韓國心理學會誌
Korean Journal of Psychology
1992. Vol. 11. No. 1, 24-40

심리 척도에 의한 성격 집단 분류에서 예측되는 분류 오류의 확률 : 항목반응이론의 응용

이 연 희 박 광 배

충북대학교 심리학과

심리학의 모든 분야에 걸쳐서 능력, 성격특질, 혹은 태도, 가치관 등이 대비되는 집단들을 구성하고, 그러한 집단들이 특정한 구성개념(construct) 상에서 어떤 차이를 가지는지를 비교검증하는 연구 및 분석설계가 많이 활용되고 있다. 그런데 집단구분 혹은 독립변인의 조작을 위하여 쓰이는 측정도구의 신뢰도가 완벽하지 않은 한, 그러한 측정도구를 이용하여 집단구분을 하는 경우 반드시 할당오류 혹은 분류오류를 범하게 된다. 이 분류오류는 말하자면 독립변인의 조작에서의 오류(error)이다. 반면에, 이러한 연구 및 분석설계를 위해서 자주 활용되는 분석기법들인 변량분석이나 회귀분석은 독립변인이 오차를 가지지 않는다는 가정을 한다. 측정도구를 이용한 독립변인의 조작에서 분류오류를 완전히 없애는 것은 불가능하지만, 특정한 척도를 집단분류를 위하여 이용하는 경우 분류오류가 어느 정도 발생할 것인지를 파악하는 것은 결과의 일반화와 해석의 일관성을 위하여 중요한 정보가 된다. 본 논문은 심리척도상의 연속적인 점수에 기초하여 집단을 분류할 경우 분류 오류의 확률이 얼마인지 파악하는 방법과 절차를 항목반응이론(item response theory) 의 개념적 틀안에서 실제 자료를 예로 들어 기술한 것이다. Hulin, Drasgow, and Parsons (1983) 가 기술한 방법에 의해 추정된 분류오류확률의 해석에 있어서 주의할 점이 논의되었다.

일반적으로 심리학 여러 분야에서 연구 설계를 하거나 자료 분석 계획을 수립하면서 연구자가 특정한 성격 변인상의 점수를 기준으로 집단을 인위적으로 분류하는 경우가 많다. 가장 일반적인 방식은 상대적인 성격 특질의 높고 낮음을 기준으로 집단들을 분류하는 것이다. 이러한 연구의 공통된 문제점은 분류 오류의 발생이다. 분류 오류는 특정한 척도상의 점수를 이용하여 피험자를 분류하는 경우, 실제에 있어서 특질이 약한 피험자가 강한 특질 집단에 잘못 분류되는 것과 실제에 있어서 특질이 강한 피험자가 약한 특질 집단에 잘못 분류되는 것을 말한다. 이러한 분류 오류가 발생하는 이유는 집단 분류에 사용되는 척도의 신뢰도가 완벽하지 않기 때문이다. 검사의 신뢰도에 대한 조작적 정의는 한 검사를 같은 피험자에게 여러번 실시했을 때 일관된 점수가 산출되는 것을 말한다 (Allen & Yen, 1979). 척도의 신뢰도가 완벽하면 피험자가 그 척도를 실시할 때마다 관찰점수가 변하지 않게되고, 이는 관찰점수(X)가 진점수(T)와

1) 이 논문은 이 연희의 석사학위논문(1992)의 일부를 발췌한 것임. 항목반응이론은 문항반응이론이라고 번안할 수도 있다. 그러나 이 이론이 문자로 기술된 설문지 문항에만 적용되는 것은 아니다. 예를 들어 농구선수가 공던지기를 10번 시도하는 경우 매번의 시도를 하나의 항목으로 간주하여 성공율에 대한 항목반응이론을 적용할 수도 있다(Masters, 1988). 따라서 본 연구에서는 문항이라는 용어보다 항목이라는 용어를 사용하고자 한다.

일치하기 때문이다. 이 경우에 한해서 관찰점수 (X)로 집단을 분류하는 경우 분류 오류가 발생하지 않는다.

집단 분류에서의 오류는 결과의 일반화를 위해서 뿐만 아니라 결과 분석을 위한 통계 방법의 적용에서도 문제가 될 수 있다. 심리 척도의 연속적인 점수에 의해 집단 분류가 이루어지는 경우 그렇게 분류된 집단들은 일반적으로 독립변인으로 간주되어 회귀분석이나 변량분석에 이용된다. 그런데 회귀분석과 변량분석등의 통계기법들은 일반적으로 독립변인에는 오차가 없다는 가정에 기초하고 있다. 그런데 이런 경우 분류오류의 존재는 곧 독립변인의 오차를 의미하고 따라서 회귀분석이나 변량분석등의 가장 기본적인 가정 중의 하나를 파기하게 된다. 물론 심리척도를 이용하여 집단분류를 하는경우 그 척도의 신뢰도가 완벽하지 않는한 분류오류를 완전히 없앨 수는 없지만 그 오류의 정도를 가늠하는 것은 결과의 해석과 일반화를 위해서 중요할 수도 있다.

현재까지 심리학 연구들에서 집단 분류에 쓰여졌던 척도는 대부분 고전적 측정 이론에 기초한 것이다. 그런데 고전적 측정 이론을 이용하여 개발한 척도로 집단 분류하는 경우 집단 분류 오류 예측을 가능하게 해 주는 이론적인 기반이 없다. 이는 고전적 측정 이론이 기본적으로 항목 성향을 파악해내기 위한 이론이고, 응답자의 특질 정도를 추정해내기 위한 것이 아니기 때문이다. 다시 말하면 고전적 측정 이론은 소위 "근본 측정(fundamental measurement)"으로서의 성격을 가지지 않는 이론이다(Andrich, 1988). 한편 Guttman과 Lazarsfeld(1950)에 의해 개발된 항목반응이론은 항목특성을 규명하는 것과 동시에 각 피험자의 잠재적 특질을 추정해내기 위해 이론적으로 발전되었으므로 그러한 이론적 속성을 이용하면 분류 오류를 어느정도 예측할 수 있게 해준다.

본 논문은 다음과 같은 목적을 가지고 수행되었다. 연속적인 척도속성을 가지는 심리척도들이 연구목적상 집단분류를 위하여 이용되는 경우 내재하는 본질적인 문제, 오류 확률을 체계적으로 파악할 수 있는 척도 개발 방법을 실제 자료에 적용하여 설명하고자 한다. 다시 말하면 주어진 척도상의 연속적인 점수를 이용하여 피험자를 특질이 강한 집단과 특질이 약한 집단으로 분류하는 경우에 있어서 실제에 있어서 강한 집단에 속할 피험자가 약한 집단에 속하게 되는 오류 확률과 실제에 있어서 약한 집단에 속할 피험자가 강한 집단에 속하는 오류 확률을 이론적으로 파악하는 방법을 기술하고, 그 방법의 문제점과 주의할 점을 제시하고자 하는 것이 본 논문의 목적이다. 이를 위하여 본 논문에서는 1881명의 고등학생 및 대학생의 자료를 분석하였다.

우선 본 논문에서 활용된 항목반응이론의 이론적 개념들을 소개하고자 한다. 물론 항목반응이론에 관한 지금까지의 이론적 발달 과정을 모두 섭렵할 수는 없으므로 본 논문의 목적에 부합하고 필요한 개념들만을 소개하고자 한다. 그런 후 자료분석방법과 결과를 기술하고 결과에 대한 논의 및 결론을 제시하고자 한다.

항목반응이론(Item Response Theory)

배경

항목반응이론(Item Response Theory, IRT)은 항목에 대한 피험자의 반응을 그 피험자의 능력 또는 측정하고자 하는 잠재적 특질과 연관시키는 확률적 모형이다. IRT가 전제하고 있는 중요한 가정은 특정한 반응이 발생할 확률은 그 개인이 가지고 있는 특질의 정도에 의해 결정된다는 것이다. 이 말은 너무나 당연한 말처럼 들리지만 문제는 개인이 가지고 있는 특질이라는 것이 직접 관찰되지 않는다는데 있다. 즉 그것은 잠재적 특질(latent trait)이다. IRT는 항목의 성격을 명확히 파악하고, 그 파악된 성격에 기초하여 항목에 대한 피험자의 반응으로부터 그 피험자의 잠재적 특질의 정도를 추정하기 위한 이론이다.

항목반응이론은 응답자의 잠재적 특질을 소위 "최대우도추정(Maximum Likelihood Estimation)"

방식에 의해 산출한다. 이 방법은 잠재적 특질을 추정하기 위하여 그 특질의 분포 양상에 대한 가정을 할 필요가 없다는 장점이 있다(Hulin, Drasgow & Parsons, 1983). 최대우도추정방식은 순환과정(iteration)을 통하여 수렴한 잠재적 특질을 산출하는데, 이 수렴 과정에서 여러개의 잠정적인 계수치들이 산출되고 이 계수치들은 소위 "점진적 정상성(asymptotic normality)" 을 가지는 속성이 있다(Bollen, 1989). 그러나 이 추정된 계수치들의 분포가 가지는 정상성은 최대우도추정방식에 의해 산출된 계수값들이 결과적으로 획득하는 속성이지, 계수 산출을 위하여 사전에 요구되는 가정(assumption)이 아니라는 사실을 주목해야 한다. 반면에 고전 측정이론에서 소위 진점수(true score)의 신뢰구간을 추정하기 위해서는 한 응답자가 같은 검사에 행한 여러번의 반응 점수들이 정상분포한다는 가정을 해야 한다. 그 이유는 측정오차(measurement error)가 정상분포하고 그 기대치가 0이라는 가정 없이는 진점수의 신뢰구간에 대한 추정이 불가능하기 때문이다(Allen & Yen, 1979, p.89). 이 때 측정오차의 분포는 개념적으로는 한 응답자의 관찰된 점수들의 분포이지만 실제 고전 측정 이론에서는 여러명의 응답자들이 한 번 행한 점수분포를 이용하여 추정하고, 다만 모든 응답자들의 오차분포 혹은 점수분포가 모두 동일하다는 가정을 한다(Allen & Yen, 1979, p.59). 따라서 결과적으로 관찰된 점수가 전집에서 정상분포한다는 가정을 가지고 있다. 이런 의미에서 IRT는 고전 진점수 이론보다 더 실증적인 이론이라고 할 수 있다.

Logistic 모형

아래의 공식(Birnbaum, 1968)은 특질 반응(Keyed Response: KR)을 할 확률(P)을 피험자 계수인 θ와 항목 계수인 a, b, c로 나타낸 것이다(Hulin, Drasgow and Parsons, 1983. P.29). θ는 잠재적 특질의 정도를 나타내고, a는 항목의 변별계수이고, b는 각 항목 난이도계수이며, c는 각 항목의 추측계수이다. 피험자 계수 θ를 횡축으로, KR을 할 확률 P를 종축으로 도표화하는 경우(항목특성곡선) c는 확률 곡선과 만나는 종축의 값이다. 정확한 의미에서의 a, b계수는 항목특성곡선의 변곡점(inflection point)에서의 기울기와 θ값이지만 c가 0이라고 가정하면 전형적으로 P가 0.50인 점에서의 기울기와 θ값으로 추정한다. c가 0이라고 가정하지 않는 경우에도 추측에 의해 정답이나 KR을 할 확률이 높지 않으면 a는 p가 0.5인 점에서의 기울기에 비례하고 b는 p가 0.5인 점에서의 θ값에 근사한다. 반면에 c가 0보다 많이 크면 그 항목은 항목 선정 과정에서 탈락될 가능성이 높아진다.

$$P = c + (1 - c)\frac{1}{1 + EXP[-1.7a(\theta - b)]}$$

p = 특질 반응(Keyed Response : KR)을 할 확률
θ = 잠재적 능력 혹은 특질
a = 변별계수 : 변곡점에서의 기울기
b = 난이도계수 : 변곡점에서의 θ
c = 추측계수 : θ가 최하인 점에서의 P

위의 공식을 이용하면 특정한 정도의 특질 θ를 가지는 사람이 a, b, c의 속성을 가지는 항목에 대하여 특질 반응(KR)을 할 확률 P를 도출할 수 있다. 위의 공식은 3모수 로지스틱 모형(The Three Parameter Logistic Model: 3PLM)으로 IRT 중에서 가장 일반적인 모델로서 Tucker(1946)와 Lord(1952)에 의해 개발되기 시작하였고 특히 능력을 측정하는 척도를 위한 이론적 모델로서 이용되어 왔다. 〈일반적〉 이라는 말은 많이 쓰인다는 뜻이 아니고 다른 IRT 모델들을 모두 함축하고 있다는 뜻이다. 이 3PLM의 특징은 능력을 측정하는 경우 능력이 전무한 사람도 때로는 추측에 의해 항목에 대한

정답을 할 확률(P)이 0보다 클 수 있다는 가정을 하는데 있다. 그러나 3PLM에서 계수 c에 대한 추정치는 불안정한 경우가 많다. 왜냐하면 특질의 정도가 극도로 낮은 응답자의 수가 일반적으로 매우 적은 것이 통례이기 때문이다. 위의 공식에서 c가 0이라고 가정하면 2모수 모델이 된다. 2모수 모델이 사용되는 경우는 태도 측정에서처럼 응답자가 각 문항에 정답을 하려고 추측할 필요가 없을 때이다 (Lazarsfeld, 1950, 1959). 한편 1모수 모델은 위의 공식에서 c가 0이고, a가 모두 같다고 가정할 경우의 모델이다. 1모수의 사용은 한 척도의 모든 항목이 변별력이 같다고 가정되는 경우에 사용한다(Rasch, 1960). 이 때 변별계수는 일반적으로 1로 가정한다. 따라서 1모수 모델에서 항목의 속성은 난이도 계수 b 하나만으로 정의된다.

최근에는 반응범주가 이분화된(예-아니오 혹은 정답-오답) 척도에 대해서 뿐만 아니라 서열적 반응범주(ordered response categories), 평정척도(rating scale), 같은 항목에 대한 여러번의 시도에서 성공하는 빈도(binomial trial), 포아송 빈도(poisson counts) 등의 반응 변인들에 대해서 1모수 모델인 Rasch 모델의 이론적 틀 안에서 반응자들의 잠재적 특질을 추정하는 모델들이 개발되고 있다 (Masters, 1988, 1980; Andrich, 1988, 1982, 1978; Samejima, 1969; Molenaar, 1983; Duncan, 1985; Wright & Masters, 1982; Rasch, 1972, 1960).

방법

위에서 언급한 세 개의 모델 중 한 개의 모델이 정해지면 IRT는 앞에서 언급한 最大尤度推定 방법에 의해 θ, a, b, 및 c를 산출해낸다. 최대우도추정은 우선 항목계수들 a, b, c 에 임의적인 숫자를 주고 尤度(likelihood)가 최대가 되도록 각 피험자의 θ(잠재적 특성)를 산출한다. 예를 들어 검사 총점이 5점인 응답자가 100명이 있다고 가정하자. 이들 중 70명이 어떤 항목에 대해 특질 반응(KR)을 하였다면 실증적 P는 0.70이 된다. 주어진 a, b, c 값과 이 실증적 P값의 전제하에 그 응답자들의 θ가 -1일 확률, -0.50일 확률, 0.05일 확률 등을 모든 θ값에 대해 산출한다. 그 중 가장 높은 확률(최대우도)을 갖는 θ값이 이 집단의 $\hat{\theta}$라고 잠정적으로 추론한다. 이 산출된 $\hat{\theta}$와 항목계수들을 위에 기술한 logistic function을 위한 공식에 대입하여 산출된 P를 연결하는 항목특성곡선을 그린다. 이 때의 산출된 P는 이론적 P이다. 동시에 각 $\hat{\theta}$수준에서 각 항목에 대해 실제로 특질반응(Keyed Response: KR)을 한 피험자의 비율을 구하여 위에서 산출한 이론적 항목특성곡선과 얼마나 잘 부합하는지를 비교한다. 두 종류의 비율들이 어느 적정 수준이상 부합하지 않으면 이번에는 $\hat{\theta}$를 고정시키고 尤度가 최대가 되도록 a, b 및 c를 산출한다. 또 다시 공식에 의한 이론적 항목특성곡선과 실제로 KR를 한 피험자의 비율을 비교하여 부합의 정도를 검증한다. 만약 그래도 부합의 정도가 좋지 않으면 이번엔 다시 항목계수들을 고정시키고 $\hat{\theta}$를 산출한다. 공식에 의한 이론적 항목특성곡선과 실제 KR를 한 피험자의 비율의 차이가 어느 적정수준 이하로 작아질때까지 이러한 절차를 반복한다 (Lord, 1980).

위의 절차에 의해 최종적으로 구해진 $\hat{\theta}$는 각 피험자의 잠재적 특질에 대한 추정치이므로 그 값을 그대로 그 피험자의 측정 점수(score)로 간주한다. 이 추정치는 최대우도추정방식에 의해 산출된 것이라는 사실이 매우 중요하다. 즉 최대우도추정방식은 앞에서 언급한 점진적 정상성 이외에 이론적으로 "잠재적"특질을 추정하는데 점진적 효율성(asymptotic efficiency)을 갖고 있다고 알려져 있다(Kendall & Stuart, 1979). 이 말은 다른 어떠한 방식으로 측정점수를 산출하여도 자료의 수가 많은 경우 최대우도추정방식에 의해 산출된 측정점수보다 직접 관찰되지 않는 잠재적 특질에 관하여 더 많은 정보(information)를 제공하지 않는다는 뜻이다(정보의 구체적 개념은 항목과 검사 정보 부분에서 자세히 설명될 것임).

그러나 추정된 $\hat{\theta}$가 실제의 잠재적 특질의 정도 그 자체를 정확하게 대변하는 것은 아니다. MLE방식에

의한 θ의 추정은 여러번의 순환을 거치게 되고 그 과정에서 추정치가 어떤 일정한 수렴치를 중심으로 계속 바뀌게 된다. 이 과정에서 소위 정보함수가 산출되고 이 정보함수로부터 추정된 $\hat{\theta}$의 표준오차(SE) 가 산출된다. 정보함수로부터 표준오차가 산출되는 자세한 과정은 곧 설명될 것이다. 이 표준오차를 이용하여 실제 잠재적 특질의 신뢰구간을 설정할 수 있다. 예를 들어 관찰되지 않는 잠재적 특질에 대한 95%신뢰구간은 (추정된 $\hat{\theta}$) ±2SE 이다.

지역 독립성과 항목계수 불가변성

항목반응이론은 중요한 가정 위에 수립되었다. 그것은 일정한 속성(a, b, c)을 가진 주어진 항목에 대한 피험자의 반응을 결정하는 요인이 오직 하나이며 그것은 바로 잠재적 특질(θ)이라는 것이다. 즉 종속변인으로서의 P는 오직 θ의 함수로서 표현되고 있다. 모든 항목에 대한 반응이 하나의 잠재적 특질의 함수로 표현되고 있으며, 이것은 곧 척도가 일차원성(unidimensional)이라는 것을 뜻하고, 척도상에 나타나는 개인차는 잠재적 특질상의 개인차에 의해서만 유발되고 다른 요인에 의해서는 유발되지 않는다는 것을 뜻한다 (Lord & Novick, 1968).

같은 내용을 조금 다른 각도에서 이해할 수있다. 만약 척도에 포함된 항목들이 일차원성이라면 모든 항목들은 동일한 유일의 잠재적 특질을 측정한다는 뜻이다. 만약 모든 항목들이 동일한 특질을 측정하고 다른 요인의 개입이 없다면 모든 항목들 사이의 상관계수는 당연히 높아야한다. 항목 Ⅰ에 대하여 KR를 한 피험자는 다른 항목 Ⅱ에 대하여도 KR를 하는 경향이 있을 것이고, Ⅰ에 非-KR를 한 피험자는 Ⅱ에도 非-KR를 하는 경향이 있을 것인데 이것은 두 피험자의 잠재 특질의 정도가 다르기 때문이며 다른 이유 때문이 아니다. 즉 만약 척도가 일차원성이라면, 항목들 사이의 높은 상관관계 또한 피험자들 사이의 상이한 잠재적 특질에 의해서만 유발되고 다른 요인에 의해서는 유발되지 않는다. 잠재적 특질의 정도가 같은 피험자들 사이에서의 항목들간의 상관계수는 어떻게 될까 ? 두 피험자가 잠재적 특질의 정도가 같다면 그들의 반응 패턴도 같을 것이다. 즉 한 피험자가 Ⅰ에 KR를 하고 Ⅱ에도 KR를 했다면 잠재적 특질이 같은 피험자도 같은 반응 패턴을 보일 것이고 결과적으로 각 항목은 잠재적 특질이 같은 피험자들 내에서는 변산이 0이 되어 항목 Ⅰ과 Ⅱ는 零인 상관계수를 보이게 된다.

잠재적 특질이 같은 피험자들 내에서 항목들간의 상관계수가 0이라야 한다는 원칙을 지역 독립성(local independence)이라고 부른다. 전체 피험자들을 대상으로 했을 때 나타나는 항목들 사이의 높은 상관계수가 오직 유일하게 잠재적 특질에 의해 유발된 것이라면, 잠재적 특질을 통제(control)하고 난 후에는 그 상관계수가 사라져야 하는 것이 당연하다. 잠재적 특질을 통제하는 한 방법은 특질의 정도가 같은 피험자 집단을 추출하는 것이며 이 집단내에서는 상관계수를 유발시키는 잠재적 특질 상의 변산이 존재하지 않으므로 항목간 상관계수 또한 존재할 수 없다. 그러나 만약 척도가 일차원성이 아니고 항목에 대한 반응에 다른 요인이 영향을 준다면 잠재적 특질이 같은 집단 내에서도 그 다른 요인의 변산에 의한 항목간 상관계수가 유발될 수 있다. 따라서 척도의 일원성은 지역독립성을 증명하는 준거가 된다.

물론 항목반응이론을 적용하여 척도개발을 하는 경우에도 연구 초기에는 심리측정적인(psychometric) 속성이 우수한 항목과 그렇지 못한 항목을 동시에 취합하여 그 중에서 우수한 항목을 선별하게 되므로 척도 개발의 초기에 일차원성인 항목들만을 확보해서 항목반응이론을 적용한다는 것은 실제적으로는 비현실적인 일이다. 따라서 척도의 일원성이라는 조건은 항목반응이론에 의해 최종적으로 확정된 척도가 충족해야할 조건이라고 사료된다.

지역독립성과 함께 IRT의 중요한 다른 특징은 문항모수 불변성 (item parameter invariance)이다. 즉 항목계수들(a, b, c)의 추정치가 표집단의 성격에 따라 좌우되지 않는다는 것이다. 예를 들어서 만약 한 항목에 KR를 할 확률이 0.50인 잠재적 특질 집단이 표집에 포함되어 있지 않으면 그 항목난이도는 표집밖에서 결정된다. 물론 이 경우 산출되는 난

이도는 수식에 입각한 이론적 난이도이며 실증적 난이도는 아니다. 이 이론적 난이도가 타당한가를 결정하기 위하여는 그 난이도를 포함하리라고 예측되는 표집을 추출하여 검증할 수 있다. 그러나 중요한 것은 이 이론적 난이도는 표집의 구속을 받지 않는다(sample free)는 것이며, 고전적 측정 이론에서의 결함을 극복하는 장점이라는 사실이다. 고전적 측정 이론에서는 항목난이도가 표집의 성향에 좌우된다.

항목정보와 검사정보 (Item and Test Information)

(1) 서론

정보는 어떤 事象의 불확실성에 의해 정의할 수 있다 (박광배, 인쇄중). 예를 들면, 어떤 사상이 일어날 불확실성이 적을수록 우리는 그 사상에 대하여 더 많은 정보를 갖는다고 할 수있다. 이러한 일반적 개념을 IRT에서 어떻게 도입해서 사용할 수 있는지를 보여주기 위해서 통계적 추론 과정을 고려해보자. 모집단 혹은 이론적인 모수에 관해 우리의 불확실성은 표본추정치의 표집분산 혹은 모수를 위한 신뢰구간의 폭에 의해 표현될 수 있다. 만일 모집단 평균(μ)을 추정하는 경우 표준오차가 적을수록 우리의 정보는 더 커지게 된다. 그러므로 모수에 관한 정보를 표준오차제곱의 역함수로 정의할 수 있다.

$$I_X = 1/\sigma_x^2$$

정보는 IRT에서 중요한 개념인데 왜냐하면 IRT 절차를 사용하여 개발한 척도는 그것들이 제공하는 추정치의 정확성 면에서 평가될 수 있기 때문이다. $\hat{\theta}$의 표준오차의 제곱의 역함수, 즉 어떤 항목이 가지는 정보는 산출된 $\hat{\theta}$의 정확성을 측정하는데 사용된다. 문항이나 검사가 가지는 정보는 모든 θ의 사람에 대해서 반드시 상수는 아니다. 특정한 θ값에서의 정보는 θ추정에 사용된 문항의 수와 속성에 달려있다. 예를 들면, 거의 모든 고능력 피험자가 올바르게 답변할 수있는 쉬운 문항은 높은 수준의 θ에서는 잠재 특성에 관한 정보를 거의 제공하지 못한다. 그러나 이 문항들은 낮은 수준의 θ에서는 잠재 특성에 관해 상당한 양의 정보를 제공할 수 있다.

(2) 항목정보와 검사정보 (Item and test information)

항목 반응을 얻고 이 반응을 $\hat{\theta}$추정치로 전환하기 전에는 개인의 특성에 관해서 우리는 전혀 정보를 갖고 있지 않다. 항목 반응들로부터 $\hat{\theta}$를 추정한 후 비로소 θ의 위치에 관해서 얼만큼의 정보를 갖고 있는지를 평가할 수 있다. 모집단 평균에 관한 정보가 $1/\sigma_x^2$에 의해 주어지는 것과 마찬가지로, θ에 관한 정보는 $I(\theta)=1/\sigma^2(\hat{\theta}|\theta)^2$ 즉 θ의 추정치 $\hat{\theta}$의 표준오차의 제곱의 역이다. 추정된 $\hat{\theta}(\theta)$의 표준오차는 동일한 θ값을 갖는 모든 사람들에 대한 $\hat{\theta}$의 분포의 표준편차로서 해석된다. 정보와 표준오차의 제곱 사이의 관계를 사용하여 $\sigma(\hat{\theta}|\theta)$ 가 $1/\sqrt{I(\theta)}$ 임을 알 수 있다. Birnbaum(1968) 과 Lord (1980)는 검사정보함수 $I(\theta)$를 동일한 특성 θ를 갖는 피험자에 대한 最大尤度 특성추정치 $\hat{\theta}$의 점근적 표집분산의 역으로 정의함으로써 검사 정보에 대한 공식을 얻는 문제에 접근하고 있다. 이것이 의미하는 바는 문항수가 커질수록 $1/\sqrt{I(\theta)}$는 최대우도법을 써서 θ를 추정하는 경우 $\hat{\theta}$의 표준오차이다.

검사 정보를 이용하여 할당 오류 확률을 결정하는 방법

여기에서 기술된 개념적 방법은 Hulin, Drasgow, and Parsons (1983) 에 설명된 내용에 기초하고 있다. IRT의 검사정보개념을 이용하여 피험자를 개별적 집단에 할당하는 경우 생길 수 있는 할당오류의 정도를 미리 알 수 있는 척도를 개발하는 일반적인 절차는 다음과 같다. 우선 연구자는 피험자를 분류하기 위해 기준이 되는 추정된 $\hat{\theta}(\theta)$값을 정한다. 예를 들면 피험자의 θ가 0.50이상이면 그 피험자는 특정한 실험 집단에 할당되고, 그렇지 않으면

---

2)기호 $\sigma^2(\hat{\theta}|\theta)$는 실제 특질의 정도가 θ인 사람의 산출된 θ가 가지는 변량을 의미한다.

그 실험 집단에 할당되지 않는다고 가정하자. 그런데 $\hat{\theta}$가 실제 θ와 언제나 일치하는 것은 아니므로 실제 θ는 0.50이하이더라도 $\hat{\theta}$가 0.50이상으로 산출되어 그 집단에 할당되는 피험자가 생길 수 있다. 이것이 할당 오류이다. 예를 들어 연구자는 실제의 θ값이 평균보다 이하인 피험자가 실험집단에 분류되는 것에 대하여 특별한 우려를 가지고 있다고 가정하자. $\hat{\theta}$는 전체 피험자의 평균이 0이 되고 표준편차가 1이 되도록 표준화된다. 그런데 어떤 응답자의 산출된 $\hat{\theta}$가 0인 경우 즉 평균치와 일치하는 경우 우리는 산출된 $\hat{\theta}$의 표준오차 $\sigma(\hat{\theta}|\theta)$를 이용하여 그 응답자의 실제 θ에 대한 신뢰구간을 구성할 수 있다. 이 신뢰구간을 구성하면서 동시에 실제 θ값이 0인 피험자에게서 산출된 $\hat{\theta}$가 0.50보다 크게 나올 확률을 결정하는 것이 할당오류를 결정하는 문제의 핵심이다. 그것을 위하여 우선 산출된 $\hat{\theta}$값이 0일 때의 각 항목의 항목정보값을 다음의 공식에 준해서 산출한다(Hulin, Drasgow & Parsons, 1983, p.34). 이 공식은 2모수 모델에서의 항목의 정보값을 산출하는 공식이며 3모수, 1모수를 사용할 경우의 항목정보값 산출 공식은 다르다.

$$I(\theta, ui) = \frac{D2ai2\ EXP\ -Dai(\theta - bi)}{1 + EXP\ -Dai(\theta - bi)\quad 2}$$

θ = 잠재적 능력 혹은 특질

I(,ui) = i번째 항목이 잠재적 특질 에서 가지는 항목정보값

D = 1.702

ai = 항목 i의 변별 계수

bi = 항목 i의 난이도 계수

위의 공식에 의하여 산출된 $\hat{\theta}$값이 0 인 지점에서의 각 항목에 대한 정보값이 정해지면 그것을 모든 항목에 대하여 합하므로서 검사정보값이 산출된다. 이 검사정보값의 역수를 취하고 제곱근을 구하면 그것이 산출된 $\hat{\theta}$값이 0 일 때의 표준오차이다. 이 표준오차를 이용하여 산출된 $\hat{\theta}$가 실제 θ를 중심으로 정상분포한다는 가정하에 실제 θ가 0 인 사람의 산출된 $\hat{\theta}$가 0.50 보다 클 확률을 추정한다.

방 법

피험자

1차 표집은 서울의 강남, 강북에 위치한 남,녀 3개 고등학교 1, 2, 3학년생 718명에게 질문지를 실시하였다. 이들 중 남자가 375명, 여자가 343명이었다. 조사 대상자 중 1 학년은 279명, 2학년은 287명, 3학년은 142명이었다(10명은 학년에 대한 응답을 하지 않았다). 2차 표집은 지방대학교에서 무선 표집된 학과의 학생들 554명에게 질문지를 실시하였다. 이들 중 남학생이 296명, 여학생이 258명이었다. 조사 대상자 중 1학년은 44명, 2학년은 123명, 3학년은 329명, 4학년은 56명이었다(2명은 학년에 대한 응답을 하지 않았다). 3 차 표집은 서울 소재 고등학교에서 609명의 학생들에게 질문지를 실시하였다. 이들 중 남자가 298명, 여자가 311명이었고, 1학년은 204명, 2학년은 200명, 3학년은 204명이었다(1명은 학년에 대한 응답을 하지 않았다).

조사도구

질문지는 각 개인의 배경 자료 및 Beck Hopelessness Scale(BHOP;Beck & Weissman, 1974)을 신 민섭과 동료들(1990)이 번안한 절망척도가 사용되었다. 절망 척도는 미래에 대한 부정적이고 비관적인 기대를 측정하고자 만들어진 20개 항목의 眞爲形 질문지이다. 이 척도는 절망이 우울, 특히 자살 의도의 중요한 요인이라고 생각하는 연구자들(Beck, Kovacs & Weissman, 1974)에 의해 관심이 모아지고, 그 후 Beck, Weissman, Lester 와 Trexler(1974)가 미래에 대한 부정적인 기대를 측정하고 계량화할 척도를 개발하고자 하는 노력에서 만

들어졌다.

이 항목들을 정교화하기 위해 우울한 집단과 우울하지 않은 집단을 무선 표집하여 이 척도를 실시하였다. 그리고 몇몇 임상가들이 문항의 안면 타당도(face validity)와 이해할 수 있는 정도를 살펴보았다. 이러한 절차를 거쳐서 최종 11개의 眞 문항과 9개의 僞 문항이 구성되었다. 절망 척도의 신뢰도를 검증하기 위해 Beck, Weissman, Lester와 Trexler (1974)는 최근 자살 시도를 한 경험이 있는 294명의 환자들에게 척도를 실시하였다. 실시 결과 내적 일관성 신뢰도 계수가 .93이었다. 절망에 대한 전체적인 임상 평정과 절망 척도 점수간의 상관관계도 정신질환자(r=.74)와 자살 환자들(r=.62)에게서 높았다. 마찬가지로 절망 척도 점수와 Beck의 우울증 검사(Beck Depression Inventory)의 비관적인 항목들로부터 산출된 점수간의 상관관계(r=.63)와 절망 척도 점수와 Stuart Future Test 점수간의 상관관계도 긍정적(r=.60)이었다. 신 민섭과 동료들(1990)은 고등학생의 자살 성향에 관한 연구를 자살 생각, 우울, 절망의 관계를 통해 밝히고자 하였다. 이 연구에서 저자들에 의해 번안된 절망 척도의 신뢰도 계수(Cronbach's α)는 .81이었고, 자살 생각 점수와 절망 척도 점수간의 상관계수는 .40이었다. 박 광배와 신 민섭(1991)은 우울 척도, 자살에 대한 생각을 측정하는 척도, 절망 척도등을 사용하여 고등학생의 스트레스와 자살 생각간의 관계를 규명하였다. 이 연구에서 저자들은 자살 생각 점수와 우울증 점수들의 상관관계를 보았다. 신 민섭과 동료들(1990)에 의해 번안된 절망 척도점수와 "살아야 하는 이유" 척도(Reason For Living Inventory) 점수간의 상관관계는 .62였고, Beck의 우울증 검사 점수와의 상관관계는 .47, 아동 우울증 검사(Children's Depression Inventory)와의 상관관계는 .44, 자살 생각(Scale for Suicide Ideation) 점수와의 상관관계는 .37이었다. 번안된 절망 척도의 타당도 계수가 전반적으로 지나치게 높지 않은 것은 구인타당도(construct-related validity)에 있어서는 좋은 지표이다. 준거 관련 타당도(criterion-related validity) 에서 산출된 상관계수와 달리, 구인타당도의 상관계수는 적절히 높아야 하며 너무 높아서는 안된다. 만약 새로운 검사가 특별한 잇점이 없으면서 이전에 사용되고 있는 척도와 높은 상관관계를 갖으면, 새로운 검사는 불필요하게된다 (Anastasi, 1988, p.154). 절망 척도는 절망이 중요한 요인이 되는 다양한 정신질환자 집단에 사용하기에 적절하다. 특히 이 척도는 자살에 관련된 생각, 의도, 행동을 구별하는데 유용하다.

신 민섭과 동료들(1990)에 의해 번안된 절망 척도의 타당도 계수가 원저자들의 타당도 계수에 비해 전반적으로 낮은 것은 번안된 절망 척도가 보고자하는 것은 구인타당도인데 반해 원저자들에 의해 보고된 것은 준거 관련 타당도이기 때문일 가능성이 높다. 원저자들은 절망척도의 타당도를 임상 집단에서 본질적으로 같은 성격을 가지는 변인과의 상관관계에 의해 살펴본 반면 번안된 절망 척도의 타당도는 청소년 집단에서 우울증 혹은 앞으로의 자살 생각과의 관련성을 파악한 구인타당도이고, 이것은 다소 성격을 달리하는 변인과의 상관계수이므로, 전반적으로 준거타당도에 비해 구성 개념 타당도가 낮은 것을 번안된 절망 척도가 타당하지 못한 것으로 해석하기는 힘들다. 앞으로 후속 연구에서는 번안된 절망 척도를 사용하여 임상 집단을 준거로 절망 척도의 준거타당도를 살펴보아야 할 것이다.

절차

1차 표집과 3차 표집은 해당 고교의 담임 교사가 학급 단위로 수업 시간에 학생들에게 실시하였고, 2차 표집은 해당 학과 학년 수업 시작전이나 종료 10분전에 연구자가 실시하였다. 질문지에 대한 응답을 완성하는데는 약 8분 정도 소요되었다.

결 과

절망 척도의 기술통계적 구조

각 항목의 평균과 표준편차는 표 1에 제시되었다.

항목반응이 0과 1로 입력이 되었으므로 표 1에 제시된 각 항목 평균은 그 숫자를 1에서 빼면 고전적 측정이론에서 의미하는 항목난이도이다. 총 응답자 중 결측치를 제외한 1619명의 절망 총점(number-right score)의 평균은 5.261이고, 표준편차는 4.141이었다. 절망 점수의 분포는 각 집단에 따라 달라질 수 있는 까닭으로 본 연구에서 사용한 표집에서 각 하위표집의 중앙치와 상하위 25% 에서의 총점을 살펴보았다(표는 이연희(1992) 논문 참고). 각 하위집단에 따라 절망 점수의 분포가 큰 차이가 나지 않으므로 본 연구에서는 표집 전체를 포괄하여 청소년 전집을 대표하는 것으로 가정하고 분석하였다. 총점의 전체적인 분포는 정적으로 편포되었다. 분포가 정적으로 편포된 것은 표 1에 제시되었듯이 거의 모든 항목난이도가 일률적으로 높은 편에 속하기 때문이다. 난이도가 높은 항목들로 구성된 척도는 특질의 정도가 강한 응답자들을 서로 잘 변별하지만 특질의 정도가 약한 응답자들은 서로 잘 변별하지 못하는 성향을 가지게 된다.

1619명의 자료에 기초한 번안된 절망 척도의 신뢰도 계수(Cronbach's α)는 0.87 이었다. Beck (1974)은 절망척도를 294명의 자살시도를 한 피험자에게 실시하여 주성분분석을 한 결과 1보다 큰 eigenvalue를 가지는 세 요인을 추출하였다. 첫번째 요인은 총변량의 41.70%를 설명하였고 두번째 요인은 6.20%, 세번째 요인은 5.60%를 설명하였다. 본 연구를 위하여 수집된 1619명의 자료를 요인분석하였다. 주성분 분석으로 3개의 요인을 추출한 결과 첫번째 요인의 eigenvalue는 6.04이었고 두번째 요인의 eigenvalue는 1.48, 세번째 요인의 eigenvalue는 1.05이었다. 첫번째 요인은 총변량의 30.20%를 두번째 요인은 7.40%를 세번째 요인은 5.20%를 설

표 1. 항목의 평균, 표준편차, 난이도 계수 및 변별도 계수 및 추정된 정보값

항목	평균	표준편차	난이도계수(b)	변별계수(a)	추정된 정보값
01	.23	.45	0.60	1.48	1.16
02	.10	.30	1.44	1.20	.26
03	.18	.38	1.71	0.52	.12
04	.49	.50	-0.16	0.05	.00
05	.43	.50	0.21	0.26	.05
06	.24	.43	0.68	1.88	1.31
07	.13	.33	1.00	2.25	.50
08	.32	.47	0.54	0.96	.60
09	.21	.41	0.87	1.27	.67
10	.25	.43	1.11	0.56	.18
11	.13	.34	1.10	1.70	.46
12	.31	.46	0.50	1.47	1.28
13	.21	.41	0.80	1.62	.91
14	.26	.44	0.63	1.53	1.15
15	.23	.42	0.72	1.82	1.17
16	.10	.30	1.28	1.58	.31
17	.24	.42	0.69	1.79	1.23
18	.35	.48	0.41	1.40	1.21
19	.16	.36	0.95	1.90	.67
20	.09	.29	1.33	1.65	.26

명하였다.[3] 세 요인 모두 eigenvalue가 1보다 크므로 원래의 절망 척도가 일차원성인 척도인지는 불분명하다. 그러나 두번째와 세번째 요인의 eigenvalue가 1 에 근사하므로 비록 명확하게 일차원성이지는 않더라도 근원적인 요인이 많이 내재해 있는 것 같지는 않다. 또한 Scree Plot으로 요인의 eigenvalue를 나타내 요인의 수를 결정하여도 1개의 요인을 명확하게 정할 수 있다(그림 1 참조).

절망 척도에 대한 요인구조의 해석에서 두가지 유념해야 할 사항이 있다. 첫째는, 척도 항목에 대한 반응을 유발하는 잠재적 특질이 연속변인인 반면 반응 자체는 이분화된 (예-아니오) 범주변인인 경우 항목들 사이의 관찰된 상관계수(ϕ계수)가 실제 잠재적 상관계수보다 낮아져서 결과적으로 항목들이 하나의 요인으로 수렴하지 못하고 따라서 실제보다 더 많은 요인 혹은 주성분이 도출될 수 있다는 사실이다. 이 사실에 수반되는 또 하나의 결과는 각 요인 혹은 주성분이 설명할 수 있는 변량이 상대적으로 적어진다는 것이다 (Harman, 1976). 따라서 절망 척도의 경우 첫번째 주성분이 설명하는 변량이 30% 정도로서 비교적 낮은 이유는 반응 범주가 이분화된 항목들로 구성된 까닭일 가능성이 있다. 두번째는, 계수를 요인분석하면 응답자의 특질을 반영하는 요인 다음으로 산출되는 요인은 응답자의 성향을 반영하는 요인이 아니라 항목난이도를 반영하는 경향이 있다 (Carroll, 1961). 이 점을 확인하기 위하여 표 1에 제시된 항목 평균과 위의 주성분 분석에서 산출된 각 항목의 요인부하량 사이의 상관계수를 구해본 결과 항목 평균이 첫번째 주성분과 가지는 상관계수는 -0.41이고, 두번째 주성분과 가지는 상관계수는 0.84이며, 세번째 주성분과 가지는 상관계수는 -0.12였다. 요인 분석 혹은 주성분 분석에서 요인이나 주성분의 의미에 대한 해석은 요인 부하량이 큰 항목과 작은 항목의 성향을 비교하므로서 이루어진다. 두번째 주성분과 항목 평균의 상관계수가 높다는 것이 의미하는 바는 난이도가 높은 항목(평균이 낮은 항목)의 요인부하량이 작고 난이도가 낮은 항목(평균이 높은 항목)의 요인부하량이 높은 것을 의미하므로 보다 합당한 의미 해석이 존재하지 않는 상황에서 두번째 주성분은 항목 난이도를 대표하는 주성분이라고 해석될 수 있다. 따라서 첫번째 주성분과 세번째 주성분이 아마도 응답자의 속성을 나타내는 것으로 보여진다. 그런데 세번째 주성분은 eigenvalue 및 설명할 수 있는 변량이 극히 미미하므로 실질적인 의미에서 원래의 절망척도가 일차원성인 척도라고 간주해도 좋을 듯하다. 표 2에 제시된 요인구조도 이 점을 뒷받침해 준다. 표 2의 요인구조를 보면 첫번째 주성분에 걸리는 모든 항목의 부하량이 균일하게 크므로 첫번째 주성분이 일반요인(general factor)이라고 보아지는 반면, 세번째 주성분의 요인부하량은 한 두개의 항목을 제외하고는 실질적인 의미를 부여하기에 지나치게 작다. 그러므로 만약 척도 개발을 위한 분석이라면 세번째 주성분에 요인부하량이 큰 두 항목을 삭제하는 것이 바람직할 것이다.

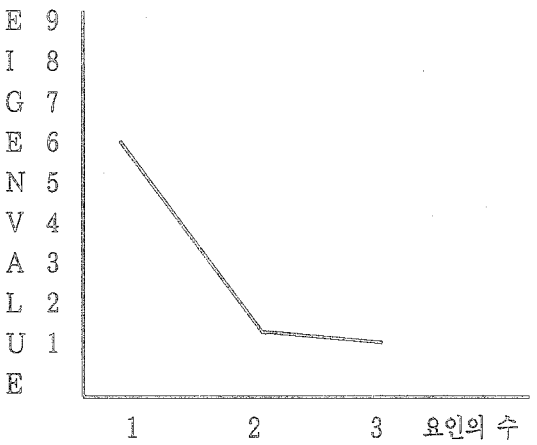

그림 1. 주성분분석에 의한 Scree Plot

결론적으로, 원래의 절망척도의 요인 구조는 비록 명확하지는 않지만 그 척도가 일차원성인 척도라는 것을 시사한다. 다만 첫번째 주성분이 설명하는 변량

3) 주성분분석을 한 이유는 원저자 Beck 의 요인분석 결과와의 비교를 위해서이다.

표 2. 절망 척도의 요인 구조

항 목	주성분 1	주성분 2	주성분 3
01	.59677	.36493	.04189
02	.47994	-.31001	.26907
03	.31353	-.12064	.70694
04	.31680	.38095	-.12688
05	.33982	.38246	.27297
06	.65161	.25940	.09418
07	.64738	-.18429	-.05052
08	.51882	.31423	.14386
09	.56434	-.15333	-.11736
10	.37546	.03165	.38212
11	.59183	-.39046	-.07679
12	.59565	-.02161	-.25317
13	.61224	.12430	-.20174
14	.60984	-.15160	-.20223
15	.61737	.26007	-.01251
16	.52985	-.47607	-.03791
17	.63171	.03266	-.10409
18	.58526	.31023	-.15577
19	.62833	-.11069	-.01262
20	.55992	-.35822	.04799
eigenvalue	6.04	1.48	1.05
변량 %	30.2	7.4	5.2

이 비교적 적은 것에 대하여 그 이유가 이분화된 반응 범주 때문인지 아니면 미국에서 개발된 절망 척도가 한국인의 절망감을 잘 반영하지 못하기 때문인지에 관하여 후속 연구가 수행될 필요가 있다.

계수 산출

IRT에 의해 산출된 각 항목난이도 계수와 변별 계수도 표 1에 제시되었다. 각 항목난이도 계수와 변별 계수 산출은 SYSTAT 컴퓨터 프로그램에서 2모수 모델을 사용한 결과이다. 2모수 모델의 사용은 본 자료는 추측에 의해 정답을 할 필요가 없으므로 추측계수 c를 상정하지 않기 때문이다. SYSTAT에서는 각 항목의 모수 추정을 위하여 단계(Step) 와 허용도(Tolerance)를 지정해 줄 수 있다. 각 단계(Step)는 두 개의 절차(Stage)로 구성된다. 절차 1(Stage 1)에서는 각 항목의 모수치는 고정되고 θ가 추정된다. 절차 1의 끝에서 0의 평균과 1의 표준편차를 갖는 잠정적 θ가 추정된다. 절차 2(Stage 2)에서는 가장 최근의 θ를 고정시키고 한 번에 한 항목씩 항목 모수치를 추정한다. 만약 필요하다면 새로운 단계(Step)가 반복된다. 허용도란 앞서 단계에서 산출된 추정치와 이후 단계에서 산출된 추정치의 차이의 허용한계를 말한다.[4] 그 차이가 0.10보다 작아지면 추

정치가 수렴한 것으로 간주하여 순환단계를 멈추고 최종적인 추정치들을 결정한다. 절망척도에선 각 항목의 모수치를 추정할 때 처음에는 단계는 낮고, 허용도 수준은 크게하였고, 이후 단계는 높히고 허용도 수준은 줄였다. 그 이유는 소위 국지적 최소(local minima)의 위험을 회피하기 위해서였다. 국지적 최소개념은 Hulin, Drasgow & Parsons(1983)에 자세히 설명되어 있다. 처음에는 $\hat{\theta}$를 산출하고 그런 후 다시 $\hat{\theta}$를 고정하고 항목계수들을 산출하므로서 1단계를 마친다. 이러한 절차에 의해 최종 추정치들을 산출하기 위하여 8개의 단계를 순환하였고 이 때 이용된 허용도는 0.10이었다. 항목의 수가 40개 이상이면 일반적으로 허용도를 좀 더 낮출 수 있고, 이것은 추정치들의 정확도에 기여한다. 그러나 절망 척도는 항목 수가 20개 뿐이어서 수차례의 분석을 반복한 결과 허용도 0.10이 적합하다는 결론에 도달하였다.

분류 오류 확률

연구자가 척도의 모든 항목을 이용한다면 이 두개의 할당오류를 합하여 그 확률이 얼마가 되는지 살펴보자. 상위 25%와 하위 25%의 추정된 $\hat{\theta}$값은 0.63과 -0.39이고, 중앙치 값은 0.14였다. 검사정보값 $I(\theta) = 1/\sigma^2(\hat{\theta}|\theta)$ 이고 전체 절망 척도의 검사정보값을 알기 때문에 표준오차값을 계산할 수 있다. 산출된 추정된 θ값이 0.14일 때 전체 절망 척도의 검사 정보값은 표 1에 제시된 정보값을 모두 합한 13.5이고 $I(\theta) = 1/\sigma^2(\hat{\theta}|\theta)$ 이므로 한 개의 표준오차값이 0.27임을 알 수 있다. 전체 절망 척도를 이용하여 연구자가 상위 25%를 절망 집단으로 분류할 때 추정된 $\hat{\theta}$값 0.63은 0.14로부터 1.81의 표준오차만큼 떨어져 있으므로 정상분포 곡선상에서 1.81보다 큰 z값이 산출될 확률은 .0351이고, 따라서 이 때의 분류 오류 확률도 .0351이 된다.[5]

비절망 집단을 분류할 때 오류 확률도 같은 방식으로 산출할 수 있다. 계산된 전체 항목의 중앙치에서의 검사정보값이 13.5이므로 상위 25%에서와 마찬가지로 한 개의 표준오차값은 0.27이다. 연구자가 전체 절망 척도를 이용하여 하위 25%를 비절망집단으로 분류할 때 추정된 θ값 -0.39는 0.14로부터 1.96의 표준오차만큼 떨어져 있으므로 정상분포상에서 1.96보다 큰 z값이 산출될 확률은 0.0250이고, 따라서 이 때의 분류 오류 확률도 0.0250 이 된다. 그러므로 연구자는 전체 항목을 이용하여 상위 25%는 절망 집단으로 하위 25% 는 비절망 집단으로 분류할 경우 0.0601의 할당 오류 확률을 범하게 된다. 그러나 모든 연구에서 연구자가 상위 25%와 하위 25%를 기준으로 집단을 분류하지는 않을 것이므로 본 연구자료에 의해 다른 상정 기준을 사용할 때의 분류 오류 확률을 제시한 것이 표 3에 있다.

표 3. 다양한 분류 기준에서의 오류확률

기준	상한	하한	분류오류확률
10%	.97	-1.21	.0011
15%	.83	-.72	.0063
20%	.71	-.49	.0273
25%	.63	-.39	.0601
30%	.51	-.27	.1496

4) 때로는 허용도가 이론적인 항목특성곡선과 실증적인 항목곡선의 차이로 개념화하기도 한다.

5) 이 논리는 최대우도추정방식에 의해 산출된 계수가 가지는 점진적 정상성의 속성에 근거한다.

사분위 점에 해당하는 척도의 총점

집단 분류를 위하여 척도의 총점이 이용되는 경우를 위하여 본 연구에서 상정한 θ의 사분위 점들이 척도의 총점상에서는 무엇에 해당하는지 알아볼 필요가 있다. 절차는 사분위 점에 해당하는 축소된 척도의 총점을 구하는 절차와 동일하다. 항목 반응 이론에 의해 산출된 추정된 θ값이 0.63일 때의 절망 척도의 회귀 분석에 의해 예측된 총점은 8.10 이며, 추정된 θ값이 -0.39일 때의 절망 척도의 회귀 분석에 의해 예측된 총점은 0.91 이다. 원래 척도로부터 추정된 θ값과 척도의 총점 사이의 상관계수는 0.9577로서 대단히 높으므로 척도 총점의 이 예측치들은 예측 오차가 극히 적은 것으로 사료된다. 그러므로 원래의 절망 척도를 실시하여 총점이 8 혹은 9를 넘으면 절망 집단에 분류하고 0 혹은 1이하이면 비절망 집단에 분류하면 된다. 이 기준들은 본 연구의 자료에서도 실제 사분위점에 근접하였다. 본 자료에서 계산된 정확한 상한 사분위점은 7.16이었고, 하한 사분위점은 1.26 이었다. 회귀 방정식에 의하지 않고 logistic 모형에서 제시된 공식에 상위 25%, 하위 25% 에서의 θ값과 각 항목의 a, b값을 대입하여 산출된 P값을 모든 항목에 대하여 더하면 절망척도에서의 분류 기준으로 사용될 총점을 알 수 있다. 이렇게 계산된 상위 25%에서의 총점은 8.2, 하위 25%에서의 총점은 2.08이었다.

논의 및 결론

본 연구에서는 심리 척도들이 연구 목적으로 쓰이는 경우 피험자의 분류와 관련된 오류 확률을 체계적으로 파악할 수 있도록 항목 반응 이론을 적용하는 절차를 기술하였다. 분류 오류의 확률을 추산하는 것이 가능한 이유는 항목반응이론이 척도 항목에 대한 응답자의 반응에 기초하여 그 응답자의 잠재적 특질의 정도를 추정하기 위한 측정 이론이기 때문이다. 주어진 척도상의 점수를 이용하여 피험자를 강한 특질 집단과 약한 특질 집단 혹은 실험 집단과 통제 집단으로 분류하는 경우에 있어서 실제 특질이 강한 피험자가 약한 특질 집단에 속하게 되는 오류 확률과 실제에 있어서 특질이 약한 피험자가 강한 특질 집단에 속하게 되는 오류 확률을 이론적으로 파악하고자 하였다.

절망 척도 점수상에서 상위 25%는 절망 집단으로 하위 25%는 비절망 집단으로 분류한 후 실제 피험자의 θ값(실제 잠재적 특질의 정도)이 중앙치인 사람이 절망 집단에 잘못 할당되거나 실제 θ값이 중앙치인 사람이 비절망 집단에 잘못 할당되는 것을 할당오류로 가정하였다. 척도의 전체 항목을 이용하여 이상 두 개의 할당오류를 합한 확률을 추정해 본 결과 0.06이 구해졌다. 그러나 상위 25%와 하위 25%를 기준으로 집단을 분류하는 것은 표집의 상대적 기준에 의해 본 연구자가 채택한 것이므로 많은 다른 연구자들이 상정하는 기준은 다를 수 있다. 그런 경우를 위하여 척도의 총점상에서 다른 기준을 적용하여 집단을 분류할 경우 예측되는 분류오류확률을 제시하였다(표 3참조). 본 연구에서 채택한 0.06이라는 분류 오류 확률은 오직 이론적인 오류 확률로서 특정한 연구에서 오류 확률이 언제나 0.06이내일 것이라고 보장하지는 않는다. 그러나 본 연구에서 채택한 것과 같은 기준을 사용하는 연구가 무수히 많이 존재한다고 가정하면 그 오류 확률이 평균적으로 0.06내외가 될 것이다.

또 한가지 언급해야 할 사항은 본 연구에서 사용한 분류 오류의 확률이라는 개념에 관해서이다. 예를 들어 어떤 연구자가 절망 집단에 30명, 비절망 집단에 30명씩의 피험자를 본 연구에서 사용한 분류 기준을 이용하여 할당하였다고 가정하자. 그런 경우 분류 오류 확률이 0.06이라는 것이 선정된 총 60명의 피험자 중 6%에 해당하는 4명 정도가 잘못 분류되었다는 것을 의미하는 것은 아니다. 또한 Hulin, Drasgow, and Parson (1983) 가 기술한 방법에 의하여 추정된 0.06 이라는 확률수치가 중앙치 이하인 불특정인이 상위집단에 잘못 분류되고, 중앙치 이상인 불특정인이 하위집단에 잘못 분류될 누적확률(cumulative probability) 로서 해석하기에도 곤란

하다. 분류 오류 확률 0.06이라는 개념은 앞서 자세히 언급하였듯이 실제 절망의 정도가 중앙치에 해당하는 피험자가 절망 집단, 비절망 집단에 잘못 할당될 확률이다. 다시 말하면 중앙치의 절망 정도를 가지는 어떤 한 개인이 두 집단 중 하나에 포함될 가능성을 의미한다. 따라서 이 확률 개념을 빈도 개념으로 전환하거나 누적확률개념으로 해석하기는 어렵다. 그러나 중앙치의 특질을 가지는 사람이 잘못 분류될 확률이 0.06 이라면 우리는 중앙치 이하의 어떤 특정한 사람이 상위집단에 잘못 분류되거나, 중앙치 이상의 어떤 특정한 사람이 하위집단에 잘못 분류될 가능성은 그보다 낮을 것으로 추정할 수 있다. 어떤 척도들에 대하여 이러한 분류 오류 가능성이 알려지면 같은 특질을 측정하는 척도가 여러 개 존재하는 경우 연구 목적에 보다 부합하는 척도를 선정하는 기준으로 이용될 수도 있고, 또한 연구 결과를 일반화하는데 어느 정도의 유보성을 두어야 하는지를 가늠하는데 큰 도움이 될 것이다.

Hulin, Drasgow, and Parson (1983) 가 기술한 방법에 의해서 산출된 분류오류확률의 의미해석이 다소 까다로운 이유는 무엇일까? 그것은 아마도 집단분류에 이용되는 심리척도의 속성과 목적의 불일치에서 기인되는 것으로 사료된다. 본 연구에서 다루어진 심리척도는 연속적인 척도속성을 가지는 것이다. 일반적으로 인위적인 집단분류를 이용하는 심리학 연구들에서도 연속적인 척도속성을 가지는 심리척도를 이용하여 집단분류를 시행하는 경우가 많다. 그러한 경우에 항목반응이론이 유용한 이유는 그것이 잠재특질이 연속적인 차원이라는 가정을 하는 척도이론이기 때문이다. 그러나 그러한 연속적인 척도를 범주적인 집단분류를 위하여 이용하는 경우 척도의 속성과 이용목적간의 불일치에 의해서 개념적인 복잡성이 발생할 수 있다. 좀 더 구체적인 예를 들면, 연속적인 잠재특질의 실제값이 -2.0 인 사람과 -1.8 인 사람은 실제 특질값이 다르므로 '상위집단 (특질값 1.5 이상)' 으로 잘못 할당될 확률이 다르다. 그러나 만약 어떤 잠재특질이 본래부터 범주적이면 실제에 있어 하위집단에 속하는 모든 사람들은 상위집단으로 잘못 분류될 확률이 동일해야 한다. 이렇게 특정한 잠재범주에 속하는 모든 불특정인들이 잘못 분류될 확률에 있어서 동일한 경우에는 그 추정된 확률의 의미해석이 훨씬 간편하고 명료하게 될 것이다. 따라서 집단분류에서의 분류오류를 추정하는 방법으로서 항목반응이론에 기초한 Hulin, Drasgow, and Parsons (1983) 의 방법은 완벽하다고 할 수는 없으나 연속적인 심리척도를 이용하여 집단분류를 하는 연구를 위해서는 현재로서는 가장 진보된 방법중의 하나라고 사료된다. 가장 이상적으로는 모든 집단분류를 위해서는 범주적인 척도를 이용하고, 이때 범주적인 잠재특질을 추정하는 측정방법 (예를 들면 잠재계층모형) 을 활용하여 분류오류확률을 추정하는 것이다.

본 연구에서 편의상 절망 집단과 비절망 집단을 상정하고, 그 집단들에 대한 분류 오류 확률을 추론하였는데, 이 때 절망 집단이 임상적인 의미를 가지는 집단은 아니라는 사실을 명백히 할 필요가 있다. 즉 본 연구에서 언급된 분류 오류 혹은 할당 오류가 진단 오류와 혼동되어서는 안된다. 본 연구에서 한국의 고등학생과 대학생을 전집 혹은 모집단으로 간주하고 그 모집단으로부터 표집된 표본에서 상위 25%에 해당하는 집단을 절망 집단, 하위 25%에 해당하는 집단을 비절망 집단이라고 명명하였다. 앞으로 임상적인 의미에서의 절망 집단이 절망 척도에서 구별되는 절단점 (cut-off point) 이 밝혀지면 그 절단점에서의 분류 오류 확률을 추론할 수 있고, 더 나아가서 그 절단점에서 변별력이 높은 항목들을 알아낼 수 있을 뿐만 아니라 그 때 추정되는 분류 오류 확률은 진단 오류 확률이라고 해석될 수 있을 것이다.

마지막으로, 확률 0.06의 오류율을 상정했지만 추정된 θ로부터 총점을 실증적으로 예측할 때 또한 예측오차가 다소 개입되므로 본 논문에서 제시한 총점들을 기준으로 집단분류하는 경우 실제 오류 확률은 0.06보다 다소 클 수도 있다. 그러나 회귀 방정식에 의한 예측오차는 항목 반응 이론에 의해 산출된 θ와 검사 총점 사이의 상관계수가 극도로 높으므로 거의 무시해도 좋은 정도이다. 다만 앞으로의 연구에서는

회귀방정식이 아닌 다른 방법에 의해 기준으로 사용된 θ에 해당하는 검사 총점이 실증적으로 정해질 필요가 있고, 이 때의 예측오차에 의한 분류 오류 확률도 정확히 추정될 필요가 있다.

참 고 문 헌

박광배 (인쇄중). 빈도 분석. 서울 : 성원사

박광배, 신민섭 (1991). 고등학생의 스트레스와 자살생각. 한국심리학회지, 17-29.

신민섭, 박광배, 오경자, 김중술 (1990). 고등학생의 자살 성향에 관한 연구 : 우울-절망-자살간의 구조적 관계에 대한 분석. 한국심리학회지 : 임상, 9, 1-19.

Allen, M. J. and Yen, W. M. (1979). *Introduction to Measurement Theory.* Brooks/Cole Pubilshing Company, Monterey, California.

Anastasi, A. (1988). *Psychological Testing.* New York: Macmillan.

Andrich, D. (1978). A rating formulation for ordered response categories *Psychometrika, 43,* 561-573.

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and diserpersion parameters. *Psychometrika, 47,* 105-113.

Andrich, D. (1988). *Rasch Models for Measurement.* Sage publications. The Publishers of Professional Social Science. Newbury Park Beverly Hills London New Delhi.

Beck, A. T. (1974). The development of depression: A cognitive model. In R. Friedman & M. Katz(Eds.), *Psychology of depression: Contemporary theory and research.* Washington, D.c.: Winston in press.

Beck, A. T., Weissman, A., Lester, D. and Trexler, L. (1974). A measurement of pessimism: The hopelessness scale. *Journal of Consulting and Clinical Psychology, 42,* 861-865.

Beck, A. T. , Kovacs, M. and Weissman, A. (1979). Assessment of suicidal intention: the scale for suicide ideation. *Journal of Consulting and Clinical Psychology, 47,* 343-352.

Birnbaum, A. (1968) Some latent trait models and their use in inferring examinee's ability. In F.M.Lord & M.R. Novick, *Statistical theories of mental test scores.* Reading, Mass.: Addison-Wesley Publishing.

Bollen, K. A. (1989). *Structural Equations with latent variables.* New York:Wiley and Sons, Inc.

Duncan, O. D. (1985). Some models of response uncertainty for panel analysis. *Social Science Research, 14,* 126-141.

Guttman, L. (1950). Chap. 2, 3, 6, 8, and 9. In S. A. Stouffer et al. (Eds.), *Measurement and prediction.* Princeton, N.J.: Princeton University Press.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago press.

Hulin, C. L., Drasgow, F. and Parsons, C. K. (1983). *Item Response Theory: Application to Psychological Measurement.* Dow Jones-Irwin. Homewood, Illinois.

Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics* (Vol.2, 4th

4th ed.). New York: Macmillan.

Lazarsfeld, P. F. (1950). Chap. 10 and 11. In S. A. Stouffer et al. (Eds.), *Measurement and prediction.* Princeton, N. J.: Princeton University Press.

Lazarsfeld, P. F. (1959). Latent Structure analysis. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3). New York : McGraw-Hill, 476-542.

Lord, F. M. (1952). A theory of test scores. *Psychometric Monography,* No. 7.

Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika, 39,* 247-264.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, N. J.: Erlbaum.

Lord, F. M. and Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, Mass.: Addition-Wesley Publishing.

Masters, G. N. (1980). A Rasch model for rating scales. *Dissertation Abstracts International, 41,* 215A-216A.

Masters, G. N. (1988). Measurement models for ordered response categories. In R. Langeheine and J. Rost (Eds.), *Latent Trait and Latent Class Models.* New York and London: Plenum Press.

Molenaar, I. (1983). Item Steps. *Heymans Bulletins Psychologische* (Report No. HB-83-630-EX). Gronigen, Netherlands: Instituten R. U.

Rasch, G. (1960). *Probablistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1972). Objectivity in the social science: A methodological problem (in Danisch), *Nationalekonomisk Tiddskrift, 110,* 161-196.

Samejima, F. (1969). Estimation of latent ability using a response patten of graded scores. *Psychometrika* (Monograph Supplement No. 17).

Tucker, L. R. (1946). Maximum validity of a test with equivalent items. *PsychometriKa, 11,* 1-13.

Wright, B. D., and Masters, G. N. (1982). Rating scale analysis. Chicago: MESA Press.

韓國心理學會誌
Korean Journal of Psychology
1992. Vol. 11. No. 1, 24-40

Expected probability of Misclassification of Subjects into Personality Groups : An Application of Item Response Theory

Yeon Hee Lee and Kwang B. Park

Chungbuk National University

The purpose of the present paper was to demonstrate a method of estimating the probability of misclassification when a Scale is used to classify respondents arbitrarily into the high and low groups based upon the scores on the scale. Determination of misclassification probability is made possible because Item Response Theory provides theoretically sound estimation of latent trait. In this study, misclassification was defined by the hypothetical incidents in which the subjects with the actual degree of traits below the median are assigned into the high group and the subjects with the actual degree of traits above the median are assigned into the low group based upon the scores on the scale. The estimated misclassification probability associated with Beck Hopelessness Scale was .06. The interpretive meaning of the misclassification probability as defined in this study was discussed. Finally, cut-off points on the raw score scale reflecting those probabilities were provided for practical consideration, and important issues for further studies were suggested.