ideal point IRT model and dominance IRT model for honesty and cooperation scales were very similar. The criterion-related validity based on individual ability estimates and grades was not significant for the three methods but the validity for the ideal point method is not better than dominant IRT model. Implications and limitations of the findings are discussed.

Key word: personality assessment, ideal point response, dominant process, test construction, classical test theory

부록 1. 공정성 30개 문항

No.	문항내용	ITC*	요인 부하량
공정01*	공부를 열심히 안 해도 부정적인 방법으로 높은 점수를 받을 수 있다.	.37	.38
공정02*	나에게 피해를 주지 않는다면 부정한 일들이 벌어져도 상관없다.	.55	.58
공정03*	법과 규칙을 항상 준수하는 것은 피곤한 일이다.	.60	.63
공정04*	부당이득을 취하는데 거리낌이 없다.	.63	.66
공정05*	법의 허점을 이용하여 세금을 피할 수 있다면 피하고 싶다.	.58	.61
공정06*	윤리라는 것은 허울 좋은 이야기일 뿐이다.	.53	.56
공정07*	친한 친구한테 빌린 돈은 제 때 갚지 않아도 된다.	.34	.36
공정08*	규칙 준수는 살아가는데 중요치 않다.	.51	.54
공정09*	남의 물건을 슬쩍 집어오더라도 크게 신경 쓰지 않는다.	.40	.41
공정10*	특별한 사정이 있다면 규칙을 잠시 어겨도 된다.	.60	.64
공정11	다른 사람들만큼은 규칙을 지키는 편이다.	.15	.16
공정12*	나의 도덕성은 평균(중간) 수준이다.	.16	.18
공정13*	뇌물은 나쁜 것이라고 생각하지만, 상황에 따라 받을 수도 있다.	.61	.65
공정14	다른 사람들만큼은 공정하다.	.25	.25
공정15	다른 사람들만큼의 도덕적 기준을 가지고 있다.	.29	.31
공정16*	욕을 먹지 않을 정도로만 규칙을 지킨다.	.61	.63
공정17*	윤리 수업이 옳다고는 생각하지만, 항상 윤리를 따르지는 않는다.	.56	.60
공정18*	때로 공정하지 못하게 상황이나 일을 처리하는 경우가 있다.	.53	.56
공정19*	보는 사람이 없을 때, 교통 법규를 지킬 때도 있고 지키지 않을 때도 있다.	.58	.64
공정20*	항상 공정한 판단을 하는 것은 아니다.	.41	.44
공정21	나에게 이익이 오지 않더라도 공정하게 판단하려고 노력한다.	.44	.45
공정22	뇌물을 받는 것은 어떤 경우에라도 용납할 수 없다.	.50	.54
공정23	도덕 및 규칙을 준수하기 위해 많은 시간과 비용을 할애할 용의가 있다.	.54	.57
공정24	도덕적 기준에 맞지 않는 일은 할 수 없다.	.60	.62
공정25	법은 언제나 공정해야 한다.	.41	.42
공정26	보는 사람이 없어도 교통 법규를 꼭 지킨다.	.62	.67
공정27	윤리적인 문제에는 엄격하게 대응한다.	.54	.56
공정28	종업원의 실수로 더 받은 거스름돈은 꼭 돌려준다.	.52	.54
공정29	사소한 법규도 어긴 일이 거의 없다.	.57	.60
공정30	의도하지 않게 규칙을 어기면 매우 죄책감이 든다.	.34	.36

N=693. *ITC=item-total correlation

*역문항(11번에서 20번은 중간수준 문항)

부록 2. 진실성 30개 문항

No.	문항내용	ITC*	요인 부하량
진실01*	상대를 이용하기 위해 마음에 없는 말을 한다.	.51	.35
진실02*	빈말을 자주 한다.	.49	.31
진실03*	다른 사람들에게 자주 속을 모르겠다는 말을 듣는다.	.37	.17
진실04*	사람들에게 겉과 다른 속마음을 들이지 않을 자신이 있다.	.38	.19
진실05*	모르는 상황에서도 아는 척 한다.	.38	.20
진실06*	의도적으로 거짓말을 할 때가 많다.	.54	.38
진실07*	성공하기 위해서 겉과 속이 다르게 행동할 때가 있다.	.64	.52
진실08*	출세를 위해 아부 하는 것은 그다지 잘못된 것이 아니다.	.41	.26
진실09*	솔직하게 감정을 표현하는 것을 꺼려한다.	.31	.10
진실10*	상황을 모면할 수 있다면 당연히 거짓말을 한다.	.49	.32
진실11	남들만큼은 진실하다.	.13	.02
진실12*	상황에 따라 내 생각을 말할 때도 있고, 그렇지 않을 때도 있다.	.17	.05
진실13*	나의 진실성은 평균수준이다.	.34	.15
진실14*	내 진심을 모두 드러내지는 않는다.	.45	.23
진실15*	다른 사람들보다 내 의견(생각, 감정)을 더 많이 표현하거나, 더 적게 표현하지 않는다.	.04	.00
진실16*	다른 사람들처럼 나도 가끔은 타인을 속일 때가 있다.	.57	.41
진실17*	다른 사람이 나에게 의견을 물으면 솔직하게 말할 때도 있고, 그렇지 않을 때도 있다.	.52	.33
진실18*	내가 할 수 없는 것을 할 수 있다고 말할 때도 있고, 솔직하게 할 수 없다고 할 때도 있다.	.34	.18
진실19*	남들 하는 정도의 아부는 한다.	.44	.28
진실20*	남들만큼 필요에 따라 약간의 거짓말은 한다.	.60	.46
진실21	나에게 이득이 된다고 해도 속마음과 다르게 행동할 수 없다.	.39	.20
진실22	원하는 것을 얻기 위해 남을 이용하지 않는다.	.57	.42
진실23	내 자신의 생각이나 기분을 솔직하게 표현하는 편이다.	.33	.10
진실24	상황 때문에 할 수 없는 것을 할 수 있다고 속이지 않는다.	.40	.21
진실25	어떤 상황에서도 싫으면 싫다고 말한다.	.33	.17
진실26	대인관계에서 진실성이 가장 중요하다고 생각한다.	.37	.16
진실27	겉과 속이 다르게 행동하는 것은 나와 어울리지 않는다.	.62	.46
진실28	마음에 없는 말은 하지 않는다.	.63	.50
진실29	창피하더라도 필요하다면, 솔직하게 도움을 청한다.	.31	.11
진실30	불만이 있으면 직설적으로 말하는 편이다.	.28	.08

N=693. *ITC=item-total correlation

*역문항(11번에서 20번은 중간수준 문항)

부록 3. 협조성 30개 문항

No.	문항내용	ITC*	요인 부하량
협조01 ^a	팀으로 일하는 것을 매우 싫어한다.	.34	.36
협조02 ^a	팀원들과 의견을 조율하는 과정은 시간과 노력의 낭비다.	.55	.59
협조03 ^a	다른 사람들과 함께 일하는 것이 힘들다.	.59	.64
협조04 ^a	사람들과의 협력이 필요한 일/과제 보다는 나 혼자만 잘하면 되는 일이 훨씬 편하다.	.48	.53
협조05 ^a	여럿이 함께 일하는 것보다 혼자서 일하는 것이 편하다.	.50	.52
협조06 ^a	조용히 혼자서 일을 진행하는 것을 선호한다.	.50	.55
협조07 ^a	다른 사람들과 협력하는 것이 내겐 중요하지 않다.	.45	.47
협조08 ^a	사람들과 함께 협력해야 하는 일은 자신이 없다.	.48	.50
협조09 ^a	사람들과 함께 의견을 조율하며 일하는 것이 불편하다.	.46	.50
협조10 ^a	협력이 필요한 래프팅을 하느니 혼자할 수 있는 등산을 하겠다.	.33	.34
협조11 ^a	다른 사람보다 더 협조적이지도 덜 협조적이지도 않다.	.42	.43
협조12	협조적일 때도 있고, 경쟁적일 때도 있다.	.33	.33
협조13 ^a	나의 팀워크 능력은 평균(중간) 수준이다.	.14	.11
협조14 ^a	팀 과업시 다른 사람들이 하는 정도만큼만 협조한다.	.44	.46
협조15	혼자 일하는 것도 좋고, 여럿이 함께 일하는 것도 좋다.	.28	.28
협조16 ^a	팀의 이익을 우선시 할 때도 있고, 내 이익을 우선시할 때도 있다.	.11	.08
협조17 ^a	나에게 피해가 없는 정도에서만 팀 활동에 참여한다.	.54	.58
협조18 ^a	팀 의견에 따르긴 하지만 나 자신을 희생하지는 않는다.	.59	.63
협조19 ^a	팀 작업시 다른 사람들이 협조하는 만큼만 협조한다.	.49	.52
협조20 ^a	팀에 피해가 없을 정도만 협조한다.	.32	.32
협조21	경쟁적이기보다는 협조적인 편이다.	.59	.64
협조22	나와 다소 성향이 다른 사람들과도 함께 일할 수 있다.	.30	.32
협조23	다른 사람들과 의견을 조율하여 중요한 사항을 결정하는 것이 어렵지 않다.	.56	.62
협조24	여러 사람이 함께 일을 할 때, 더 좋은 성과를 낼 수 있다.	.59	.65
협조25	함께 일하는 사람들에 맞추어서 생각하고 행동한다.	.51	.56
협조26	혼자 일하는 것보다 함께 일하는 것이 즐겁다.	.52	.57
협조27	많은 시간이 걸리더라도 팀원들과 함께 일하는 것은 매우 가치 있다.	.52	.57
협조28	집단의 구성원으로서 동료들과 함께 일하는 것을 매우 좋아한다.	.48	.53
협조29	팀의 원활한 작업을 위해서 팀에 적극적으로 협조한다.	.58	.62
협조30	나는 매우 협조적인 사람이다.	.54	.57

N=693. *ITC=item-total correlation

^a역문항(11번에서 20번은 중간수준 문항)