한국심리학회지 : 일반
Korean Journal of Psychology: General
2009, Vol. 28, No. 3, 659-686

급내 상관과 일반화가능도 이론 간의 이론적 및 경험적 비교

김 지 윤 여 성 칠 이 순 묵†

성균관대학교 심리학과 건국대학교 응용 통계학과 성균관대학교 심리학과/인재개발학과

자료의 신뢰도를 추정하기 위해 연구자들은 여러 가지 방법을 통해 신뢰도 계수를 산출한다. 신뢰도를 나타내는 대표적인 방법은 α계수, 급내 상관(intraclass correlation), 그리고 일반화가능도 이론(Generalizability theory)이다. 세 이론 모두 고전검사이론에 기초한 추정 방법으로서 공분산의 모형(α계수), 분산분석(ANOVA)의 모형(급내 상관, 일반화가능도 이론)을 적용하므로 많은 공통점이 있지만, 모두가 서로 다른 가정들에 근거하여 각각의 신뢰도 값을 산출한다. 그럼에도 불구하고 산출된 신뢰도 값이 동일하다는 점에서, 해석상의 문제를 제기한다. 특히 급내 상관과 일반화가능도 이론에 의한 α계수의 해석이 그러하다. α계수는 사용되기 시작한지 60년에 이르는 긴 역사와, 분야를 가리지 않고 보고되는 보편성을 가지고 있다. 그에 비해 상대적으로 복잡성과 포괄성을 더해가는 급내 상관과 일반화가능도 이론은 응용 연구에서 아직은 제한적으로 사용되고 있으며 그 논리가 명확하지 않은 부분이 있다. 특히 급내 상관의 경우, 연구자에 따라 신뢰도에 대한 정의와 해석에서 서로 다른 부분이 확인 되었다. 따라서 본 연구에서는 급내 상관계수의 개념을 서로 다르게 정의한 두 연구를 비교 및 검토하였다. 또한 대부분의 연구에서 신뢰도의 지표로서 제시되는 α계수와 비교하여 급내 상관계수와 일반화가능도 이론에서 산출되는 신뢰도 값들 간의 관계를 개념적으로 정리하였다. 특히 급내 상관과 일반화가능도 이론에 따라 다르게 해석되었던 α계수가 어떠한 효과로 정의된 신뢰도인지 제시하고, α계수의 값이 동일하게 산출되더라도 그 해석은 자료수집의 모형을 반영하여야 함을 보였다. 따라서 급내 상관, 일반화가능도 이론, 그리고 α계수의 관계에 대해 정리를 하고, 실제 자료를 토대로 각각의 신뢰도 계수를 산출하여 세 이론적 개념간 비교를 명확히 하였다.

주요어 : 신뢰도, 신뢰도 α 계수, 급내 상관, 일반화가능도 이론

---

† 교신저자: 이순묵, 성균관대학교 심리학과/인재개발학과, 서울시 종로구 명륜동 3가
Tel: 02-760-0492, E-mail: smlyhl@chol.com

평가 장면에서, 평가자들 사이의 일관성(consistency)과 일치도(agreement)는 행동과학에서 측정 결과로 수집된 자료의 질적 수준에 대해 중요한 지표를 제공한다. 측정의 질적 수준을 나타내기 위하여 측정의 신뢰도라는 용어를 사용하는데, 이것은 일반적으로 심리적인 속성에 대한 평가에서의 일관성으로 간주된다(Anastasi, 1982). 최근에는 기존의 신뢰도 개념에 일치도를 추가함으로서 확장되고 있다. 일관성은 측정치들 사이의 일관성 즉, 측정 대상의 상대적 위치가 두 측정치들간에 유지되는 정도로서, 규준지향(norm-referenced) 신뢰도라고 알려져 있다. 일치도는 측정 대상의 상대적 위치의 개념을 넘어 절대 수준에 부여된 값이 가리키는 절대적 위치가 두 측정치들간에 일치하는 정도로서, 준거지향(criterion - referenced) 신뢰도라고 알려져 있다(McGraw & Wong, 1996). McGraw와 Wong 이전에는 신뢰도라고 할 때 일관성만을 가리키고 일치도는 또 다른 개념으로 사용한 경우도 있다. 그러나 이 글에서는 McGraw와 Wong 이후의 추세에 따라 신뢰도의 범위에 일관성과 일치도를 포함해서 논의하였다.

신뢰도를 산출하는 많은 방법들 중에서도 고전검사이론(Classical Test Theory)에 기초한 α 계수는 응답들간에 공변하는 정도(covariance)에 기초하여 신뢰도를 정의한 것인데, 공분산은 일관성을 추론하는 근거가 되지만 일치도를 추론할 수는 없다. 반면에 급내 상관(Intraclass Correlation)은 다수의 평가자들에 의해 측정된 결과에서 진점수가 차지하는 부분을 알아보기 위하여 분산분석(ANOVA, analysis of variance)모형을 적용하였다. 이것은 다수의 평가자들이 응답자들을 평가한 결과에 대한 일관성과 일치도를 평가하게 해준다. 신뢰도를 산출하기 위한 방법으로 사용되는 또 다른 방법은 일반화가능도 이론(Generalizability Theory)을 이용하는 것이다. 일반화가능도 이론은 급내 상관과 마찬가지로 고전검사이론의 확대 및 발전 형태지만, 급내 상관에 비해서 훨씬 더 다양하게 오차 점수의 원천을 구별해 내고, 각각의 원천이 갖는 상대적인 영향력을 판별해 낼 수 있는 개념적인 틀과 방법론으로서, 일관성과 일치도를 구분해서 제시한다. 신뢰도에 대한 지표로서 1945년 Guttman이 최초로 제시하고(McDonald, 1999) Cronbach의 1951년 논문 이후 널리 쓰이게 된 α계수(Guttman-Cronbach alpha 라고도 불림)는 이미 사용이 정착되었으나, 개념을 확대 및 추가하는 급내 상관과 일반화가능도 이론은 아직 제한적으로 사용하고 있으며 그 논리에 혼동을 주는 부분이 있다. 따라서 이 글에서는 이 두 이론 사이에 유사점과 상이점 및 관련된 이슈를 분명히 하기 위해 두 이론에 대한 개념, 정의, 그리고 가정을 대비하였다.

우선 두 이론사이의 논점은 크게 세 가지인데, 첫째는 두 이론에서 상이한 가정을 가지고 있음에도 그에 근거하여 산출되는 신뢰도 값들이 같다는 것, 둘째로 이 두 확대 이론들의 기본 개념은 α계수에 대하여 상이한 해석을 제시하고 있다는 것, 마지막으로 급내 상관의 개념 내에서 연구자에 따라 상이하게 신뢰도를 정의하고 있다는 것이다.

이 연구에서는 위의 세 가지 논점에 대하여 분석적, 경험적으로 접근하여 혼동의 이유를 밝히고, 궁극적으로는 급내 상관과 일반화가능도 이론에서 각각 산출되는 신뢰도 계수를 사용함에 있어서 명확한 기준을 제시하고자 한다. 첫째로, 급내 상관 이론에서 신뢰도 정의가 연구자에 따라 다른 부분을 정리하여

일관성 있는 기준을 제시한다(연구 1). 둘째로, 대부분의 연구에서 신뢰도 추정을 위해 일반적으로 사용되는 α계수(Cronbach, 1951; Guttman, 1945)의 개념에 비추어, 급내 상관과 일반화가능도 이론의 두 방법으로 추정된 신뢰도 값들의 관계를 논의하였다. 이 논의에서는 분산분석으로 볼 때 2원 모형을 중심으로 정리하였다. 급내 상관의 2원 모형이면, 일반화가능도 이론에서는 단일국면이다. 그 후 실제 자료에 급내 상관, 일반화가능도 이론, 그리고 α계수를 적용하여 신뢰도를 추정한 후, 각 이론들로부터 산출된 신뢰도 값들이 개념적으로 정리한 내용과 동일한 결과를 나타내는 지 확인하였다(연구 2).

고전검사이론에서의 신뢰도

신뢰도(reliability)

오늘날 사용하고 있는 신뢰도의 개념은 고전검사이론(Classical Test Theory)을 기초로 하고 있다. 고전검사이론의 기본 가정은 하나의 검사에서 진점수 T와 오차점수 e의 상관이 0이라는 가정과, 오차점수 e는 상호간에 상관이 없으며 다른 검사의 진점수와도 상관이 없다는 것이다. 또한 고전검사이론은 한 사람의 응답자에게 검사를 2회 시행하여 얻은 관찰점수 X와 X'가 각각 진점수(true score) T와 T', 오차점수(error score) e와 e'의 합으로 이루어진다고 설명하며, 이는 아래 식 (1)로 표현가능하다. 여기서 동일한 응답자 이므로 T = T'이지만, $e \neq e'$이다.

$$X = T + e$$

$$X' = T' + e' \tag{1}$$

특히, 평행검사(parallel test)로 가정된 두 검사에서의 관찰점수 X와 X'는 응답자에 대하여 동일한 진점수(T=T')를 가지며, 응답자 집단으로 보면, 동일한 오차분산($\sigma_e^2 = \sigma_{e'}^2$)을 가진다는 조건을 만족한다. 또한 관찰점수의 분산($\sigma_X^2 = \sigma_{X'}^2$)이 동일하다는 의미도 포함한다. 여기서 하나의 개념에 대한 두 번의 관찰 X와 X′간에 상관이 높다면 그 검사는 신뢰할 수 있다고 간주되므로, 신뢰도는 두 평행검사의 관찰점수간의 상관계수로 정의된다(Allen & Yen, 1979). 이 때 관찰점수 X와 X′의 공분산은 아래 식 (2)가 되고 상관은 식 (3)과 같다.

$$\mathrm{Cov}(X, X') = \mathrm{Cov}(T + e, T' + e') \tag{2}$$

$$= \sigma_{TT'} + \sigma_{Te'} + \sigma_{T'e} + \sigma_{ee'} = \sigma_T^2$$

$$\rho_{XX'} = \frac{\mathrm{Cov}(X, X')}{\sigma_X \sigma_{X'}} = \frac{\sigma_T^2}{\sigma_X \sigma_{X'}} \tag{3}$$

$$= \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2} = \text{신뢰도의 정의}$$

식 (3)으로 알 수 있듯이, 평행한 두 검사간의 상관계수는 측정 결과가 관찰대상의 진점수를 반영하는 정도를 의미한다(Allen & Yen, 1979). 즉, 신뢰도의 정의를 보여주고 있다.

신뢰도 α계수

신뢰도 α계수는 검사의 내적 일관성을 나타내는 값으로서, 한 검사를 이루고 있는 모

든 문항들이 동일한 잠재특성을 측정하고 있다는 가정에 근거한다(Cronbach, 1951; Guttman, 1945). Cronbach은 그가 활동하던 시기에 이 지수의 앞에 자신의 이름이 붙는 것에 부담을 표명하기도 하였다. 그래서 검사이론가들 사이에서는 Guttman-Cronbach 알파 또는 그냥 신뢰도 "알파계수"로 부른다(McDonald, 1999).

만약, 두 개의 평행한 검사를 시행하여 얻은 관찰점수 X와 X'가 문항점수(Y_i)들의 합으로 이루어졌다고 가정하면, 식 (1)은 식 (4)와 같이 재표현 할 수 있으며, α계수의 정의 식은 아래 식 (5)로 표현가능하다.

$$X = Y_1 + Y_2 + \ldots + Y_n \qquad (4)$$
$$X' = Y'_1 + Y'_2 + \ldots + Y'_n$$

$$\alpha\text{계수} = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N}\sigma^2_{Yi}}{\sigma^2_X}\right) \qquad (5)$$

평행성(parallelism)에 근거한 신뢰도의 정의식 (3)과 α계수의 식 (5)에서, 식 (6)이 도출된다(Allen & Yen, 1979, p.79 참조).

$$\frac{\sigma^2_T}{\sigma^2_X} \geq \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N}\sigma^2_{Yi}}{\sigma^2_X}\right) \qquad (6)$$

즉, α계수는 신뢰도의 하한값(lower bound)이다. 식 (4)에서 각 검사에 포함되는 문항점수(Y_i)들이 평행일 때, 식 (6)의 등호(=)가 성립하게 된다. 즉, α계수는 1) 두 검사 X와 X'가 평행하고 2) 두 검사에 포함되는 각각의 문항점수(Y_i)들이 평행할 때 산출되는 신뢰도이며, 그 외의 경우에는 실제 신뢰도보다 작은 값이 산출된다.

일반화가능도 이론에서의 신뢰도

개념

식 (1)의 오차점수는 여러 오차들이 합쳐져 있음에도 구별되지 않는 하나의 값으로 표현되고 있으므로, 식 (3)의 정의로는 측정상황에서 발생할 수 있는 많은 오차 요인에 대한 설명이 불충분하다. 이러한 고전 검사 이론의 한계는 진점수와 여러 오차의 합으로 검사점수를 표현하는 일반화가능도 이론에 의하여 어느 정도 해결 되었으며, 이는 식 (7)과 같이 표현가능하다. 즉, 일반화가능도 이론에서는 여러 오차의 원인을 고려하여 그에 의한 영향을 분리하여 구별하는 것이 가능하다.

$$X = T + (e_1 + e_2 + \ldots + e_k) \qquad (7)$$

일반화가능도 이론에서는 측정 결과를 일반화시키기 위한 대상전집을 나타낼 때 고전검사이론에서 사용하는 모집단의 개념이 아닌 전집이라는 개념을 적용한다. 식 (7)에서 e_1에서 e_k까지의 오차원인은 국면(facet) 이라고 하고, 각 국면에서 모든 가능한 수준(예: 문항수효가 하나의 국면이라면 모든 가능한 문항수효가 문항 국면의 전집)을 전집이라고 한다. 국면내 '수준'을 '조건'이라고도 한다. 여기서 측정의 한 국면이란, 측정 조건의 집합을 의미하며, 분산분석 설계에서 요인(factor)과 유사한 개념이다(김성숙, 1993). 특히 전집점수라는

용어는 고전검사이론에서의 진점수와 유사한 개념으로 일반화 대상이 되는 국면의 전집에서 어느 한 측정대상이 가지는 평균점수를 의미한다(김성숙, 김양분, 2001). 또한 고전검사이론에서는 오차점수를 오직 상대평가에서의 오차 개념으로 고려하지만, 일반화가능도 이론에서는 상대평가와 절대평가에서의 오차를 구분하여 상대오차, 절대오차라고 한다. 상대평가는 '관찰대상 간의 차이 파악'에 관심이 있는 반면, 절대평가는 '관찰대상인 행위 자체의 수준'에 관심이 있다. 일반화가능도 이론에서 상대오차를 고려하여 산출되는 신뢰도 추정치인 일반화가능도 계수(generalizability coefficient)는 고건검사이론에서의 신뢰도와 비슷한 개념이며, 전집점수(universe score) 분산을 전집점수 분산과 상대오차분산(relative error variance)의 합으로 나눈 값이다. 절대평가에서의 오차를 고려하여 산출되는 의존가능도 계수(dependability coefficient)는 전집점수 분산을 전집점수 분산과 절대오차분산(absolute error variance)의 합으로 나눈 값이다.

일반화가능도 계수 (8)

$$= \frac{\text{전집점수분산}}{\text{전집점수분산} + \text{상대오차분산}}$$

의존가능도 계수 (9)

$$= \frac{\text{전집점수분산}}{\text{전집점수분산} + \text{절대오차분산}}$$

절대오차란 특정 응답자에 대한 관찰점수가 전집점수에 비해서 차이나는 정도이므로, 절대오차분산은 측정대상의 관찰점수와 전집점수간 차이에 대한 분산으로서 응답자 분산을 제외한 모든 분산성분의 합으로 표현된다. 반면 상대오차는 특정 응답자의 관찰점수와 그 응답자가 속한 집단 평균과의 차이로서 개념적으로는 응답자간 차이와 같다. 상대오차분산은 관찰점수와 전집점수 평균간 차이 즉, 상대적 위치 차이의 분산으로서, 응답자 요인과 상호작용하는 요인들의 분산합으로 표현된다(김성숙, 김양분, 2001, p. 18 참조). 상대오차분산에 포함되는 분산성분의 수가 절대오차분산에 포함되는 분산성분의 수보다 작기 때문에 일반적으로 상대오차분산은 절대오차분산보다 작다. 따라서 동일한 자료에서 일반화가능도 계수의 값은 의존가능도 계수의 값보다 크게 된다. 또한 이 두 계수는 검사점수가 절대평가에 사용되는지, 상대평가에 사용되는지에 따라 해석이 구분된다. 즉, 특정 응답자의 점수를 점수가 산출된 영역에 대비하여 해석하기 위해서는 절대오차 분산을 사용하여 계산되는 의존가능도 계수가 사용되고, 특정 응답자의 점수를 그 응답자가 속한 집단의 점수와 비교하여 해석하기 위해서는 상대오차분산을 사용하여 계산되는 일반화가능도 계수가 사용된다.

일반화가능도 이론에서는 일반화가능도 연구(G연구)와 결정연구(D연구)로 구별된다. G연구의 목적은 측정의 분산원천에 대해 가능한 많은 정보를 제공하는 것이다. G연구를 통해 연구 설계모형이 교차(crossed)모형인지 내재(nested)모형인지를 결정하여 분산분석모형을 적용하고, 이를 통해 분산성분을 추정하여 오차분산성분의 상대적 크기를 비교하고 각 오차원의 영향력을 분석할 수 있다. G연구 설계에서는 단일국면에서부터 2국면이상 다 국면 설계에 이르기까지 범위가 다양하다. D연구에서는 G연구를 통해 얻어진 결과를 토대로 이용한다. D연구에서는 연구자의 관심에

따라 일반화하고자 하는 국면의 수와 폭을 정의한 후, 측정이 상대적인 것인지 절대적인 것인지를 명백하게 하여 오차의 원천을 파악한다. 또한 D연구에서 산출해내는 두 가지의 신뢰도들을 검토함으로써, 오차를 최소화하고 신뢰도를 최대화하기 위한 각 측정국면의 조건을 파악할 수 있고, 연구 목적에 맞는 적정 수준의 신뢰도계수를 산출하기 위한 최적의 조건을 제공한다. 즉, 일반화가능도 이론은 연구자의 의도에 따라 신뢰도를 추정할 수 있는 모형을 다양하게 설계할 수 있다.

본 연구에서는 다양한 일반화가능도 이론의 모형 중에서 급내 상관과의 비교가 가능한 단일국면 교차설계를 설명하도록 한다.

단일국면 교차설계 : p × i 설계

여기서 p는 “person”을, i는 “item”을 나타낸다. 즉 측정대상인 응답자들에게 문항을 주어 반응을 구하는 설계이다. 측정오차를 고려하기 위해 사용된 국면은 ‘문항’으로서, 단일국면 설계이다. 그러나 측정대상도 하나의 국면으로 보면, 응답자 국면의 모든 조건들이 문항 국면의 모든 조건들과 함께 관찰되는 ‘교차(crossed)’된 측정이므로, p×i 설계로 표기되었다. 이 설계에서 교차되는 조건(cell)의 표본 크기는 1이다. 심리학에서 사용되는 일반적 실험설계와 달리 세부조건(cell)의 표본 크기가 1이 되는 것은 각 문항을 하나의 검사로 간주한다는 것을 의미한다. 단일국면 교차설계는 표 1과 같이 표현 가능하다.

단일국면 교차설계에는 분산의 원천이 4가지가 있다: 측정대상(표 1의 p, 응답자) 간의 분산(σ_p^2), 문항들 간의 분산(σ_i^2), 응답자와 문항의 상호작용 분산(σ_{pi}^2), 파악되지 않은 오차 분산(σ_e^2). 이 중 응답자와 문항의 상호작용 분산과 파악되지 않은 오차분산은 단일국면 교차설계에서는 구별되지 않는다. 즉, 하나의 변산원(pi,e)이 되고 분산도 $\sigma_{pi.e}^2$로 표시한다. 그 이유는 응답자 국면과 문항 국면간에 교차되는 조건에는 한 번의 관찰밖에 없기 때문이다. 따라서 이 두 가지의 분산을 묶어 잔차분산이라 부른다. 교차설계에 대한 벤다이어그램과 응답자가 하나의 문항에 응답한 점수(X_{pi})는 그림 1과 식 (10)으로 표시한다.

$$
\begin{aligned}
X_{\pi} &= \mu & &[\text{총평균}] \\
&+ \mu_p - \mu & &[\text{응답자효과}] \\
&+ \mu_i - \mu & &[\text{문항효과}] \\
&+ X_{pi} - \mu_p - \mu_i + \mu & &[\text{잔차}]
\end{aligned}
$$

표 1. 단일국면 교차설계에서 분산분석 표

변산원	분산표기	자유도	평균 제곱치 [a]	평균 제곱치(MS)의 기대치
p	σ_p^2	n_p-1	MS_p	$\sigma_{pi.e}^2 + n_i\sigma_p^2$
i	σ_i^2	n_i-1	MS_i	$\sigma_{pi.e}^2 + n_p\sigma_i^2$
pi,e	$\sigma_{pi.e}^2$	(n_p-1)(n_i-1)	$MS_{\pi.e}$	$\sigma_{pi.e}^2$

주. a. MS_P = 응답자의 평균 제곱치, MS_i = 문항의 평균 제곱치, $MS_{pi.e}$ = 오차분산의 평균 제곱. 다음 절에서의 급내 상관에서 사용하는 부호로 바꾸면, $MS_p = MS_R$, $MS_i = MS_C$, $MS_{pi.e} = MS_E$이다.

(10)

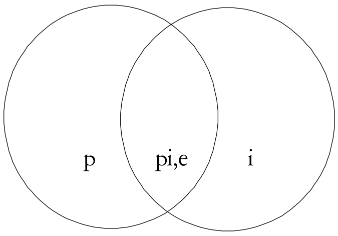

그림 1. 단일국면 교차설계

일반화가능도 이론은 분산분석 모형 중에서 무선효과(random effect) 모형을 기반으로 발전된 이론이므로 국면의 개념 역시 '무선' 효과 국면이 기본이다. 측정대상(응답자) 국면은 언제나 무선효과 국면이고, 오차국면 중 적어도 하나는 무선효과이어야 한다. 오차국면 모두가 고정효과이면, 오차분산이 발생하지 않기 때문이다.

급내 상관(Intraclass Correlation)과 신뢰도

통상적인 상관은 정확히 말하면 급간(interclass) 상관이다. 예로써, 임상가의 판단과 내담자의 반응간의 상관을 구할 때 두 변수는 각각 임상가 집단과 내담자 집단을 의미하는 변수로서 유목 또는 급(class) 변수라고 한다. 따라서 두 변수간 상관은 유목간 상관 또는 급간 상관(interclass correlation)인데 통상 상관계수로 부른다. 그러나 유목내 구성원들 즉 '임상가의 판단'에서 복수의 임상가들의 판단이 있고 그 판단들간 상관을 본다면 그것은 유목내 상관 또는 급내(intraclass) 상관이다. 즉, 급내 상관(혹은 유목내 상관)이란 급내에서 구한 관찰치들간의 관계를 나타내주는 측정치로서, 어떤 유목 변수내 구성원들간의 상관을 의미한다(McGraw & Wong, 1996).

급내 상관계수를 설명하기 위해, 급 변수 또는 유목 변수(A)가 하나인 식 (11)과 같은 분산분석 모형을 가정해보자. A가 평가자를 가리킨다면, $A_1, A_2, \ldots, A_n$은 각각의 상이한 평가자가 매긴 진점수이다.

$$X_{ij} = \mu + A_j + e_{ij}, \tag{11}$$

i=1, ...n번째 측정대상; j=1, ..., k번째 평가자

X_{ij} = i번째 측정대상에 대한 j번째 평가자의 관찰치

μ = 총 평균(상수)

A_j = j번째 평가자의 진점수

e_{ij} = 오차

각 평가자의 진점수 분산은 동일:

$$\sigma^2(A_1) = \cdots = \sigma^2(A_k)$$

식 (11)에서 볼 때, 식 (1)의 통합된 오차 e 보다는 특정의 오차국면이 하나 증가(A_j)되었고 식 (7)의 세분된 오차 $(e_1 + e_2 + \ldots + e_k)$에 비해서는 오차국면을 충분히 고려하지 않는다는 특징이 있다. 급내 상관계수는 고전검사이론을 기초로 하고 있다. 즉 고전검사이론의 기본 가정인 각 개인은 어떤 두 검사에서 동일한 진점수($A_j = A_{j'}$)를 가지며 오차점수 e 는 상호간에 상관이 없고 다른 검사의 진점수와도 상관이 없다는 것을 동일하게 가진다. 따라서 식 (4)의 모형에서 j번째 평가자가 매긴 점수와 j'번째 평가자가 매긴 점수간의 상관인 급내 상관을 구하기 위한 공분산은 아래 식 (12)로 표현이 가능하다.

$$Cov(X_{ij},\ X_{ij'}) \quad (12)$$
$$= Cov(\mu + A_j + e_{ij}, \mu + A_{j'} + e_{ij'})$$
$$= Cov(A_j + e_{ij}, A_{j'} + e_{ij'})$$
$$= \sigma^2(A_j A_{j'}) + \sigma(A_j e_{ij'}) + \sigma(A_{j'} e_{ij})$$
$$+ \sigma(e_{ij} e_{ij'}) = \sigma^2(A_j)$$

식 (12)을 사용하여 급내 상관계수를 계산하면 아래 식 (13)과 같다.

$$\text{급내상관계수} = \frac{Cov(X_{ij},\ X_{ij'})}{\sigma(X_{ij})\sigma(X_{ij'})} \quad (13)$$
$$= \frac{Cov(A_j,\ A_{j'})}{\sigma(X_{ij})\sigma(X_{ij'})} = \frac{\sigma^2(A_j)}{\sigma^2(X_{ij})}$$

즉, 급내 상관계수의 식은 전체분산 중에서 급내 구성원이 유발하는 진점수 분산의 비율로 정의가 되므로, 신뢰도계수가 정의되는 식 (3)과 그 의미가 동일하다.

식 (11)에 포함되는 유목변수의 수에 따라 변수의 효과에 대한 가정이 달라진다. 하지만 이러한 가정에 의해 정의되는 개념과 급내 상관의 분류 기준이 연구자 마다 달라 급내 상관의 사용과 해석에 혼동을 가져올 수 있다. 이러한 문제점을 해결하기 위하여 연구 1에서 지금까지의 급내 상관을 분석적 접근으로 정리하여 명확한 기준을 제시한다. 그리고 연구 2에서 연구 1을 통해 정리된 급내 상관과 일반화가능도 이론, 그리고 α계수를 분석적으로 접근하여 정리하고, 실제 자료를 사용해서 산출된 신뢰도 값을 확인 및 검토하여 경험적으로 비교한다.

연구 1: 급내 상관에 대한 서로 다른 정의에서의 일관성

연구문제

급내 상관에는 다양한 모형이 있다. 하지만, 동일한 모형에서 전제되는 가정이 동일함에도 불구하고 연구자에 따라 신뢰도에 대한 정의들이 서로 다르다. 여기서 오는 혼동을 극복하기 위해 그 이유를 밝히고, 급내 상관의 일관성 있는 논의를 제시하고자 한다.

연구방법

우선 동일한 모형과 동일한 가정에도 불구하고, 급내 상관에 대해 상이한 정의를 제시하는 여러 연구들 중에서 가장 대표적인 두 연구 즉, Shrout와 Fleiss(1979)와 McGraw와 Wong(1996)를 중심으로 비교 및 검토하는 분석적 접근을 하고자 한다. 이러한 분석적 접근을 통해 두 연구의 공통점과 차이점을 밝히고 급내 상관 개념의 일관성 있는 논리를 제시하겠다. 여기서 언급되는 자료 행렬에서 행(row)은 측정대상을, 열(column)은 평가자를 가리킨다.

Shrout와 Fleiss(1979) 및 McGraw와 Wong (1996)의 정의

급내 상관을 구분하는 기준에는 다음의 4가지 논점이 있다. 처음 두 논점은 Shrout와 Fleiss가 제시하는 것으로, 1) 분석의 단위로서 개별점수 또는 평균점수를 사용하였는지 여부,

표 2. 모형 별 급내 상관계수 정리 및 비교

Shrout와 Fleiss의 모형 번호[1]		McGraw와 Wong의 모형 번호[2]		상호작용[3]	측정치간 관계[4]	상관계수 $\hat{\rho}$의 추정 공식[5] 분석단위	
개별점수	평균점수	개별점수	평균점수			개별점수	평균점수
1원 무선모형(one-way model) : $x_{ij}=\mu+r_i+w_{ij}$ [6] $i=1,\ldots,n$명의 측정대상, $j=1,\ldots,k$명의 평가자							
ICC (1, 1)	ICC (1, k)	모형 1 ICC (1, 1)	모형 1 ICC (1, k)			$\frac{MS_R-MS_W}{MS_R+(k-1)MS_W}$	$\frac{MS_R-MS_W}{MS_R}$
2원 무선모형(two-way random model) : $x_{ij}=\mu+r_i+c_j+(rc_{ij})+e_{ij}$ [7] $i=1,\ldots,n$명의 측정대상(무선효과), $j=1,\ldots,k$명의 평가자(무선효과)							
		모형 2 ICC (C, 1)	모형 2 ICC (C, k)	O	일관성	$\frac{MS_R-MS_E}{MS_R+(k-1)MS_E}$	$\frac{MS_R-MS_E}{MS_R}$
ICC (2 ,1)	ICC (2, k)	모형 2 ICC (A, 1)	모형 2 ICC (A, k)	O	일치도	$\frac{MS_R-MS_E}{MS_R+(k-1)MS_E+\frac{k}{n}(MS_C-MS_E)}$	$\frac{MS_R-MS_E}{MS_R+\frac{1}{n}(MS_C-MS_E)}$
		모형 2A ICC (C, 1)	모형 2A ICC (C, k)	X	일관성	$\frac{MS_R-MS_E}{MS_R+(k-1)MS_E}$	$\frac{MS_R-MS_E}{MS_R}$
		모형 2A ICC (A, 1)	모형 2A ICC (A, k)	X	일치도	$\frac{MS_R-MS_E}{MS_R+(k-1)MS_E+\frac{k}{n}(MS_C-MS_E)}$	$\frac{MS_R-MS_E}{MS_R+\frac{1}{n}(MS_C-MS_E)}$
2원 혼합모형(two-way mixed model) : $x_{ij}=\mu+r_i+c_j+(rc_{ij})+e_{ij}$ $i=1,\ldots,n$명의 측정대상(무선효과), $j=1,\ldots,k$명의 평가자(고정효과)							
ICC (3, 1)		모형 3 ICC (C, 1)	모형 3 ICC (C, k)	O	일관성	$\frac{MS_R-MS_E}{MS_R+(k-1)MS_E}$	추정 불가능
		모형 3 ICC (A, 1)	모형 3 ICC (A, k)	O	일치도	$\frac{MS_R-MS_E}{MS_R+(k-1)MS_E+\frac{k}{n}(MS_C-MS_E)}$	추정 불가능
	ICC (3, k)	모형 3A ICC (C, 1)	모형 3A ICC (C, k)	X	일관성	$\frac{MS_R-MS_E}{MS_R+(k-1)MS_E}$	$\frac{MS_R-MS_E}{MS_R}$
		모형 3A ICC (A, 1)	모형 3A ICC (A, k)	X	일치도	$\frac{MS_R-MS_E}{MS_R+(k-1)MS_E+\frac{k}{n}(MS_C-MS_E)}$	$\frac{MS_R-MS_E}{MS_R+\frac{1}{n}(MS_C-MS_E)}$

주. 1. Shrout와 Fleiss(1979)가 제시한 모형 번호, 2. McGraw와 Wong(1996)이 제시한 모형 번호로, 'C'의 경우에는 측정치간 관계가 일관성을, 'A'의 경우에는 일치도를 나타냄, 3. (측정 대상 X 평가자)의 상호작용 존재 유무를 나타내며, 상호작용이 없는 경우 해당 모형에서 상호작용 효과(rc_{ij}) 생략, 4. 측정치간 관계가 일관성인지 일치도인지의 여부, 5. MS_R= 행 변수인 측정대상의 평균 제곱합, MS_W= 잔차 분산의 평균 제곱합, MS_C= 열 변수인 평가자의 평균 제곱합, MS_E= 오차분산의 평균 제곱합, 6. x_{ij}= i번째 측정 대상에 대한 j번째 평가자의 평가, μ= 모든 측정에 대한 총 평균, r_i= i번째 측정 대상 효과(평균 0, 분산 σ_r^2), w_{ij}= 잔차 효과(평균 0, 분산 σ_w^2), 7. c_j= j번째 평가자 효과(평균 0, 분산 σ_c^2), rc_{ij}= i번째 측정대상과 j번째 평가자의 상호작용 효과(평균 0, 분산 σ_{rc}^2), e_{ij}= 잔차 효과(평균 0, 분산 σ_e^2)

2) 자료 행렬내 열(column) 변수를 무선효과 또는 고정효과로 사용하였는지 여부로서, 이 두 기준에 따라 6가지 모형을 구분하였다(아래 표 2 참고). 하지만 Shrout와 Fleiss는 2원 모형 즉, 2원 무선모형과 2원 혼합모형에서 (측정 대상 X 평가자)의 상호작용을 일관성 있게 고려하지 않았다. 이러한 문제를 보완하기 위해 McGraw와 Wong이 나중 두 기준인 1) 2원 모형에서 (측정 대상 X 평가자)의 상호작용을 포함하는지의 여부, 2) 측정사이의 관계를 상대적인 일관성과 절대적인(absolute) 일치도로 구분하는 지 여부의 두 논점을 추가하였다. 따라서 총 4가지 논점에서 각 2가지 분류가 있어 16가지 모형으로 급내 상관을 구분하였다(표 2 참고).

표 2에서 McGraw와 Wong이 제시한 4가지 논점은 모형의 선택, (측정 대상 X 평가자)의 상호작용의 존재 유무, 측정치간 관계, 및 분석단위로서의 개별/평균 점수이다. 이에 근거하여 Shrout와 Fleiss가 제시한 개념을 살펴본 후, McGraw와 Wong이 제시한 개념과 비교 및 평가하도록 하겠다. 이 글에서 1원 무선모형은 논의의 대상이 아니므로, 2원 무선모형과 2원 혼합모형에서 혼동이 안될 경우 무선모형과 혼합모형으로 부르기로 한다. 또한 (측정 대상 X 평가자)의 상호작용이 존재하는 지의 유무를 구분하기 위하여 상호작용이 존재하지 않는 모형(absence of interation)은 '모형 2A', '모형 3A'로 표현하였다.

분석단위가 개별 점수인 경우와 평균 점수인 경우

Shrout와 Fleiss는 개인 점수는 대체적으로 신뢰할 수 없는 경우가 대부분이므로, 분석의 단위로 개인 점수가 사용되는 것은 거의 드문 상황이라고 제안하였다. 이것과 동일한 맥락으로 행동과학의 연구에서는 다문항 척도에서 개별 문항의 응답을 평균하여 척도점수로 사용하는 경우가 많으며, 이것은 개별 점수가 아닌 평균 점수를 사용하는 경우이다. 또한 신뢰도로서 α계수를 산출할 때의 암묵적 가정은 주어진 문항들을 평균하거나 합산해서 하나의 척도 값으로 사용한다는 것이다. McGraw와 Wong 또한 분석의 단위로 평균 점수를 사용한 경우와 개인 점수를 사용한 두 가지의 경우에서 산출된 계수의 해석은 다르지만, 개별 점수를 사용하여 계산된 계수는 평균 점수를 사용하여 계산된 계수의 특수한 경우라고 언급하였다. 이것은 표 2의 우측에 제시한 분석단위가 평균 점수인 경우의 추정식에 k명 평가자의 평균값이 아닌 k명 각각의 값을 대입해보면 분석단위가 개별 점수인 경우의 추정식과 동일해짐이 확인가능하다(부록 1 참고).

2원 무선 효과 모형과 2원 혼합 효과 모형 사이의 선택 여부

Shrout와 Fleiss는 자료 행렬내 열변수인 평가자 효과가 무선 효과인지 고정 효과인지 여부에 따라 2원 무선모형과 2원 혼합모형을 구분하였다. 즉, 무선모형에서는 특정 모집단 안에서 k명의 평가자들을 무선적으로 추출하였다는 암묵적 가정을 바탕으로한 일반화가 목적이지만, 혼합모형에서는 특정한 또는 임의로 선정된 단일 평가자 또는 k명의 평가자들이 평가(rating)를 한다는 가정이 들어 있다. 즉, 평가자는 일종의 고정효과 변수이다. 만약, 최종 연구에서 모든 자료들이 동시에 분석된다면, 평가자들의 효과와 (측정 대상 X 평가자)의 상호작용 효과는 평가의 오차에 기여하게

되고, 무선모형과 연관 되어 있는 신뢰도 계수로서 ICC(2, 1) 또는 ICC(2, k)가 산출 된다. 하지만, 평가자들의 평가를 각각 개별적으로 분석하고 분석된 결과를 합치게 되면 (측정 대상 X 평가자)의 상호작용 효과가 평가의 오차에 포함되지 않으므로 혼합모형과 연관되어 있는 신뢰도 계수인 ICC(3, 1) 또는 ICC(3, k)가 적절하다(Shrout & Fleiss, 1979). McGraw와 Wong은 위에 설명한 Shrout와 Fleiss의 개념을 받아들여 무선-고정 효과의 구별에서 중요한 점은 각 모형에서의 신뢰도 계수의 계산이 아니라 각 모형에서 해당되는 효과들의 해석이라고 덧붙였다.

측정치간 관계의 정도가 일관성(consistency)인지 완벽한 일치도(absolute agreement)인지의 여부

Shrout와 Fleiss는 평가자 국면이 무선인 경우의 신뢰도를 일치도로, 평가자 국면이 고정효과인 경우의 신뢰도를 일관성으로 간주하였다(Shrout & Fleiss, 1979). 만약 연구자의 의도에 따라 일정하게 선정된 평가자들만이 동일한 n명의 대상들을 평가한다면, 평가자들은 고정효과로 간주되며, 평가의 일관성은 혼합모형의 ICC(3, 1) 또는 ICC(3, k)에 의해 추정된다고 제시하였다. 일관성을 측정하는 것은 평가자들의 점수들 사이의 경향성을 강조하므로 신뢰도 ICC(3, 1) 또는 ICC(3, k)을 'ICC (consistency)'라고 표현하였다. 또한, 무한한 모집단에서 평가자들을 무선적으로 k명 추출하여 동일한 n명의 대상들을 평가한다면, 평가자들이 상호 교환적(interchangeable)이 되어, 무선효과처럼 간주되기 때문에 무선모형의 ICC(2, 1) 또는 ICC(2, k)가 적당하다고 제시하였다. 일치도를 추정하는 것은 평가자들의 상호교환성을 강조하므로 신뢰도 ICC(2, 1) 또는 ICC(2, k)는 'ICC(agreement)'로 설명한다. 즉, Shrout와 Fleiss는 일반화의 목적 외에 평가자들이 상호교환적인지의 질문에 의해 계수 ICC(2, 1) 또는 ICC(2, k)과 계수 ICC(3, 1) 또는 ICC(3, k)이 구분되는 것이라 주장하였다.

반면, McGraw와 Wong은 일관성과 일치도 사이의 구분이 그들의 공식에서 분모에 평가자간 분산을 오차 분산으로 포함시킬 것인지의 여부에 달려있다고 제시하였다. 일관성의 경우, 평가자들이 동일한 대상에게 부여한 평가 점수들의 값이 정확히 일치하지 않더라도 평가 점수들이 개인간 상대 서열을 유지하면 일관성 있으므로 세로열의 분산(column variance)을 오차분산에서 제외하였다. 하지만, '완벽한 일치도'의 경우는 평가자들이 부여한 점수들이 정확히 일치하지 않아 발생하는 분산 즉, 평가자들 사이에 의견 차이(disagreements)는 오차분산으로 간주된다. 예를 들어, 세 명의 평가자가 두 명의 응답자에게 (2, 4), (4, 6), (6, 8)의 점수들을 각각 부여하였다고 가정한다면, 일관성의 개념에 의해서는 평가자들 사이에 완벽한 일관성($\hat{\rho}$=1.00)이 있지만, 일치도의 개념에 따르면 평가자들은 중간 정도의 일치도 ($\hat{\rho}$=.67)가 있게 된다.

(측정 대상 X 평가자)의 상호작용의 존재 여부

표 2에서 Shrout와 Fleiss의 모형과 McGraw와 Wong의 모형을 비교해놓은 부분을 보면, Shrout와 Fleiss는 2원 혼합모형에서 분석단위가 평균점수인 경우는 예외지만, 무선모형과 혼합모형에서 (측정 대상 X 평가자)의 상호작용이 있는 경우의 모형만을 고려하였다. 무선모형의 경우에서는 상호작용들이 상호간에 독립

적이라고 가정하기 때문에 모형에 상호작용이 포함된다. 반면 혼합모형에서 분석단위가 평균점수일 때, 서로 다른 대상들을 평가하고 있는 경우만, 즉 서로 다른 평가자들이 서로 다른 대상들을 평가한 경우에만 독립이라고 가정한다. 예를 들어, i번째 대상을 (j-1)번째 평가자와 j번째 평가자가 동일하게 평가한 경우($(ab)_{i(j-1)}$,$(ab)_{ij}$)는 독립이 아니며 부적으로 상관된다. 따라서 대상에 대한 평가들의 공분산은 분석단위가 개별점수인 경우 $\sigma_r^2 - \sigma_{rc}^2/(k-1)$가 되지만, 분석단위가 k명의 평가자들에 기초한 두 개의 평균점수인 경우에는 평가들의 공분산이 σ_r^2가 되어 혼합모형의 경우에 추정이 불가능하게 된다. 이러한 이유로 인해 Shrout와 Fleiss는 혼합모형에서 분석단위가 평균점수인 경우는 (측정 대상 X 평가자)의 상호작용이 없다고 가정하고서 신뢰도 계수에 대한 추정 공식을 유도하였다.

Shrout와 Fleiss가 고려하지 않은 모형들을 McGraw와 Wong이 지적하면서, 그 모형들이 고려되지 않았던 이유는 전체 분산에서 공분산이 차지하는 비율을 봤을 때 상관이 거의 없는 것으로 간주되기 때문일 것으로 설명하였다. 그러나 이들 값은 관계의 정도를 추정할 때 고려할만한 경험적인 값으로 간주되므로 Shrout와 Fleiss가 고려하지 않은 모형들 즉, 2원 무선모형에서 분석단위가 개별점수와 평균점수인 모든 경우에서 (측정 대상 X 평가자)의 상호작용이 없는 경우의 일관성 모형도 모형에 포함되어야 한다고 제안하였다.

급내 상관의 정리 및 기준 제시

동일한 급내 상관의 개념에서 Shrout와 Fleiss (1979)가 정의내린 것과 McGraw와 Wong(1996)이 정의내린 것이 서로 다르다. 이것은 Shrout와 Fleiss가 상호작용이 없는 모형을 충분히 고려하지 않았고, 측정치간의 관계(일관성/일치도)와 분석단위(개별/평균점수)에 따라 일부의 모형만을 고려했기 때문이다. McGraw와 Wong이 제안한 모형들의 입장에서 보면, 1원 무선모형을 제외하였을 때, Shrout와 Fleiss는 2원 무선모형에서는 분석단위가 개별점수 또는 평균점수인 모든 경우에 (측정 대상 X 평가자)의 상호작용이 있는 일치도 모형을 제시하였다. 혼합모형에서는 분석단위가 개별점수인 경우에는 (측정 대상 X 평가자)의 상호작용이 있는 일관성 모형, 그리고 분석단위가 평균점수인 경우에는 (측정 대상 X 평가자)의 상호작용이 없는 일관성 모형만을 고려한 것이다. 결론적으로, Shrout와 Fleiss(1979)는 각 모형에서 4가지 논점중 일부만을 고려함에 급내 상관의 구분이 충분하지 않다.

또한, 표 2를 통해 무선모형과 혼합모형 간에는 평가자 요인이 각각 무선효과, 고정효과임에도 불구하고 급내 상관계수 추정식이 모형간에 동일하다(혼합모형에서 추정불가능 경우 제외). 동일하게 되는 과정을 부록2에서 수식으로 제시하였다. 즉, 급내 상관계수의 정의식인 식 (13)을 가지고 McGraw와 Wong은 각 모형에 맞춰 제곱평균기대치(EMS)의 추정식을 도출하였다. 급내 상관 ρ는 도출된 제곱평균기대치의 추정식을 적용하여 각 모형에 맞게 정의된 것이며, ρ의 정의식(표 2에서 우측)이 무선모형과 혼합모형 간에, 2원 무선모형 내에서, 또는 2원 혼합모형 내에서 우연히 동일한 것이다. 즉, 실제 급내 상관계수를 계산하는 계산식과 계산된 값은 동일하지만, 가정하고 있는 모형이 각각 다르므로 그 해석이 다르게 되어야 한다.

연구 2: 급내 상관과 일반화가능도 이론에 대한 고찰

연구문제

신뢰도를 산출하는 많은 방법들 중에서도 고전검사이론에 기초한 α계수는 응답자의 응답들간에 공변하는 정도에 기초하여 신뢰도가 정의된 상대평가의 신뢰도 지수이다. 고전검사이론의 확대 및 발전 형태인 급내 상관과 일반화가능도 이론에서도 상대평가의 일관성에 대한 지수가 산출된다. 그런데 급내 상관과 일반화가능도 이론 각각에서 상이한 가정에 근거하여 산출되는 신뢰도 값들이 동일하다. 즉, 일반화가능도 이론에서 평가자들을 “무선효과”의 가정으로 산출되는 일반화가능도 계수 값과, 급내 상관에서 평가자들을 “고정효과”의 가정으로 한 모형에서 분석단위가 평균 점수일 때 산출되는(Shrout & Fleiss, 1979) 일관성의 값이 α계수 값과 동일하다. 그 결과로 α계수에 대한 해석이 상이해진다. 따라서 본 연구는 이러한 문제점을 검토하고자 급내 상관과 일반화가능도 이론, α계수를 비교 및 검토하였다.

연구방법

급내 상관 및 일반화가능도 이론을 사용하여 상대평가의 신뢰도를 계산하는 경우, 두 이론에서 상대평가에 대한 상이한 해석을 제공함으로써 그 기초가 되는 고전검사이론의 이해에 많은 연구자들에게 큰 혼동을 주게 된다. 따라서 이러한 논점에 대하여 분석적으로 접근하여 혼동의 이유를 밝히고, 급내 상관과 일반화가능도 이론에서 각각 산출되는 상대평가의 신뢰도 계수를 사용할 때 명확한 기준을 제시하고자 한다. 또한 경험적 접근에서 실제 자료에 급내 상관, 일반화가능도 이론, 그리고 α계수를 적용하여 신뢰도를 추정한 후, 각 이론들로부터 산출된 신뢰도 값들이 개념적으로 정리한 내용과 동일한 결과를 나타내는 지 확인하였다.

분석적 접근

개념적 비교

1950년대와 1960년대에 급내 상관의 공식은 상대평가의 신뢰도 이론에 적용되었다. Ebel(1951), Haggard(1958)와 Buros(1963)는 급내 상관이 어떻게 통계학에서 신뢰도처럼 사용되는지 보여주었다. 그리고 Cronbach 등(1972)은 급내 상관을 일반화가능도 이론에 포함하였다(Cook, 2000에서 재인용).

앞서 설명한 것처럼, 일반화가능도 이론과 급내 상관은 고전검사이론을 기초로 측정모형에 분산분석모형 체계를 적용한 이론들이기 때문에 신뢰도의 추정방법에서 많은 공통점들을 가지고 있다. Shrout와 Fleiss(1979), McGraw와 Wong(1996)이 제시한 급내 상관의 모형들은 일반화가능도 이론에서 단일국면 G연구의 특수한 경우로 보여 진다(Cronbach et al, 1972). G연구의 분석처럼, 급내 상관도 다양한 분산성분의 모수(parameter)에 의하여 정의된다. 단일국면 연구의 일반화가능도 계수의 결과는 분산성분의 추정값이 부적인 경우를 제외하고는 Shrout와 Fleiss가 제안한 평균 제곱치(MS)에

일치한다. 만일 음수로 나온 분산성분의 값을 0으로 대체한다면 일반화가능도 계수는 Shrout와 Fleiss의 평균제곱치와 약간 달라진다. 특히, 일반화가능도 계수의 정의 식인 식 (8)에서 분모인 '전집 점수 분산 + 상대오차분산'은 '전체 분산'과 동일하므로, 그 의미가 급내 상관계수의 정의식인 식 (13)과 동일하게 되고, 급내 상관에서 가정한 모형이 무선 효과 모형이라면, 급내 상관계수는 일반화가능도 계수와 동일한 의미를 가지게 된다.

또한, 급내 상관은 모형이 1원(one-way)과 2원(two-way) 일 경우에서만 적용이 가능한데, 그 경우에만 급내 상관의 신뢰 구간과 검증 통계치들을 정의할 수 있기 때문이다. 반면 일반화가능도 이론에서는 단일국면(one facet)을 비롯한 2국면, 3국면은 물론 더 많은 수의 국면에까지 적용이 가능하여, 급내 상관보다 다원적 분석이 가능하다. 급내 상관의 평가자 국면이 무선효과(표 2에서 2원 무선효과모형)이면, 일반화가능도 이론에서의 단일국면(평가자 국면) 분석이 된다. 따라서 이 두 가지 이론에서 사용되는 방법에서의 차이점 비교는 급내 상관의 2원 모형(2원 무선효과모형)과 일반화가능도 이론의 단일국면 모형인 경우에 가능하다. 자료 구성을 볼 때, 행(row) 변수(예: 측정대상)와 열(column) 변수(예: 평가자 또는 문항)를 하나씩 가지는 2원 자료에서 행 변수인 측정 대상을 항상 무선효과로 가정하는 것은 급내 상관과 일반화가능도 이론에서 모두 동일하지만, 열 변수에 대한 가정이 급내 상관과 일반화가능도 이론 간에 다를 수 있다. 열변수를 급내 상관에서는 고정효과 또는 무선효과로 선택적으로 가정하는 반면에, 일반화가능도 이론은 열 변수가 하나밖에 없을 때는 반드시 무선효과로 가정 한다[1]. 열 변수가 여러 개이면 그 중 적어도 하나는 무선효과이다. 따라서 하나밖에 없는 열 변수를 고정효과라고 가정 한다면, 일반화가능도 이론에서는 오차 분산의 추정이 불가능하고 분석 불가능한 모형이다.

일반화가능도 이론은 급내 상관의 모형들 중에 표 2의 2원 혼합모형은 다룰 수 없고, 2원 무선모형에서도 상호작용이 존재하지 않는다고 가정하는 모형(모형 2A)을 고려하지 않는다. 일반화가능도 이론은 상호작용 항이 존재한다고 인정은 하지만, 평가자라는 단일 국면만 다루는 경우에는 상호작용 항과 오차 항이 분리되지 않는다. 그러나 급내 상관은 측정 대상 국면과 평가자 국면이라는 2국면으로 다루므로 상호작용 항의 추정이 가능하고, 상호작용이 있는 무선모형과 주효과만 있는 모형("모형 2A")의 구분이 가능하다. 따라서 일반화가능도 이론의 경우는 급내 상관에서 상호작용이 존재한다는 무선모형에 해당하지만 상호작용 효과가 오차 항에 모두 포함된다.

이와 같은 설명들을 종합해보면, 일반화가능도 이론에서 단일국면 모형은 급내 상관의 모형들 중에서 상호작용이 존재하는 무선모형에 해당되며, 열 변수를 고정효과로 가정하는 것이 불가능 하다. 무선모형의 관점에서 비교해보면, 일반화가능도 이론을 통해 산출되는 두 계수 즉, 상대평가의 신뢰도인 일반화가능도 계수가 급내 상관에서의 일관성 계수에, 절대평가의 신뢰도인 의존가능도 계수가 급내 상관의 일치도 계수에 해당된다.

---

1) 이것은 2원 자료에 대한 내용이다. 3원 자료부터는 열 변수 중 적어도 하나 이상이 무선효과이면 된다.

α 계수와 비교

Shrout와 Fleiss(1979)은 평가자들을 고정효과로 간주하는 2원 혼합모형에서, 분석단위로 평균점수를 사용한 신뢰도 값[ICC(3, k)]이 α 계수와 동일하다고 제시하였다. 즉, Shrout와 Fleiss의 2원 무선모형에서는 일치도밖에 없으므로 일관성을 산출하는 2원 혼합모형에서의 일관성 계수[ICC(3, k)]를 α계수와 동일하다고 한 것이라 해석된다.

하지만, McGraw와 Wong(1996)은 평가자들을 무선효과로 간주하고 측정 대상과 평가자 간의 상호작용이 존재한다고 가정하는 무선모형에서 분석의 단위로 평균점수를 사용한 일관성의 계수 값[ICC(C,k)]이 α계수와 동일하다고 설명한다. 이 같은 McGraw와 Wong의 제안은 일반화가능도 이론에서 단일국면에서의 일반화가능도 계수 값이 α계수와 동일하다고 제시한 것과 동일한데, 일반화가능도 이론의 단일국면에서평가자는 무선효과이기 때문이다.

앞서 제시한 것처럼, 일반화가능도 이론에서 상대평가를 위한 신뢰도로서 산출되는 일반화가능도 계수는 고전검사이론의 α계수와 개념적으로 유사하다(Cronbach et al., 1972). 일반화가능도 계수는 관찰 점수 분산에 대한 전집점수분산의 비율로 정의될 수 있는데, α계수는 일반화가능도 이론에서 오차국면이 하나인 단일국면에서의 G연구일 경우, 일반화가능도 계수와 일치하게 된다. 다시 말하면 α계수는 일반화가능도 이론에서 아주 제한된 경우의 일반화가능도 계수라고 정의될 수 있다. 또한 단일국면 교차설계에서 일반화가능도 계수는 고전검사이론에서 문항을 분석 단위로 하는 α 계수와 이론적으로 같은 기댓값을 가지며 실증 연구에서도 이들이 같음을 보고하고 있다(양지승, 이규민, 2007).

따라서 일반화가능도 이론에서 단일국면에서의 일반화가능도 계수와 α계수가 동일하고, 급내 상관에서 무선효과를 가정하는 2원 무선모형의 일관성 계수[ICC(C,k)]가 α계수와 동일하게 된다. 이 때의추정 공식은 표 2에서 $(MS_R$ - $MS_E)/MS_R$이다. 이러한 결론들로부터 α계수는 무선효과를 가정하는 신뢰도임이 확인 가능하다. 그런데 Shrout와 Fleiss가 α계수라고 주장하는 2원 혼합모형의 "모형 3A" 일관성 계수[ICC(C, k)]의 추정 공식 역시 $(MS_R$ - $MS_E)/MS_R$로 동일하다는 흥미 있는 현상이 나타난다. 즉, 실제 계산에서는 동일한 값이 되어 실질적인 문제는 없어진다. 물론 이론적으로는 McGraw와 Wong의 주장이 더 적절하다.

경험적 접근

연구자료

본 연구에서는 김성숙과 김양분(2001, p.78)의 자료를 사용하여 급내 상관계수, 일반화가능도 이론의 두 가지 계수인 일반화가능도 계수와 의존가능도 계수 및 α계수를 산출하여 그 값들을 비교해보기로 한다. 김성숙과 김양분(2001)의 자료는 3명의 평가자들이 6명의 응답자에 대한 평가점수들로 이루진 p(응답자) X r(평가자) 설계로 아래 표 3에 제시한다. 또한 평가자와 응답자 모두 평가자의 전집과 응답자의 전집에서 나온 임의 표본으로 간주한다.

표 3. p X r 설계의 김성숙과 김양분(2001, p. 78)의 자료

응답자	평가자		
	1	2	3
1	5	4	3
2	7	5	5
3	3	3	1
4	5	4	3
5	9	3	5
6	3	6	1
7	7	5	5
8	5	5	1
9	8	6	1
10	4	6	1

자료 분석 방법

2원 모형의 급내 상관계수와 고전검사이론의 대표적 신뢰도인 α값을 산출하기 위해 McGraw와 Wong(1996)의 이론이 적용되어 있는 SPSS 15.0[2)]프로그램을 사용하였다. G연구는 일반화가능도 이론을 적용한 p(응답자) X r(평가자) 무선효과 모형의 분산성분 추정에 필요하고 D연구는 G연구의 결과를 바탕으로 일반화전집을 추리하여야 하므로, 이러한 분석 결과를 산출해주는 GENOVA 프로그램[3)]이 사용되었다.

2) SPSS 프로그램은 상호작용이 있는 경우의 급내 상관계수만을 산출하여 제공한다.
3) GENOVA 프로그램은 완전균형설계의 단일변량 일반화가능도 분석을 위한 ANSI FORTRAN 컴퓨터 프로그램이다. 이 프로그램은 G연구와 D연구 둘 다 가능하다. GENOVA 프로그램은 1980년도 초에 Brennan과 Crick에 의해 만들어졌다.

분산분석 결과

측정 대상과 평가자 국면 모두 무선효과로 가정한 무선모형의 급내 상관 분석용 분산분석표는 표 4에 제시하고, 표 5는 표 3의 자료에 대한 분산분석의 결과(평가자 무선효과)를 나타내고 있다. 이는 SPSS 15.0 프로그램을 통해 급내 상관에 관한 분석을 하였을 때 계산되는 분산분석의 결과인 표 4와 동일하다. 표 5에 대해 자세히 설명을 하면, 전집 점수 분산과 각 오차 점수 분산 성분의 크기를 상대적으로 비교하여 각 요인이 관찰 점수 분산에 미치는 영향력을 설명하고 있다. 총 분산($\widehat{\sigma^2}(X)$, 3.27)에 대해 전집 점수 분산($\widehat{\sigma^2}(p)$)은 매우 작은 편(0.16, 5%)으로 응답자간의 수행 능력에 차이가 없음을 알려준다. 평가자의 분산성분($\widehat{\sigma^2}(r)$) 역시 작게 나타났다(0.59, 18%). 이는 평가자이 비교적 동질적으로 응답자 들에 대해 평가하고 있음을 나타내어준다. 잔차분산 성분으로서, 평가자그룹과 응답자그룹의 상호작용 및 오차분산($\widehat{\sigma^2}(pr)$)은 매우 크게(2.52, 77%) 나타났다. 일반화가능도 이론 연구에서 잔차 분산 성분이 전체 점수 분산에서 차지하는 비중이 큰 것은 일반적인 현상이다.

신뢰도 산출 결과

표 6은 일반화가능도 이론에서 산출된 일반화가능도 계수 값, 의존가능도 계수 값과 고전검사이론에서의 신뢰도인 α계수가 나타나 있다. α계수와 일반화가능도 계수 는 모두 상대평가의 신뢰도로서 일반화가능도 이론에서 평가자의 무선효과를 가정할 때, 두 값이 동

표 4. p(응답자) X r(평가자) 설계에 대한 분산분석 (급내 상관 분석)

변산원		자유도	제곱합	평균 제곱합	F	p
응답자 간		9	35.633	3.959		
응답자 내	평가자 간	2	47.400	23.700	9.424	.002
	오차	18	45.267	2.515		
	전체	20	92.667	4.633		
전체		29	128.300	4.424		

주. 전체 평균 = 4.30, 사용된 프로그램 : SPSS 15.0.

표 5. p(응답자) X r(평가자)설계에 대한 분산분석(ANOVA)

분산의 원천	자유도	제곱합	평균 제곱합	분산성분 추정치(%)
응답자	9	35.63	3.96	0.16(5)
평가자	2	47.40	23.70	0.59(18)
(응답자X평가자 상호작용 및 오차)	18	45.27	2.52	2.52(77)
전체	29	128.30	4.42	3.27(100)

주. 소수점 세 자리 수에서 반올림 하였음, 사용된 프로그램 : GENOVA.

표 6. p X r 설계에 대한 일반화가능도 계수와 의존가능도 계수, α계수

일반화가능도 이론				α**계수**
무선효과		**고정효과**		
$E\rho^2$	Φ	$E\rho^2$	Φ	
0.37	**0.24**	**-**	**-**	**0.37**

주. $E\rho^2$=일반화가능도 계수, Φ=의존가능도 계수, 소수점 세 자리 수에서 반올림 하였음, 일반화가능도 이론에 사용된 프로그램은 GENOVA이고, α계수 산출에 사용된 프로그램은 SPSS 15.0이다.

일하게 나왔음을 표 6에서 확인할 수 있다. 하지만 국면의 요인을 고정효과로 가정하였을 때에는 일반화가능도 계수와 의존가능도 계수가 산출되지 않았는데, 앞서 언급하였듯이 일반화가능도 이론에서는 국면 중 적어도 하나는 반드시 무선효과로 가정되어야 하기 때문에 하나밖에 없는 국면을 고정효과로 가정하면 GENOVA 프로그램에서는 각각의 계수 값들을 산출하지 못한다.

또한 표 7은 각각의 모형에서 각 모형 별로 급내 상관계수 값이 제시되어있다. 여기서 두 모형 즉, 무선모형과 혼합모형의 조건 별로 산출된 값이 동일하게 나타남을 알 수 있는데, 그것에 대해 McGraw와 Wong(1996)이 언급 한

표 7. p X r 설계에 대한 급내 상관계수 값

2원 무선모형				2원 혼합모형			
일치도		일관성		일치도		일관성	
개인점수	평균점수	개인점수	평균점수	개인점수	평균점수	개인점수	평균점수
0.094	**0.238**	**0.161**	**0.365**	**0.094**	**0.238**	**0.161**	**0.365**

주. 2원 무선모형=평가자 변수가 무선효과, 2원 혼합모형=평가자 변수가 고정효과.

것과 마찬가지로 국면의 요인이 무선효과로 가정되었다면 하나의 자료 범위를 넘어서 일반화 될 수 있지만, 고정효과로 가정되었을 경우에는 일반화 될 수 없다는 해석상의 차이만 있을 뿐이다.

표 6과 표 7의 결과를 비교해보면, 일반화가능도 계수(0.37)는 무선모형과 혼합모형의 일관성에서 분석단위가 평균점수일 경우의 신뢰도 값(0.365)과 동일하고, 의존가능도 계수(0.24)는 무선모형과 혼합모형의 일치도에서 분석단위가 평균점수일 경우의 신뢰도 값(0.238)과 동일함을 확인해 볼 수 있다. 이 같은 결과는 연구자의 모형이 2원 모형이고 무선효과 설계일 경우에는 일반화가능도 이론에서 산출해내는 계수의 값들과 급내 상관계수 값들이 동일하다는 결론을 이끌어 낼 수 있다. 하지만 만약 국면의 요인이 고정효과일 경우에는 일반화가능도 계수와 의존가능도 계수가 산출이 되지 않기 때문에 비교가 불가능하다.

결론 및 향후 연구방법

본 연구의 결과 및 추후 연구에 대한 논의를 정리하면 다음과 같다.

첫째로, 일반화가능도 이론이 국내에 소개가 되면서, 일반적으로 연구자가 자신의 연구결과가 신뢰로운지 알아보기 위해 일반화가능도 이론을 적용하는 경우가 많다. 하지만 앞의 결과에서 보는 것과 같이, 2원 모형에서의 일반화가능도 이론은 모든 변수가 무선효과로 가정되었을 경우에만 의미 있는 값을 제공한다. 즉, 2원 모형에서 고정효과의 가정을 지닌 변수가 모형에 포함되어 있는 경우에는 일반화가능도 계수와 의존가능도 계수가 산출되지 않으므로 이 경우에는 급내 상관방법을 이용하여 신뢰도 값을 산출하여야 한다.

둘째로, 일반화가능도 이론은 분산분석 모형 중 무선효과 모형을 기초로 하고 있으므로, 단일국면에서 평가자는 반드시 무선효과로 가정한다. 하지만 급내 상관은 일반화가능도 이론의 단일국면에 해당하는 2원 모형에서 평가자 효과를 무선효과와 고정효과로 가정하는 것이 모두 가능하다. 또한 일반화가능도 이론에서는 대상 X 평가자의 상호작용 효과의 존재를 인정하긴 하지만, 단일국면일 경우 오차와 구분하지 않으므로 실질적으로 계산이 불가능하다. 하지만 급내 상관은 무선효과와 고정효과를 가정하는 2원 모형 모두에서 상호작용이 존재를 인정하는 경우와 인정하지 않는 경우 모두를 고려하고 있다. 따라서 급내 상관의 모형들 중에서 상호작용이 존재하는 무선효과 모형이 일반화가능도 이론의 단일국면 모형과 동일하다. 그 중에서 급내 상관에서

일관성 계수가 일반화가능도 이론의 일반화가능도 계수에, 일치도 계수가 의존가능도 계수에 해당된다.

셋째로, Shrout와 Fleiss(1979)가 2원 혼합효과 모형에서 분석단위가 평균점수인 경우의 신뢰도 값이 α계수와 동일하다고 한 것은, McGraw와 Wong(1996)이 Shrout와 Fleiss가 언급하지 않은 모형까지 포함하여 급내 상관 모형을 확장함으로써 일반화가능도 이론에서의 α계수에 대한 정의와 급내 상관방법에서의 정의가 일치 되었다. 즉, α계수는 2원 모형에서 모든 변수가 무선효과로 가정되고 분석단위가 평균점수를 사용하였을 때 산출되는 신뢰도 값이다.

마지막으로, McGraw와 Wong은 2원 무선설계와 혼합설계에서 일치도와 일관성이 동일한 수식과 값을 가지며 해석상의 차이만 있을 뿐이라고 언급하였다. 하지만 이것은 2원 모형에서 대상을 평가자가 반복 평가가 아닌 단 한 번의 평가를 한 경우에만 해당한다. 단 한 번의 평가를 하면 대상 X 평가자의 효과가 오차와 구별이 되지 않는 혼입의 결과가 발생한다. 따라서 반복 측정을 통해 상호작용 효과와 오차가 구별이 되는 상황에서는 일치도와 일관성의 값이 동일하지 않다. 무선변수가 모형에 포함되어 있다면 고정변수만 포함되어 있을 경우보다 변수의 유의한 정도가 상당히 감소된다. Clark(1973)는 언어재료를 사용한 과거 연구들에 대해 언어재료를 고정효과가 아닌 무선효과로 놓고 재분석한 결과, 유의한 것으로 보고된 많은 연구들이 실제로는 유의하지 않거나 유의하더라도 상당히 과대 평가되었음을 지적한 바 있다(이종구, 2001에서 재인용). 따라서 반복 측정을 한 경우의 연구라면 단일 측정의 경우보다 더 세심한 주의가 필요하다.

위의 결과와 관련하여 본 연구에서는 다루지 못한 반복 측정을 한 경우의 연구에서 2원 무선효과 모형과 혼합효과 모형의 일치도와 일관성에 대한 급내 상관계수 값이 실제 자료에서 어떠한 차이가 있고 그 차이의 정도가 얼마나 되는지, 또한 수식에서의 변화가 있는지 등을 알아보는 후속 연구가 이루어져 급내 상관 방법에 대한 체계적이고 명확한 정리를 하여야 할 것이다.

참고문헌

강승호, 김양분 (2004). **신뢰도**. 서울: 교육과학사.

김성숙, 김양분 (2001). **일반화가능도 이론**. 서울: 교육과학사.

김성숙 (1993). 관찰을 통한 교수 평가 체계에 대한 측정의 일반화 가능도 연구. **교육학연구, 31**(1), 23-40.

양지승, 이규민 (2007). 일반화가능도 이론을 적용한 단위검사 구성 검사점수의 신뢰도 추정. **교육평가연구, 20**(1), 119-139.

이규민 (2003). 단위검사 개념의 적용–일반화가능도 이론을 중심으로–. **교육평가연구, 16**(1), 53-70.

이규민 (2006). 학생들의 동료평가를 활용한 수행평가 결과의 일반화가능도 분석. **교육평가연구, 19**(3), 107-121.

이종성 (1989). **일반화가능도 이론**. 연세대학교 출판부.

이종구 (2001). **SAS와 통계자료 분석**. 학지사.

Allen, M. J & Yen, W. M. (1979). *Introduction to Measurement Theory* (1st ed.). Long Grove, Illinois: Waveland Press. Inc.

Anastasi, Anne (1982). *Psychological Testing* (6th ed.). NY: Macmillan.

Brennan, R. L. (2001). *Generalizability theory*. NY: Springer.

Brennan, R. L. (1983). *Elements of generalizability theory*. Iowa City, IA: American College Testing Program.

Buros, O. K. (1963). *Schematization of old and new concepts of test reliability based upon parametric models*. New Brunswick, NJ: Gryphon Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335-359.

Cook, C. (2000). A review of intraclass correlation. The Annual Meeting of the Southwest *Educational Research Association*. (Dallas, Tx, January 27-29, 2000). 1-35.

Crick, J. E. & Brennan, R. L. (1983). *Mannual for GENOVA* : A generalized analysis of variation system. Mid-Western Educational Researcher, v9. n2 p.2-4 Spr.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika 16*(3), 297-334..

Croncach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. NY : Wiley.

Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika*, *16*(4), 407-424.

Fagot, R. F. (1991). Reliability of rating for multiple judges: Intraclass correlation and metric scales, *Applied Psychological Measurement*. *15*(1), 1-11.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282.

Haggard, E. A. (1958). *Intraclass correlation and the analysis of variance*. NY: The Dryden Press.

McDonald, R. P. (1999). *Test Theory: A unified treatment*. Mahwah, NJ: Erlbaum.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30-46.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19(1)*, 231-240.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). NY: McGraw-Hill.

1차원고접수 : 2009. 10. 10.
수정원고접수 : 2009. 12. 09.
최종게재결정 : 2009. 12. 21.

한국심리학회지 : 일반
Korean Journal of Psychology: General
2009, Vol. 28, No. 3, 659-686

A Theoretical and Empirical Comparison between Intraclass Correlation and Generalizability

Ji Yoon Kim
Sungkyunkwan University

Sung Chil Yeo
Konkuk University

Soon Mook Lee
Sungkyunkwan University

Researches report many different measures of reliability to establish accuracy of measurement. Intraclass correlation coefficient and generalizability theory are based on classical test theory and widely used for obtaining reliability measurement. They share some communality, however, they have different assumptions for estimation of reliability measures. If was observed that some measures rooted in different assumptions yield identical values. Also, it was found that definitions of reliability are different across different researchers in the context of intraclass correlation coefficients. In the present study we examined two studies defining the intraclass correlation differently. Also relationships between the reliability measures provided by intraclass correlation coefficient and generalizzbility theory are interpreted in the context of reliability alpha coefficient that is most widely used for a reliability measure.

Key words : reliability index, Cronbach의 α, Intraclass Correlation Coefficient, Generalizability

부록 1. 1원 무선모형에서 급내 상관계수의 추정공식 증명 - 분석단위에 따라 (표 2)

1-1. 1원 무선모형 (분석단위 : 개별 점수)

$$x_{ij} = \mu + r_i + w_{ij},\ i = 1, \cdots, n \text{ and } j = i, \cdots, k$$

$$MS_R = \frac{1}{n-1} SS_R, \qquad E(MS_R) = k\sigma_r^2 + \sigma_w^2$$

①

$$M_{S_w} = \frac{1}{n(k-1)} SS_w, \quad E(MS_w) = \sigma_w^2 \qquad ②$$

$$\hat{\rho} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \hat{\sigma}_w^2} = \frac{MS_R - MS_W}{MS_R + (k-1)MS_W} \qquad ③$$

(증명) ② 로부터 $\hat{\sigma}_w^2 = MS_W$ ④

④를 ①에 대입하면

$$\hat{\sigma}_r^2 = \frac{1}{k}(MS_R - MS_W) \qquad ⑤$$

따라서 ④와 ⑤를 식 ③에 대입하면 주어진 결과를 얻는다.

1-2. 1원 무선모형 (분석단위 : 평균 점수)

$$x_{ij} = \mu + r_i + w_{ij},\ i = 1, \cdots, n \text{ and } j = i, \cdots, k$$

$$MS_R = \frac{1}{n-1} SS_R, \qquad E(MS_R) = k\sigma_r^2 + \sigma_w^2$$

①

$$M_{S_w} = \frac{1}{n(k-1)} SS_w, \quad E(MS_w) = \sigma_w^2 \qquad ②$$

$$\hat{\rho} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \frac{\hat{\sigma}_w^2}{k}} = \frac{MS_R - MS_W}{MS_R} \qquad ③$$

(증명) ② 로부터 $\hat{\sigma}_w^2 = MS_W$ ④

④를 ①에 대입하면

$$\hat{\sigma}_r^2 = \frac{1}{k}(MS_R - MS_W) \qquad ⑤$$

따라서 ④와 ⑤를 식 ③에 대입하면 주어진 결과를 얻는다.

부록 2. 2원 무선모형과 2원 혼합모형에서 급내 상관계수의 추정공식 증명 (분석단위가 개별 점수인 경우)

2-1. 상호작용 있는 2원 무선모형 (일관성)

$$x_{ij} = \mu + r_i + c_j + rc_{ij} + e_{ij}, \quad i = 1, \cdots, n \text{ and } j = 1, \cdots, k$$

$$E(MS_R) = k\sigma_r^2 + \sigma_{rc}^2 + \sigma_e^2 \quad ①$$

$$E(MS_W) = \sigma_c^2 + \sigma_{rc}^2 + \sigma_e^2 \quad ②$$

$$E(MS_C) = n\sigma_c^2 + \sigma_{rc}^2 + \sigma_e^2 \quad ③$$

$$E(MS_E) = \sigma_{rc}^2 + \sigma_e^2 \quad ④$$

$$\hat{\rho} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + (\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2)} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E} \quad ⑤$$

(증명)

④로부터 $\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2 = MS_E$ ⑥

④를 ①에 대입하면

$$\hat{\sigma}_r^2 = \frac{1}{k}(MS_R - MS_E) \quad ⑦$$

⑥과 ⑦을 ⑤에서 분산 추정치를 나타내는 항에 대입하면 주어진 결과를 얻는다.

2-2. 상호작용 있는 2원 무선모형 (일치도)

$$x_{ij} = \mu + r_i + c_j + rc_{ij} + e_{ij}, \quad i = 1, \cdots, n \text{ and } j = 1, \cdots, k$$

$$E(MS_R) = k\sigma_r^2 + \sigma_{rc}^2 + \sigma_e^2 \quad ①$$

$$E(MS_W) = \sigma_c^2 + \sigma_{rc}^2 + \sigma_e^2 \quad ②$$

$$E(MS_C) = n\sigma_c^2 + \sigma_{rc}^2 + \sigma_e^2 \quad ③$$

$$E(MS_E) = \sigma_{rc}^2 + \sigma_e^2 \quad ④$$

$$\hat{\rho} = \frac{\widehat{\sigma_r^2}}{\widehat{\sigma_r^2} + \widehat{\sigma_c^2} + (\widehat{\sigma_{rc}^2} + \widehat{\sigma_e^2})} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)} \quad ⑤$$

(증명)

④로부터 $\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2 = MS_E$ ⑥

④를 ①에 대입하면

$$\hat{\sigma}_r^2 = \frac{1}{k}(MS_R - MS_E) \quad ⑦$$

④를 ③에 대입하면

$$\hat{\sigma}_c^2 = \frac{1}{n}(MS_C - MS_E) \quad ⑧$$

⑥과 ⑦, ⑧을 ⑤에 대입하면 주어진 결과를 얻는다.

2-3. 상호작용 없는 2원 무선모형 (일관성)

$$x_{ij} = \mu + r_i + c_j + e_{ij}, \quad i = 1, \cdots, n \text{ and } j = 1, \cdots, k$$

$$E(MS_R) = k\sigma_r^2 + \sigma_e^2 \quad ①$$

$$E(MS_W) = \sigma_c^2 + \sigma_e^2 \quad ②$$

$$E(MS_C) = n\sigma_c^2 + \sigma_e^2 \quad ③$$

$$E(MS_E) = \sigma_e^2 \quad ④$$

$$\hat{\rho} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \hat{\sigma}_e^2} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E} \quad ⑤$$

(증명)

④로부터 $\hat{\sigma}_e^2 = MS_E$ ⑥

④를 ①에 대입하면

$$\hat{\sigma}_r^2 = \frac{1}{k}(MS_R - MS_E) \quad ⑦$$

⑥과 ⑦을 ⑤에 대입하면 주어진 결과를 얻는다.

2-4. 상호작용 없는 2원 무선모형 (일치도)

$$x_{ij} = \mu + r_i + c_j + e_{ij}, \quad i = 1, \cdots, n \text{ and } j = 1, \cdots, k$$

$$E(MS_R) = k\sigma_r^2 + \sigma_e^2 \quad ①$$

$$E(MS_W) = \sigma_c^2 + \sigma_e^2 \qquad ②$$
$$E(MS_C) = n\sigma_c^2 + \sigma_e^2 \qquad ③$$
$$E(MS_E) = \sigma_e^2 \qquad ④$$
$$\hat{\rho} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \hat{\sigma}_c^2 + \hat{\sigma}_e^2} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)} \qquad ⑤$$

(증명)

④로부터 $\hat{\sigma}_e^2 = MS_E$ ⑥

④를 ①에 대입하면

$$\hat{\sigma}_r^2 = \frac{1}{k}(MS_R - MS_E) \qquad ⑦$$

④를 ③에 대입하면

$$\hat{\sigma}_c^2 = \frac{1}{n}(MS_C - MS_E) \qquad ⑧$$

⑥과 ⑦, ⑧을 ⑤에 대입하면 주어진 결과를 얻는다.

3-1. 상호작용 있는 2원 혼합모형 (일관성)

$$x_{ij} = \mu + r_i + c_j + rc_{ij} + e_{ij}, \quad i = 1, \cdots, n \text{ and } j = 1, \cdots, k$$
$$E(MS_R) = k\sigma_r^2 + \sigma_e^2 \qquad ①$$
$$E(MS_W) = \theta_c^2 + \frac{k}{k-1}\sigma_{rc}^2 + \sigma_e^2 \qquad ②$$
$$E(MS_C) = n\theta_c^2 + \frac{k}{k-1}\sigma_{rc}^2 + \sigma_e^2 \qquad ③$$
$$E(MS_E) = \frac{k}{k-1}\sigma_{rc}^2 + \sigma_e^2 \qquad ④$$
$$\hat{\rho} = \frac{\hat{\sigma}_r^2 - \hat{\sigma}_{rc}^2/(k-1)}{\hat{\sigma}_r^2 + (\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2)} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E} \qquad ⑤$$

(증명)

식①에서 식④를 빼면 $MS_R - MS_E = k\hat{\sigma}_r^2 - \frac{k}{k-1}\hat{\sigma}_{rc}^2$

따라서 $\hat{\sigma}_r^2 - \frac{1}{k-1}\hat{\sigma}_{rc}^2 = \frac{1}{k}(MS_R - MS_E)$ ⑥

①과 ④로부터

$$MS_R + (k-1)MS_E = k\hat{\sigma}_r^2 + \hat{\sigma}_e^2 + (k-1)\left(\frac{k}{k-1}\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2\right)$$
$$= k(\hat{\sigma}_r^2 + (\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2))$$

따라서 $\hat{\sigma}_r^2 + (\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2) = \frac{MS_R + (k-1)MS_e}{k}$ ⑦

⑥와 ⑦의 결과를 ⑤에서 제시된 ρ의 정의식에 대입하면 ⑤의 오른편식을 얻는다.

3-2. 상호작용 있는 2원 혼합모형 (일치도)

$$x_{ij} = \mu + r_i + c_j + rc_{ij} + e_{ij}, \quad i = 1, \cdots, n \text{ and } j = 1, \cdots, k$$

$$E(MS_R) = k\sigma_r^2 + \sigma_e^2 \quad ①$$

$$E(MS_W) = \theta_c^2 + \frac{k}{k-1}\sigma_{rc}^2 + \sigma_e^2 \quad ②$$

$$E(MS_C) = n\theta_c^2 + \frac{k}{k-1}\sigma_{rc}^2 + \sigma_e^2 \quad ③$$

$$E(MS_E) = \frac{k}{k-1}\sigma_{rc}^2 + \sigma_e^2 \quad ④$$

$$\hat{\rho} = \frac{\hat{\sigma}_r^2 - \hat{\sigma}_{rc}^2/(k-1)}{\hat{\sigma}_r^2 + \hat{\theta}_c^2 + (\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2)} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)} \quad ⑤$$

(증명)

식①에서 식④를 빼면 $MS_R - MS_E = k\hat{\sigma}_r^2 - \frac{k}{k-1}\hat{\sigma}_{rc}^2$

따라서 $\hat{\sigma}_r^2 - \frac{1}{k-1}\hat{\sigma}_{rc}^2 = \frac{1}{k}(MS_R - MS_E)$ ⑥

$$MS_R + (k-1)MS_E = k\hat{\sigma}_r^2 + \hat{\sigma}_e^2 + (k-1)\left(\frac{k}{k-1}\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2\right)$$
$$= k(\hat{\sigma}_r^2 + (\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2))$$

따라서 $\hat{\sigma}_r^2 + (\hat{\sigma}_{rc}^2 + \hat{\sigma}_e^2) = \frac{MS_R + (k-1)MS_e}{k}$ ⑦

또한 ④를 ③에 대입하면

$$\hat{\theta}_c^2 = \frac{1}{n}(MS_C - MS_E) \qquad ⑧$$

⑥과 ⑦, ⑧의 결과를 ⑤에서 제시된 ρ의 정의식에 대입하면 ⑤의 오른편식을 얻는다.

3-3. 상호작용 없는 2원 혼합모형 (일관성)

$$x_{ij} = \mu + r_i + c_j + e_{ij}, \quad i = 1, \cdots, n \text{ and } j = 1, \cdots, k$$

$$E(MS_R) = k\sigma_r^2 + \sigma_e^2 \qquad ①$$

$$E(MS_W) = \theta_c^2 + \sigma_e^2 \qquad ②$$

$$E(MS_C) = n\theta_c^2 + \sigma_e^2 \qquad ③$$

$$E(MS_E) = \sigma_e^2 \qquad ④$$

$$\hat{\rho} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \hat{\sigma}_e^2} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E} \qquad ⑤$$

(증명)

④로부터 $\hat{\sigma}_e^2 = MS_E$ ⑥

④를 ①에 대입하면

$$\hat{\sigma}_r^2 = \frac{1}{k}(MS_R - MS_E) \qquad ⑦$$

⑥과 ⑦을 ⑤에 대입하면 주어진 결과를 얻는다.

3-4. 상호작용 없는 2원 혼합모형 (일치도)

$$x_{ij} = \mu + r_i + c_j + e_{ij}, \quad i = 1, \cdots, n \text{ and } j = 1, \cdots, k$$

$$E(MS_R) = k\sigma_r^2 + \sigma_e^2 \qquad ①$$

$$E(MS_W) = \theta_c^2 + \sigma_e^2 \qquad ②$$

$$E(MS_C) = n\theta_c^2 + \sigma_e^2 \qquad ③$$

$$E(MS_E) = \sigma_e^2 \qquad ④$$

$$\hat{\rho} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \hat{\theta}_c^2 + \hat{\sigma}_e^2} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)} \quad ⑤$$

(증명)

④로부터 $\hat{\sigma}_e^2 = MS_E$ ⑥

④를 ①에 대입하면

$$\hat{\sigma}_r^2 = \frac{1}{k}(MS_R - MS_E) \quad ⑦$$

④를 ③에 대입하면

$$\hat{\theta}_c^2 = \frac{1}{n}(MS_C - MS_E) \quad ⑧$$

⑥과 ⑦, ⑧을 ⑤에 대입하면 주어진 결과를 얻는다.