The Korean Journal of Applied Statistics
2024, Vol. 37, No. 1, 61–72

DOI: https://doi.org/10.5351/KJAS.2024.37.1.061
Print ISSN1225-066X / Online ISSN 2383-5818

Reliability in longitudinal study

Jinuk Kim[1,a]

[a]Department of Sport and Exercise Sciences, Kunsan National University

Abstract

The purpose of this study is to investigate retest reliabilities in longitudinal study, the same test is administered repeatedly over time. Linear mixed models were used to establish various situations of tests occurred in longitudinal study. Combination of two types of true value and three types of systematic error was considered. In order to apply the models to real longitudinal data, height data from the Berkeley growth study and vocabulary score data from the University of Chicago experimental school were used. Using the mixed model, there is an advantage that the reliability can be determined by selecting the covariance structure of the true value and the error separately. However, in order to properly analyze the reliability, researchers need to consider variations that can occur in measurement, such as characteristics of subject, the test, and the the treatment applied in the study. And the proper model should be selected and the quality of the measurement should be evaluated for each trial.

Keywords: longitudinal study, reliability, linear mixed model

1. 서론

오차(error)는 모든 측정값(observed value)에 항상 포함되어 있다. 측정값은 의사결정을 위한 추정(estimation)과 검정(test) 과정에 사용되는데 여기에 포함된 오차는 편의(bias)로 작용하기 때문에 만일 오차가 크다면 잘못된 판단을 할 가능성이 높아진다. 따라서 측정값에서 오차를 제외한 부분인 신뢰도(reliability)와 타당도(validity)를 측정의 평가 도구로 사용해 왔다.

우리가 정말로 측정하고자 하는 것을 참값(true value)이라 할 때 고전검사이론(classical test theory) (cLord와 Novick, 1968/2008)에 의하면 오차는 측정값과 참값의 차이로 정의된다. 여기서 신뢰도와 직접적으로 관련 있는 오차는 무수히 많은 시행(trial)에 의해 정의될 수 있다 (Guttman, 1945). 즉, 동일 측정 도구로 동일 측정 대상을 동일 절차로 시간에 따라 반복 시행하였을 때 방향성 없이 우연히 나타나는 오차를 특히 임의오차(random error)라 하는데 측정값에서 임의오차가 차지하는 부분이 작다면 신뢰도는 높은 것으로 평가된다.

반복 시행의 양적자료(quantitative data) 재검사 신뢰도(retest reliability)는 상관계수로 평가하는 것으로 많은 연구자들이 알고 있다. 측정값은 서로 독립적인 참값과 오차의 합으로 이루어졌는데 반복 시행에서 측정 대상의 참값은 변하지 않거나 일정하게 변하고 오차 간은 서로 독립이라 가정에 근거한다. 그러나 실제로 참값의 변화가 일정하지 않다면 이 가정하에서 참값의 변화는 오차로 간주 될 수밖에 없으며 참값과 오차의 독립이 성립되지 않는다. 그리고 오차 간 독립도 더 이상 성립되지 않는다. 중요한 것은 반복 시행에서 오차의 독립에 영향을 주는 요인에 더 있다는 점이다. 모든 측정 대상을 동일 순서로 시간에 따라 측정하기 때문에 발생하는 요인이다. 참값의 일정하지 않은 변화를 나타내는 불안정성(instability)과 오차의 관련성(correlated error)에 의해 신뢰도 계수로서의 상관계수는 편의된 결과를 가져올 수 있다 (Blok과 Saris, 1983; Werts 등,

[1]Department of Sport and Exercise Sciences, Kunsan National University, 558 Daehak-Ro, Gunsan 54150, Korea. E-mail: jinuk@kunsan.ac.kr.

Published 29 February 2024 / journal homepage: http://kjas.or.kr

1980). 참값의 변화를 고려하지 않으면 과소 추정되고 (Werts 등, 1980), 오차의 관련성을 고려하지 않으면 과대 추정된다 (Bost, 1995; Maxwell, 1968; Smith와 Luecht, 1992).

반복 시행에서 참값의 불안정성과 오차의 관련성이라는 두 문제를 고려한 적절한 신뢰도 계수를 구하기 위한 연구는 Heise (1969), Wiley와 Wiley (1970)에 의해 시작되었다. 심플렉스 패턴(simplex pattern) (Jöreskog, 1970) 모형을 사용하였으며 Heise (1969)는 모든 시행에서 신뢰도는 동일하다는 가정, Wiley와 Wiley (1970)는 오차 분산이 동일하다는 가정을 하였다. 이후에 여러 모형이 제안되었으며 (Blok과 Saris, 1983; Deshon 등, 1998; Jagodzinski와 Kühnel, 1987; Marcoulides, 2019; Marsh와 Grayson, 1994; Palmquist와 Green, 1992; Raffalovich와 Bohrnstedt, 1987; Smith, 1974; Smith와 Luecht, 1992; Werts 등, 1980) 주로 인자분석(factor analysis)과 구조방정식 모형(structural equation model)을 기반으로 하며 종단적 연구(longitudinal study)에 의한 자료의 신뢰도를 다루었다.

종단적 연구의 측정 절차는 재검사 신뢰도의 절차와 유사하다. 종단적 연구는 시행 간격이 상대적으로 크고, 실험의 처리에 따른 변화를 보는 것에 주로 초점이 맞춰져 있으며, 참값의 변화가 일반적으로 일어난다. 종단적 연구의 신뢰도도 각 시행에서 달라지는 것으로 보는 것이 타당하며 자료의 특성에 맞는 적절한 모형으로 신뢰도를 추정할 필요가 있다. 선형혼합모형(linear mixed model) (Fitzmaurice 등, 2011)은 종단적 자료에서 신뢰도 평가를 위한 참값과 오차 구조의 모형화에 최적화된 접근법이다. 그리고 구체적인 모형의 수립을 위해서 우리는 반드시 또 다른 형태의 오차를 고려해야 한다. 이는 계통오차(systematic error)로 반복 시행에서 주로 같은 방향으로 지속적 또는 체계적으로 나타나는 형태의 오차이다. 계통오차를 세분화하여 모형에 고려하면 신뢰도에 영향을 주는 여러 요인이 명확하게 설명되며 참값과 오차 간의 관계도 더 유연하게 설명이 가능하다.

재검사 신뢰도를 위한 측정 모형은 고려해야 할 것이 많고 그 만큼 복잡할 수 밖에 없었다. 따라서 전통적으로 여러 가정에 의한 아주 간략화된 모형이 사용되어 왔다. 그러나 분석 모형의 발전으로 인해 가정은 완화하면서 상황에 맞게 적용시켜 볼 필요가 있다. 본 논문의 구성은 다음과 같다. 연구 방법의 첫 부분은 하나의 측정 도구로 반복 시행하여 측정하는 종단적 연구에서 다양한 상황의 측정을 선형혼합모형으로 설명하고 신뢰도 계수를 정의하는 것이다. 이를 위해 세부적인 참값과 계통오차를 모형에 고려하였다. 두 번째 부분은 실제 자료에 모형을 적용하여 다양한 신뢰도 계수를 추정하여 비교 분석하였다.

2. 연구방법

2.1. 모형

동일 검사 반복 시행이라도 검사의 특성과 환경에 따라 여러 형태의 측정 결과가 존재한다. 하나의 측정 모형으로 모든 재검사 신뢰도를 설명하는 것은 한계가 있다. 세분화된 모형을 위해 계통오차에 대한 정의를 먼저 살펴봐야 한다. Taylor (1997)는 계통오차가 임의오차와 다르게 반복 측정에서 드러나지 않으며 주로 측정 도구의 캘리브레이션 오차(miscalibration)에 의한 것이라 하였다. 그러나 그는 계통오차와 임의오차 구분이 항상 명확한 것은 아니라고 하였다. 이를 반영하듯 Sideridis 등 (2019)은 계통오차는 도구의 특성, 측정 과정, 측정 대상의 특성, 그리고 이 세 가지의 조합에 의해 나타난다고 하였다. 종합해 보면 계통오차는 도구 특성 외에 측정 대상, 시행 모든 요인에 의해 나타날 가능성이 있는 것으로 규정하고 있다.

본 연구에서 측정 대상은 인간에게 국한된 것은 아니지만 피험자(subject)라 부르기로 하며, 시행은 시점으로 연속형이 아닌 이산형으로 간주한다. 그리고 모든 피험자는 동일 순서로 각 시행에서 한 번씩 측정하는 모형을 고려한다. 따라서 측정값은 피험자와 시행의 교차설계(crossed design)에 의한 결과이며, 측정값 모형은 반드시 시행의 평균 효과가 포함되어야 한다. 즉, 피험자는 모집단에서 추출된 표본이지만 시행은 그러한 특성을 가지지 않는다. 본 연구의 모형은 반복 시행에서 참값의 안정성(stability)과 불안정성으로 크게 구분하였다. 계통오차는 특성에 따라 세부적으로 세 가지로 나누었다.

Table 1: Covariance matrix models and reliability coefficients on longitudinal data

True	Systematic error	Covariance matrix: True	Covariance matrix: Trial	Covariance matrix: Error	Reliability coefficients
Stability				$\mathbf{\Theta} = \begin{pmatrix} \theta^2 & 0 & \cdots & 0 \\ 0 & \theta^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta^2 \end{pmatrix}$	$\frac{\tau^2}{\tau^2+\theta^2}$
	d (2.1)	$\boldsymbol{T} = \begin{pmatrix} \tau^2 & \tau^2 & \cdots & \tau^2 \\ \tau^2 & \tau^2 & \cdots & \tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau^2 & \tau^2 & \cdots & \tau^2 \end{pmatrix}$	$\begin{pmatrix} o^2 & 0 & \cdots & 0 \\ 0 & o^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & o^2 \end{pmatrix}$		$\frac{\tau^2}{\tau^2+o^2+\theta^2}$ †
		$\mathbf{\Lambda} = \boldsymbol{T} + \mathbf{\Delta}$			$\frac{\lambda^2}{\lambda^2+\theta^2}$
	d_i (2.2)	$= \begin{pmatrix} \lambda^2 & \lambda^2 & \cdots & \lambda^2 \\ \lambda^2 & \lambda^2 & \cdots & \lambda^2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^2 & \lambda^2 & \cdots & \lambda^2 \end{pmatrix}$	$\begin{pmatrix} o^2 & 0 & \cdots & 0 \\ 0 & o^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & o^2 \end{pmatrix}$		$\frac{\lambda^2}{\lambda+o^2+\theta^2}$ †
	$(d\gamma)_{ik}$ (2.3)	$\boldsymbol{T} = \begin{pmatrix} \tau^2 & \tau^2 & \cdots & \tau^2 \\ \tau^2 & \tau^2 & \cdots & \tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau^2 & \tau^2 & \cdots & \tau^2 \end{pmatrix}$		$\mathbf{\Psi}^* = \mathbf{\Delta}^* + \mathbf{\Theta}$ $= \begin{pmatrix} \delta_1^2 & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & \delta_2^2 & \cdots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & \delta_n^2 \end{pmatrix} + \begin{pmatrix} \theta^2 & 0 & \cdots & 0 \\ 0 & \theta^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta^2 \end{pmatrix}$	$\frac{\tau_k^2}{\tau_k^2+\psi_k^2}$ ‡
Instability	d (2.4)			$\mathbf{\Psi}^* = \boldsymbol{T}^* + \mathbf{\Theta}$ $= \begin{pmatrix} \tau^2 & \tau_{12} & \cdots & \tau_{1n} \\ \tau_{21} & \tau_2^2 & \cdots & \tau_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{n1} & \tau_{n2} & \cdots & \tau_n^2 \end{pmatrix} + \begin{pmatrix} \theta^2 & 0 & \cdots & 0 \\ 0 & \theta^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta^2 \end{pmatrix}$	$\frac{\lambda^2}{\lambda_k^2+\psi_k^2}$
	d_i (2.5)	$\mathbf{\Delta} = \begin{pmatrix} \delta^2 & \delta^2 & \cdots & \delta^2 \\ \delta^2 & \delta^2 & \cdots & \delta^2 \\ \vdots & \vdots & \ddots & \vdots \\ \delta^2 & \delta^2 & \cdots & \delta^2 \end{pmatrix}$			
	$(d\gamma)_{ik}$ (2.6)			$\mathbf{\Psi}^* = \boldsymbol{T}^* + \mathbf{\Delta}^* + \mathbf{\Theta}$ $= \begin{pmatrix} \tau_1^2 & \tau_{12} & \cdots & \tau_{1n} \\ \tau_{21} & \tau_2^2 & \cdots & \tau_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{n1} & \tau_{n2} & \cdots & \tau_n^2 \end{pmatrix} + \begin{pmatrix} \delta_1^2 & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & \delta_2^2 & \cdots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & \delta_n^2 \end{pmatrix} + \begin{pmatrix} \theta^2 & 0 & \cdots & 0 \\ 0 & \theta^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta^2 \end{pmatrix}$	

Note. d = constant systematic error; d_i = item-specific factor; $(d\gamma)_{ik}$=time-specific factor; T = True value covariance matrix; Δ = systematic error covariance matrix; Θ = random error covariance matrix; Λ = Covariance matrix of the components assumed to be true value; Ψ = Covariance matrix of the components assumed to be error.

$*$ = unstructured covariance matrix

† = agreement coefficient

‡= lower bound of reliability coefficient

2.1.1. 참값의 안정성

반복 측정이 이루어지는 간격이 비교적 짧고 시행 간에 어떠한 처리 또는 성숙이 가해지지 않기 때문에 참값은 변하지 않거나 또는 처리에 의한 변화가 일어나더라도 모든 피험자가 각 시행마다 동일한 크기로 변하는 것을 의미한다. 따라서 각 시행에서 참값의 분산은 동일하며 시행 간 참값의 공분산도 참값 분산과 같다. 계통오차가 일정한 경우 i 번째($i = 1, \ldots, m$) 피험자의 k 번째($k = 1, \ldots, n$) 시행의 측정값을 Y_{ik}라 할 때 모형은 다음과 같다.

$$Y_{ik} = \gamma_0 + \gamma_k + (t_i + d) + e_{ik}, \tag{2.1}$$

γ_0과 γ_k는 각각 총평균, k 번째 시행의 평균으로 고정효과(fixed effect), t_i는 i 번째 피험자의 참값으로 임의효과(random effect), d는 계통오차, e_{ik}는 임의오차이다. 위의 식 (2.1)은 고정효과와 임의효과로 이루어진 선형혼합모형이다. 임의효과에 t_i에 의해 발생하는 변동은 피험자 간 분산(between-subjects variance)으로 참값 분산, 임의오차 e_{ik}에 의해 발생하는 변동은 피험자 내 분산(within-subjects variance)으로 오차 분산이다.
일반적으로 분산은 $n \times n$ 공분산행렬로 나타내며 Table 1에 참값 t_i와 임의오차 e_{ik}의 공분산행렬 $\boldsymbol{T}, \boldsymbol{\Theta}$ 가 있다. 전술한 바와 같이 모든 시행에서 참값 분산은 τ^2이며(대각성분) 시행 간 참값의 공분산도 τ^2이다(비대각성분). 동일한 측정 도구로 반복 시행하기 때문에 임의오차 분산은 각 시행에서 θ^2로 동일하며(대각성분) 임의오차 특성상 서로 독립이라 가정하는 것이 타당하다(비대각성분은 0). 계통오차가 일정하면(d) 변동에 기여하지 않으므로 어느 효과에 포함되어도 무방하나 위 모형은 t_i에 포함되어 있다. 가장 고전적인 모형이며 신뢰도 계수로서 측정값 간의 상관계수 또는 ICC (intraclass correlation coefficient)를 위한 모형이다. 계통오차가 일정한 경우는 주로 물리적 존재(physical existence)에 대한 측정에서 나타난다 할 수 있다(예를 들어, 질량, 거리). 다음은 계통오차가 피험자마다 다르게 나타날 경우 (d_i)의 모형이다.

$$Y_{ik} = \gamma_0 + \gamma_k + (t_i + d_i) + e_{ik}, \tag{2.2}$$

d_i는 참값에 포함되었다. 이는 참값처럼 피험자 간에 변동이 나타난다는 의미이며 주로 형이상학적 존재(metaphysical being)의 측정에서 나타난다 할 수 있다(예를 들면, 심리, 사회적 반응). 만일 측정 도구가 문항(item)의 형태라면 계통오차 d_i는 참값 이외에 그 문항을 피험자마다 다르게 느끼거나 인지하는 문항 자체의 특성이다. 그러므로 d_i는 문항특이성 인자(item-specific factor) (Bentler, 2017; Cronbach 등, 1963; Marsh와 Grayson, 1994; Raffalovich와 Bohrnstedt, 1987; Sideridis 등, 2019)로 볼 수 있으며 신뢰도에 관한 한 임의오차와 구분하여 참값과 같이 묶어 다루는 것이 맞다. 그러므로 Table 1에서 참값 분산은 계통오차 분산 δ^2이 포함된 λ^2이다. 다음은 계통오차가 피험자와 시행에서 다르게 나타나는 경우의 모형이다.

$$Y_{ik} = \gamma_0 + \gamma_k + t_i + \{d_i + (d\gamma)_{ik} + e_{ik}\}, \tag{2.3}$$

$(d\gamma)_{ik}$는 γ_k와 d_i의 교호작용(interaction effect)으로 피험자는 동일 문항을 반복해서 접하지만 시행마다 신체적 또는 심리적 상태에 의해 조금씩 달리 반응하는 것으로 방향은 같으나 크기가 다른 형태의 오차를 의미한다. 이를 시간특이성 인자(time-specific factor) (Cronbach 등, 1963; Marsh와 Grayson, 1994) 또는 일시적 오차(transient error) (Schmidt 등, 2003; Vispoel 등, 2018)로 보는 것이 타당할 것이다. 이 교호작용은 아래에 살펴볼 참값의 불안정성을 설명하는 참값의 교호작용과 별개로 나타날 수 있으므로 분리하는 것이 좋다. $(d\gamma)_{ik}$는 부분적으로 임의오차 특성과 유사하다고 볼 수 있다. 모든 피험자는 각 시행에서 한 번 측정하기 때문에 $d_i + (d\gamma)_{ik}$의 분산과 e_{ij}의 분산을 구분할 수 없으며 $d_i + (d\gamma)_{ik}$를 오차로 간주할 수 밖에 없다. d_i는 사실상 모형에 포함하지 않아도 된다. 오차인 $(d\gamma)_{ik} + e_{ik}$(d_i 생략)의 공분산행렬은 Table 1의 $\boldsymbol{\Psi}$와 같이 시간특이성의 비구조적(unstructured) 공분산행렬 $\boldsymbol{\Delta}^*$와 임의오차 공분산행렬 $\boldsymbol{\Theta}$가 합쳐져 비구조적 형태를 가진다($\boldsymbol{\Psi}^* = \boldsymbol{\Delta}^* + \boldsymbol{\Theta}$).
신뢰도 계수는 전체 분산에서 임의오차 분산이 차지하는 부분을 제외한 분산의 비율로 정의된다. 그러므로 계통오차는 임의오차와 분리하여 참값과 같이 다룬다. 각 시행의 신뢰도 계수는 Table 1에 나타나 있다. 식 (2.1),

(2.2)에 의한 신뢰도 계수는 모든 시행에서 동일한 값을 가지지만 식 (2.3)에 의한 신뢰도 계수는 시행마다 다르다. 특히, 식 (2.3)은 계통오차가 임의오차와 같이 묶여 있어 신뢰도 계수의 과소추정 가능성이 있다.
이 모형들은 참값의 평균 변화는 γ_k가 설명한다. γ_k는 이 외에 시행 절차에서 발생하는 변화도 포함된다. 시행에서 나타나는 평균 변화인 차이를 감안한 상태에서 일치의 정도를 나타낸 것이 일관성(consistency)이며, 고정효과인 γ_k를 임의효과라 가정하여 얻어진 분산 o^2(Table 1의 시행공분산행렬 $\mathbf{O}$)을 동시에 고려한 신뢰도는 절대적인 일치의 정도를 나타낸 일치도(agreement)가 된다. 일치도는 참값의 안정성 가정하의 식 (2.1), (2.2)에서 알아보는 것이 적절하다.

2.1.2. 참값의 불안정성

피험자의 참값 변화가 각 시행에서 다른 크기로 일어나는 것을 의미한다. 예를 들면, 3개월 운동 수행 동안 질량의 변화, 삶의 만족도 변화와 같이 운동 처리에 의한 확실한 참값의 변화가 피험자와 시행 모두 다르게 나타나는 경우이다. 피험자 간 참값의 상대위치(relative position) 크기(magnitude)와 방향(direction)이 시행마다 변한다. 위에서 고려한 세 가지 계통오차를 적용한 모형은 다음과 같다.

$$Y_{ik} = \gamma_0 + \gamma_k + \{t_i + (t\gamma)_{ik} + d + e_{ik}\}. \tag{2.4}$$

$$Y_{ik} = \gamma_0 + \gamma_k + d_i + \{t_i + (t\gamma)_{ik} + e_{ik}\}. \tag{2.5}$$

$$Y_{ik} = \gamma_0 + \gamma_k + \{t_i + (t\gamma)_{ik} + d_i + (d\gamma)_{ik} + e_{ik}\}. \tag{2.6}$$

이 모형들에서 참값 평균 변화는 γ_k가, 피험자의 개별적 변화는 $t_i + (t\gamma)_{ik}$가 설명한다. t_i도 d_i와 마찬가지로 모형에서 생략해도 무방하다. 참값 $(t\gamma)_{ik}$는 모두 오차에 포함되고 식 (2.5)의 d_i만 오차와 구분된다. $(t\gamma)_{ik}$의 공분산행렬 $\boldsymbol{T}$는 비구조적 형태를 가진다(Table 1의 $\boldsymbol{T}^*$). 특히 (2–3)에서$(t\gamma)_{ik}$, $(d\gamma)_{ik}$가 임의오차 e_{ik}와 효과가 혼재되어 있다. 시행의 분산을 구분할 수 없고, $(t\gamma)_{ik}$와 $(d\gamma)_{ik}$의 공분산을 구분할 수 없다($\boldsymbol{T}^*, \boldsymbol{\Delta}^*, \boldsymbol{\Theta}$). 이들을 분리하기 위해서는 가정이 필요하다. 본 연구는 참값의 불안정성은 측정 시행 간격이 긴 종단적 연구에서 주로 나타나는 것으로 가정하였다. 종단적 연구는 처리에 의한 효과인 참값의 변화가 대체로 발생하며 피험자마다 효과가 다르게 나타나는 것을 경험적으로 알 수 있다. 그리고 시행의 간격이 아주 크기 때문에 시간특이성 인자의 작용은 없는 $(d\gamma)_{ik} = 0$을 가정하였다. 종단적 연구에서는 시간특이성의 효과보다는 참값 변화인 불안정성이 나타날 가능성이 훨씬 크기 때문에 이러한 가정은 타당한 것으로 생각된다. 그러면 식 (2.6)은 식 (2.5)가 된다. 그러나 여전히 문제가 남아 있다. d_i를 제외하고 모두 오차로 묶여 있고, 식(2.4)은 심지어 모두 오차로 묶여 있다. 이럴 경우에 우선적으로 Jöreskog (1971)의 동류 검사(congeneric tests) 모형을 오차에 적용하여 분산을 분리할 수 있다. 그리고 오차의 구조를 $\boldsymbol{\Theta}$처럼 대각행렬로 모형화하여 공분산은 참값의 공분산행렬에서 설명하는 것으로 시도해 볼 수 있다.

2.1.3. 공분산 구조

참값이나 오차의 비구조적 공분산행렬은 각각 $n(n+1)/2$개의 독립적인 모수가 있으며 추정을 위해 이보다 적은 수의 모수를 가진 공분산행렬 모형을 이용해야 한다. Table 2는 본 연구의 비구조적 공분산행렬에 적용된 모형으로 n번의 시행이 있다고 가정한 일반적인 공분산 구조를 나타낸 것이다. 각각 척도화 항등(scaled identity; SI), 대각(diagonal; DG), 인자분석(factor analytic; FA), 1차 자기회귀(first-order autoregressive; AR1), 이분산 1차 자기회귀(heterogenous first-order autoregressive; ARH1), 이분산 복합대칭(heterogenous compound symmetry; CSH)을 의미한다. 종단적 연구 신뢰도의 특징은 한 번 시행의 여러 문항 합에 대한 신뢰도인 내적일관성(internal consistency)과 다르게 추정된 공분산 성분에 의해 신뢰도가 직접 결정될 일이 없다는 점이다. 그러나 측정값의 분산이 동일하지 않으면 참값과 오차의 분산을 독립적으로 추정할 수 없기 때문에 (Box, 1954; Maxwell, 1968) 간접적으로 관련성은 있다고 볼 수 있다.

Table 2: Covariance structure models

Model	The number of parameters	Covariance matrix structure
SI	1	$\sigma^2\begin{pmatrix}1 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1\end{pmatrix}$
DG	n	$\begin{pmatrix}\sigma_1^2 & 0 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & 0 & \cdots & 0\\ 0 & 0 & \sigma_3^2 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & \sigma_n^2\end{pmatrix}$
FA	n	$\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2 & \sigma_1\sigma_3 & \cdots & \sigma_1\sigma_n\\ \sigma_2\sigma_1 & \sigma_2^2 & \sigma_2\sigma_3 & \cdots & \sigma_2\sigma_n\\ \sigma_3\sigma_1 & \sigma_3\sigma_2 & \sigma_3^2 & \cdots & \sigma_3\sigma_n\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \sigma_n\sigma_1 & \sigma_n\sigma_2 & \sigma_n\sigma_3 & \cdots & \sigma_n^2\end{pmatrix}$
ARI	2	$\sigma^2\begin{pmatrix}1 & \rho & \rho^2 & \cdots & \rho^{n-1}\\ \rho & 1 & \rho & \cdots & \rho^{n-2}\\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1\end{pmatrix}$
ARHI	n+1	$\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \cdots & \rho^{n-1}\sigma_1\sigma_n\\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \cdots & \rho^{n-2}\sigma_2\sigma_n\\ \rho^2\sigma_3\sigma_1 & \rho\sigma_3\sigma_2 & \sigma_3^2 & \cdots & \rho^{n-3}\sigma_3\sigma_n\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \rho^{n-1}\sigma_n\sigma_1 & \rho^{n-2}\sigma_n\sigma_2 & \rho^{n-3}\sigma_n\sigma_3 & \cdots & \sigma_n^2\end{pmatrix}$
CSH	n+1	$\rho\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2 & \sigma_1\sigma_3 & \cdots & \sigma_1\sigma_n\\ \sigma_2\sigma_1 & \sigma_2^2 & \sigma_2\sigma_3 & \cdots & \sigma_2\sigma_n\\ \sigma_3\sigma_1 & \sigma_3\sigma_2 & \sigma_3^2 & \cdots & \sigma_3\sigma_n\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \sigma_n\sigma_1 & \sigma_n\sigma_2 & \sigma_n\sigma_3 & \cdots & \sigma_n^2\end{pmatrix}$

Note. SI = scaled identity; DG = diagonal; FA = factor analytic; AR1 = first-order autoregressive; ARH1 = first-ord er autoregressive with heterogenous variances; CSH = compound symmetry with heterogenous variances.

2.2. 자료의 분석

두 가지 형태의 종단적 연구자료로 신뢰도 계수를 구하였다. 첫 번째 자료는 버클리 성장연구(Berkeley growth study) (Tuddenham과 Snyder, 1954)에 사용된 것으로 39명의 소년과 54명의 소녀의 키(height)를 1세부터 18세까지 측정한 것이다. 그중 소년의 1, 5, 10, 15, 18세 자료를 이용했다. 원자료는 프로그램 R의 패키지 'fda'에 포함되어 있다. 두 번째 자료는 1975년 시카고 대학의 실험학교(the laboratory school of the university of Chicago)에서 연구한 8–11학년의 읽기 시험 중 어휘(vocabulary) 부분의 점수를 이용했다. 어휘 점수는 비교를 위해 임의의 공통 단위로 변환된 것이며 자료는 R의 패키지 'heplots'에서 얻을 수 있다. 분산성분(variance

Table 3: Observed variances, covariances and correlation coefficients on height data (m = 39)

Trial (year)	1 (1)	2 (5)	3 (10)	4 (15)	5 (18)
1	7.65	.491	.395	.386	.379
2	4.95	13.28	.919	.748	.785
3	5.97	18.32	29.90	.878	.905
4	8.43	21.55	37.96	62.50	.858
5	6.85	18.66	32.29	44.25	42.57
Mean	76.10	111.59	142.169	174.01	180.223

Note. Diagonal elements are variances, lower off-diagonal elements are covariances and upper off-diagonal elements are correlation coefficients.

Table 4: Reliability coefficients on height data by true and error covariance matrix models

	True/Error			
Trial	FA/SI	Stability/SI	ARH1/SI	CSH/SI
1	.2110	.6390	.3462	.2540
2	.6446	.6390	.7252	.6856
3	.8362	.6390	.8718	.8571
4	.9142	.6390	.6553	.6282
5	.8763	.6390	.9062	.8946
$\chi^2(df)$	48.44(9)	158.06(13)	46.37(8)	47.25(8)
p		< 0.001		

Note. Stability corresponds to stability's $\boldsymbol{T}$ or $\boldsymbol{\Lambda}$ in Table 1.

component)의 추정은 SPSS ver 27(IBM)의 선형혼합모형 절차를 이용하였다. 또한 모형의 적합도 평가를 위해 비구조적 공분산행렬 모형과 비교한 가능도비 검정(likelihood ratio test) (Jennrich와 Schluchter, 1986)을 수행하였다.

Table 3은 첫 번째 자료인 키의 통계량으로 평균, 분산, 공분산, 상관계수를 나타낸 것이다. 성숙에 따라 키 관측값의 평균이 점점 커지며, 분산은 점점 증가하다 마지막엔 감소하는 형태를 보였다. 공분산의 크기는 각각 다르며 상관계수를 보더라도 편차가 크다는 것을 알 수 있다. 참값 안정성은 성립되지 않을 것으로 생각되며, 측정 도구가 저울이란 점을 감안할 때 오차는 독립이 적절할 것으로 판단 된다. 오차는 분산이 동일하고 서로 독립인 SI와 참값 공분산행렬의 여러 모형을 적용하여 구한 각 시행의 신뢰도 계수는 Table 4와 같다.

Table 4의 참값/오차 두 번째 열의 안정성/SI에 의한 신뢰도 계수는 고전적인 모형에 의한 것으로 .6390을 보였다. 모든 시행에서 신뢰도 계수가 일정하며 χ^2 = 158.06이고 p < .001이므로 적합성이 떨어진다. 비록 나머지 조합들도 적합도 검정에서 만족스런 결과를 보이진 않았지만 시행마다 다른 신뢰도 계수를 보여 조금은 의미 있는 오차 분석이 가능하다. 시행의 관측 분산 크기만을 보았을 때(Table 3) FA/SI가 이를 잘 반영한 것으로 보인다. FA/SI는 Jöreskog (1971)의 동류 검사 모형과 유사하며 오차를 SI로 적용시킨점이 다르다. ARH1/SI와 CSH/SI는 모두 4번째 보다는 5번째 시행의 신뢰도 계수 크기가 FA/SI와 비교해 보았을 때 크다. 초기에는 참값의 안정성이 낮고 나이가 듦에 따라 안정성이 커진 것으로 판단된다. 이를 FA, ARH1, CSH가 잘 설명하지 못한 것으로 생각한다. 따라서 참값의 불안정성이 오차로 포함되어 크기가 증가되었을 것이다. 그리고 1세 유아의 키를 측정할 때 임의오차도 많이 포함되었을 것으로 생각한다. 1세의 신뢰도가 낮은 이유이다.

Table 5는 두 번째 자료인 어휘 점수의 통계량으로 평균은 점진적으로 증가하며 분산은 Table 3과 유사한 추이를 보였다. 상관계수는 키 자료에 비해 약간 높은 값을 보였다. Table 6은 참값과 오차의 여러 공분산 구조

Table 5: Observed variances, covariances and correlation coefficients on vocabulary data (m = 64)

Trial (Grade)	1 (8)	2 (9)	3 (10)	4 (11)
1	3.568	.810	.867	.785
2	3.190	4.347	.785	.757
3	3.551	3.548	4.704	.811
4	2.853	3.039	3.389	3.708
Mean	1.137	2.542	2.988	3.472

Note. Diagonal elements are variances, lower off-diagonal elements are covariances and upper off-diagonal elements are correlation coefficients.

Table 6: Reliability coefficients on vocabulary data by true and error covariance matrix models

	True/Error						
Trial	FA/DG	FA/SI	Stability/DG	Stability/SI	AR1/DG	ARH1/AR1	AR1/AR1
1	.8625	.7858	.8752	.7991	.9439	.8395	.8444
2	.7426	.8092	.7494	.7991	.7601	.8626	.8444
3	.8642	.8316	.8055	.7991	.8318	.8754	.8444
4	.7440	.7803	.7743	.7991	.8331	.7510	.8444
$\chi^2(df)$	1.648(2)	7.903(5)	6.607(5)	12.167(8)	5.303(4)	7.404(3)	11.714(6)
p	.439	.162	.252	.144	.258	.060	.069

Note. Stability corresponds to stability's $\boldsymbol{T}$ or $\boldsymbol{\Lambda}$ in Table 1.

모형에 의한 네 번 시행의 신뢰도 계수와 적합도 검정을 나타낸 것이다.

Table 6을 보면 고전적인 모형인 안정성/SI의 모든 시행에서 신뢰도 계수는 .7991을 보였으며 나머지는 이에 비해 시행마다 다른 값을 보였다. 가장 좋은 적합도를 보인 것은 FA/DG의 χ^2 = 1.648, p = .439이며 이는 Jöreskog (1971)의 동류 검사 모형을 적용한 것이다. 참값의 불안정성 모형이며 오차는 독립을 가정한 것이다. 오차의 독립을 완화한 ARH1/AR1은 FA/DG와 비교했을 때 2번째 시행의 계수가 차이가 있을 뿐 유사한 결과를 보였으나 적합도는 약간 떨어지는 것으로 나타났다.

반복 시행에서 상관계수가 신뢰도 계수로써 적절한 경우도 있다. 물리적 존재에 대한 측정이 그러할 것이다. Werts 등 (1978)의 연구와 같이 모든 시행에서 신뢰도는 일정할 수도 있다. 그러나 대부분 동일 검사를 비교적 짧은 시간에 반복 시행한다면 피험자의 피로 등에 의한 오차가 개입되어 오차가 독립이 아닐 가능성은 커진다. 동일 검사라도 종단적 연구와 같이 시행 간격이 아주 크다면 이러한 오차 개입 가능성은 낮아지겠지만 학습에 의한 참값의 변화가 피험자마다 다르게 나타날 수 있다. 상관계수가 적절하지 않은 이유이다.

본 연구의 모형은 검사가 타당하지 않으면 계통오차가 작용하고 이로 인해 문항특이성 인자뿐만 아니라 그 검사의 반복 시행으로 인한 심리적, 생리적 상태에 따른 일시적 오차인 시간특이성 인자도 작용하는 것으로 보았다. 즉, 특이성 인자의 근원을 계통오차로 보았다. 모형에서 계통오차는 참값의 변동 이외의 다양한 변동을 설명하기 위해서 꼭 필요하다. 계통오차는 참값과 유사한 특성을 가지고 있다. 기본적으로 반복 시행에서 모두 지속적으로 작용한다. 참값과 계통오차가 안정성 있게 작용한다면 분산과 공분산이 동일한 형태의 공분산행렬 $\boldsymbol{T}, \boldsymbol{\Delta}$가 나타난다. 그러나 불안정성이라면 공분산행렬은 비구조적인 $\boldsymbol{T}^*, \boldsymbol{\Delta}^*$로 나타난다. 문제는 이 비구조적 공분산행렬이 오차에 포함될 수 밖에 없다는 것이다. 더 큰 문제점은 동시에 참값과 계통오차의 불안정성이 발생하는 경우이다(식 2.6). 분리 추정하려면 하나는 안정성을 가정해야 하는데 종단적 연구에서는 시간특이성 인자가 작용하지 않는 것으로 가정하는 것이 타당한 것으로 생각된다. 만일 측정에서 참값의 안정성이 있을 것으로 연구자가 가정한다면 시간특이성 인자의 작용이 없음을 가정하지 않아도 된다. 사실

시간특이성 인자는 필요 없는 오차로 볼 수 있으므로 측정 과정에서 발생하지 않도록 유의해야 한다. 시간특이성 인자의 작용이 없다면 문항특이성 인자는 참값과 묶여 다뤄진다. 이럴 경우 오차의 독립은 성립되고 오차는 임의오차만 남게 되어 참값의 공분산 구조만 잘 선택하면 된다. 모든 특이성 인자의 분산을 분리하여 추정하려면 동일 개념의 여러 문항을 여러 번 반복 시행해야 하지만 여기에도 교호작용으로 인한 계통오차가 임의오차에 포함될 수 있는 가능성을 배제할 수 없다. 신뢰도는 모수와 마찬가지로 인간이 알 수 없는 영역일 수 있다.

종단적 자료에 내적일관성으로 사용되는 동형(parallel), 동등(equivalent), 동류 등의 모형을 적용시킨 연구 (Jagodzinski와 Kühnel, 1987; Miller, 1995; Tisak과 Tisak, 1996)도 있다. 본 연구도 Jöreskog (1971)과 유사한 모형을 적용시켜 타당한 결과를 보여주었다. 종단적 연구의 신뢰도에서 오차 공분산 구조의 선택과 별개로 참값 공분산 구조는 Jöreskog의 동류 모형이 적절하여 우선적으로 선택해도 좋을 듯 하다. 일반화가능도 이론 (generalizability theory) (Cronbach 등, 1963; Briesch 등, 2014; Vispoel 등, 2018; Deshon 등, 1998)에서 다루는 모형은 선형혼합모형과 유사하다. 그러나 변동은 시행에서 동일하며 하나의 신뢰도 계수로 평가를 한다. 즉, 오차영역과 참값 분산이 모든 시행에서 동일한 것으로 가정하며 각 시행에서의 신뢰도 보다는 전체 시행을 고려한 하나의 신뢰도로 측정을 평가한다. 이처럼 종단적 모형에서 시행 횟수까지 고려한 총 신뢰도 (Marcoulides, 2019)는 그리 좋은 정보가 아니라 생각된다. 시행마다 오차의 개입 정도를 알아보는 것이 측정의 질을 평가하기 좋다. 만일 참값의 안정성을 가정한다면 시간의 함수로 나타낸 신뢰도 (Blok과 Saris, 1983; Tisak과 Tisak, 1996)로 설명하는 것도 한 가지 방법일 것이다.

선형혼합모형은 신뢰도를 이루는 참값과 오차의 분산을 분리하여 추정하고 모형화할 수 있기 때문에 신뢰도 분석에 적절하다 (Kim, 2022). 종단적 연구에서 공분산 구조 선택은 고정효과 추론의 정밀성을 위해서 중요하다. 임의효과나 오차 중 어떤 것이라도 공분산 구조를 적절하게 선택하여 조합하면 그만이다. 그러나 신뢰도의 경우에는 임의효과에 해당하는 참값과 오차의 적절한 공분산 구조를 동시에 고려해야 한다. 혼합모형을 이용하면 임의효과, 오차 또는 양쪽 모두의 공분산 구조를 선택할 수 있다 (Littell 등, 2000; Wolfinger, 1993). 그러나 어떤 조합이건 허용되는 것은 아니다. 본 연구의 자료 분석에서 SPSS에서 제공하는 참값과 오차 공분산행렬 모형의 조합에서 해에 수렴하지 못하는 경우도 있었다. 분산을 분리하는 과정에서 나타나는 식별문제(identifiability problems)로 생각된다. 즉, 주어진 정보로는 시간에 따라 달라지는 임의효과들의 분산을 분리하여 추정하기 불충분할 수 있다 (Fitzmaurice 등, 2011). 반복 계산 과정 중에 역행렬(inverse matrix) 계산에서 발생하는 나쁜조건(ill-conditioned)으로 인해 최대가능도추정값(maximum likelihood estimate)은 존재하지 않기 때문이다 (Jennrich와 Schluchter, 1986).

Cronbach (1951)는 동일 검사 반복 시행에서 시행 간격이 ‘0’이 아닌 어떠한 것이든 안정성 계수(coefficient of stability), 간격이 ‘0’으로 접근하면 정밀성 계수(coefficient of precision)로 구분하였다. 두 계수 모두 단기간 시행의 신뢰도에 국한된 것으로 생각된다. 정밀성 계수는 가장 이론적인 신뢰도 계수라 할 수 있으며 안정성 계수는 본 연구에서 사용되는 안정성의 의미와는 다르기 때문에 종단적 연구에서 안정성 계수라는 용어는 적절하지 않아 보인다.

연구자는 종단적 연구에서 신뢰도를 알아보기 위해서는 자료의 특성에 맞는 가정을 해야 한다. 우선 타당한 측정 도구를 선택하는 것이 가장 중요하다. 그리고 측정 시 피험자의 반응이 올바를 수 있게 해야 한다. 단기간의 저울에 의한 질량 측정값의 반복 시행에서와 같이 피험자 질량 참값의 상대위치가 시행마다 유지된다고 가정하면 안정성 참값 모형으로 접근하고 오차의 모형에 집중하는 것이 좋다. 그러나 장기간의 운동에 의한 질량의 변화를 다루는 종단적 연구와 같이 시행마다 피험자의 질량 참값의 상대위치가 달라진다고 가정하면 불안정성 참값 모형으로 접근하는 것이 좋다. 이때 시간특이성 인자의 작용은 없는 것으로 가정하며 동류 모형이 권장된다. 그리고 오차는 비교적 단순한 구조일수록 좋다. 공분산 구조의 선택은 적합도 검정을 참조하지만 전적으로 이에 의해서 결정 되어서는 안된다. 연구자는 측정의 성격, 예측되는 참값의 평균 또는 개별적 변화, 오차의 특성 등을 균형 있게 고려해야 적절한 신뢰도 모형 선택에 도움을 줄 것이다.

3. 결론

검사의 반복 시행이 이루어지는 종단적 연구에서 신뢰도를 위한 다양한 측정의 선형혼합모형을 수립하였다. 측정값의 변동에 기여하는 여러 요인을 임의효과인 참값은 안정성과 불안정성, 계통오차는 세 가지 형태로 구분하여 각 조합으로 설명하였다. 참값과 오차의 여러 공분산 구조를 실제 자료에 적용하여 각 시행에서의 신뢰도를 알아보았다. 종단적 자료의 적절한 신뢰도 분석을 위해 연구자는 연구에서 가해지는 처리, 피험자, 그리고 검사의 특성을 꼼꼼히 살펴보고 측정 절차에서 발생가능한 모든 변동 요인을 잘 생각해 볼 필요가 있다. 그리고 이에 맞는 모형을 선택하고 시행마다 다른 신뢰도 계수로 측정의 질을 평가해야 한다.

References

Bentler PM (2017). Specificity-enhanced reliability coefficients, *Psychological Methods*, **22**, 527–540.

Blok H and Saris WE (1983). Using longitudinal data to estimate reliability, *Applied Psychological Measurement*, **7**, 295–301.

Bost JE (1995). The effects of correlated errors on generalizability and dependability coefficients, *Applied Psychological Measurement*, **19**, 191–203.

Box GE (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and of correlation between errors in the two-way classification, *The Annals of Mathematical Statistics*, **25**, 484–498.

Briesch AM, Swaminathan H, Welsh M, and Chafouleas SM (2014). Generalizability theory: A practical guide to study design, implementation, and interpretation, *Journal of School Psychology*, **52**, 13–35.

Cronbach LJ (1951). Coefficient alpha and the internal structure of tests, *Psychometrika*, **16**, 297–334.

Cronbach LJ, Rajaratnam N, and Gleser GC (1963). Theory of generalizability: A liberalization of reliability theory, *British Journal of Statistical Psychology*, **16**, 137–163.

Deshon RP, Ployhart RE, and Sacco JM (1998). The estimation of reliability in longitudinal models, *International Journal of Behavioral Development*, **22**, 493–515.

Fitzmaurice GM, Laird NM, and Ware JH (2011). *Applied Longitudinal Analysis* (2nd ed), John Wiley & Sons, Hoboken, NJ.

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

Heise DR (1969). Separating reliability and stability in test-retest correlation, *American Sociological Review*, **34**, 93–101.

Jagodzinski W and Kühnel SM (1987). Estimation of reliability and stability in single-indicator multiple-wave models, *Sociological Methods and Research*, **15**, 219–258.

Jennrich RI and Schluchter MD (1986). Unbalanced repeated-measures models with structured covariance matrices, *Biometrics*, **42**, 805–820.

Jöreskog KG (1970). Estimation and testing of simplex models, *British Journal of Mathematical and Statistical Psychology*, **23**, 121–145.

Jöreskog KG (1971). Statistical analysis of sets of congeneric tests, *Psychometrika*, **36**, 109–133.

Kim J (2022). Reliability coefficients of measurement using linear mixed model, *The Korean Journal of Physical Education*, **61**, 359–370.

Littell RC, Pendergast J, and Natarajan R (2000). Modelling covariance structure in the analysis of repeated measures data, *Statistics in Medicine*, **19**, 1793–1819.

Lord FM and Novick MR (2008). *Statistical Theories of Mental Test Scores*, Information Age Publishing (Original work published 1968).

Marcoulides KM (2019). Reliability estimation in longitudinal studies using latent growth curve modeling, *Measurement: Interdisciplinary Research and Perspectives*, **17**, 67–77.

Marsh HW and Grayson D (1994). Longitudinal confirmatory factor analysis: Common, time-specific, item-specific, and residual-error components of variance, *Structural Equation Modeling*, **1**, 116–145.

Maxwell AE (1968). The effect of correlated errors on estimates of reliability coefficients, *Educational and Psychological Measurement*, **28**, 803–811.

Miller MB (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling, *Structural Equation Modeling*, **2**, 255–273.

Palmquist B and Green DP (1992). Estimation of models with correlated measurement errors from panel data, *Sociological Methodology*, **22**, 119–146.

Raffalovich LE and Bohrnstedt GW (1987). Common, specific, and error variance components of factor models: Estimation with longitudinal data, *Sociological Methods* & *Research*, **15**, 385–405.

Schmidt FL, Le H, and Ilies R (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs, *Psychological Methods*, **8**, 206–224.

Sideridis GD, Tsaousis I, and Al-Sadaawi A (2019). An application of reliability estimation in longitudinal designs through modeling item-specific error variance, *Educational and Psychological Measurement*, **79**, 1038–1063.

Smith KW (1974). On estimating the reliability of composite indexes through factor analysis, *Sociological Methods* & *Research*, **2**, 485–510.

Smith PL and Luecht RM (1992). Correlated effects in generalizability studies, *Applied Psychological Measurement*, **16**, 229–235.

Taylor JR (1997). *An Introduction to Error Analysis* (2nd ed), University Science Books.

Tisak J and Tisak MS (1996). Longitudinal models of reliability and validity: A latent curve approach, *Applied Psychological Measurement*, **20**, 275–288.

Tuddenham RD and Snyder MM (1954). Physical growth of California boys and girls from birth to age 18, *University of California Publications in Child Development*, **1**, 183–364.

Vispoel WP, Morris CA, and Kilinc M (2018). Applications of generalizability theory and their relations to classical test theory and structural equation modeling, *Psychological Methods*, **23**, 1–26.

Werts CE, Breland HM, Grandy J, and Rock DR (1980). Using longitudinal data to estimate reliability in the presence of correlated measurement errors, *Educational and Psychological Measurement*, **40**, 19–29.

Werts C, Linn RL, and Jöreskog KG (1978). Reliability of college grades from longitudinal data, *Educational and Psychological Measurement*, **38**, 89–95.

Wiley DE and Wiley JA (1970). The estimation of measurement error in panel data, *American Sociological Review*, **35**, 112–117.

Wolfinger R (1993). Covariance structure selection in general mixed models, *Communications in Statistics: Simulation and Computation*, **22**, 1079–1106.

Received September 3, 2023; Revised December 1, 2023; Accepted December 21, 2023

종단적 연구의 신뢰도

김진욱[1,a]

[a]군산대학교 체육학과

요 약

본 연구의 목적은 동일 검사의 반복 시행이 이루어지는 종단적 연구의 재검사 신뢰도를 알아보는 것이다. 종단적 연구에서 나타날 수 있는 다양한 상황의 측정을 혼합모형을 이용하여 수립하였다. 참값은 안정성과 불안정성, 계통오차는 세 가지 형태의 조합에 의한 측정 모형을 고려하였으며 신뢰도를 정의하였다. 실제 종단적 자료의 적용과 신뢰도 분석을 위해 버클리 성장연구에서의 키 자료와 시카고 대학 실험학교에서의 어휘 점수 자료를 이용했다. 혼합모형을 이용하면 참값과 오차의 공분산 구조를 따로 선택하여 신뢰도를 알아볼 수 있는 장점이 있다. 그러나 연구자는 종단적 연구의 적절한 신뢰도 분석을 위해 피험자와 검사, 그리고 연구에서 가해지는 처리의 특성과 같이 측정에서 발생가능한 변동 요인을 잘 살펴볼 필요가 있다. 그리고 이에 맞는 모형을 선택하고 시행마다 다른 신뢰도 계수로 측정의 질을 평가해야 한다.

주요용어: 종단적 연구, 신뢰도, 선형혼합모형

[1](54150) 전라북도 군산시 대학로 558, 군산대학교 체육학과. E-mail: jinuk@kunsan.ac.kr.