PSYCHOMETRIKA—VOL. 89, NO. 1, 296–316
MARCH 2024
https://doi.org/10.1007/s11336-023-09946-1

A MULTIDIMENSIONAL MODEL TO FACILITATE WITHIN PERSON COMPARISON OF ATTRIBUTES

MARK L. DAVISON

UNIVERSITY OF MINNESOTA

SEUNGWON CHUNG

US FOOD AND DRUG ADMINISTRATION

NIDHI KOHLI AND ERNEST C. DAVENPORT JR.

UNIVERSITY OF MINNESOTA

In psychological research and practice, a person's scores on two different traits or abilities are often compared. Such within-person comparisons require that measurements have equal units (EU) and/or equal origins: an assumption rarely validated. We describe a multidimensional SEM/IRT model from the literature and, using principles of conjoint measurement, show that its expected response variables satisfy the axioms of additive conjoint measurement for measurement on a common scale. In an application to Quality of Life data, the EU analysis is used as a pre-processing step to derive a simple structure Quality of Life model with three dimensions expressed in equal units. The results are used to address questions that can only be addressed by scores expressed in equal units. When the EU model fits the data, scores in the corresponding simple structure model will have added validity in that they can address questions that cannot otherwise be addressed. Limitations and the need for further research are discussed.

Key words: structural equation modeling, item response theory, psychological measurement, educational measurement, latent variable modeling, additive conjoint measurement, quality of life.

Standard 1.14 of the *Standards for Educational and Psychological Testing* American Educational Research Association (2014), p. 27 says "When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretations should be provided." It goes on to say "Moreover, evidence for the validity of interpretations of two or more separate scores would not necessarily justify a statistical or substantive interpretation of the difference between them." At a minimum, the rationale for interpreting differences between two scores for the same person (e.g. interest in engineering versus interest in medical careers) would require evidence that the scores are on the same measurement scale. Being on the same scale means being expressed relative to the same unit and origin. One approach to placing scores on the same scale is to standardize them so they have the same mean and standard deviation in a norming population, but this approach requires the assumption that both dimensions have the same true mean and standard deviation in the norming population. A second approach is to compare the percentile ranks of the two scores, but this requires an assumption that equal ranks correspond to equivalent raw scores. In this paper, we consider the first of these two problems: expressing scores of two measurement dimensions in the same unit and a first step toward placing measurement dimensions on the same scale.

We begin by describing four axioms from the additive conjoint measurement (ACM) literature Campbell (1920, 1928); Krantz et al. (1971); Luce & Tukey (1964); Michell (1990), followed

Correspondence should be made to Mark L. Davison, Department of Educational Psychology, University of Minnesota, Minneapolis 55455, USA. Email: mld@umn.edu

by an introduction to what we call the Equal Units (EU) model, which has appeared in the structural equation model (SEM) literature (e.g., Bradlow et al. (1999); Jeon et al. (2018); Rijmen (2010); Thissen (2013)). We then describe a linear EU model whose expected response variables will satisfy the axioms. Next, we describe several nonlinear multidimensional item response theory (IRT) versions of the EU model that also yield measurements whose expected response probabilities will satisfy the axioms. When the axioms are satisfied, the measurements of A, B, and their conjoint effect (A, B) can be placed on a common scale (Luce & Tukey, 1964; Krantz et al., 1971; Domingue, 2014; Michell, 1990). In the special case of our model, all of the factors or dimensions in the model can be placed on a common scale, or at least the expected response variables satisfy axioms as if they were on a common scale.

1. Additive Conjoint Measurement

Luce and Tukey (1964) introduced the axioms of ACM into psychology. Here, we give only a brief description of the four axioms essential to our developments below. For a more complete and readable introduction to ACM, see (Michell, 1990). For ACM to exist, four axioms must be satisfied: single cancelation, double cancelation, solvability, and the Archimedean principle. Let $\circ$ represent concatenation [i.e., $(A \circ B)$]. Let $f_{AB}(A \circ B)$, $f_A(A)$, and $f_B(B)$ be increasing functions of $A \circ B$, A, and B respectively such that $f_{AB}(A \circ B) = f_A(A) + f_B(B)$.

1.1. Single Cancelation

Let a_1 and a_2 be instances of variable A, and let b_1 and b_2 be instances of variable B. When the single cancelation axiom holds,

$$f_{AB}(a_1 \circ b_1) \geq f_{AB}(a_1 \circ b_2) \quad \Leftrightarrow \quad f_{AB}(a_2 \circ b_1) \geq f_{AB}(a_2 \circ b_2) \tag{1a}$$

Likewise,

$$f_{AB}(a_1 \circ b_1) \geq f_{AB}(a_2 \circ b_1) \quad \Leftrightarrow \quad f_{AB}(a_1 \circ b_2) \geq f_{AB}(a_2 \circ b_2) \tag{1b}$$

1.2. Double Cancelation

Let a_1, a_2, and a_3 be instances of the variable A, and let b_1, b_2, and b_3 be instances of variable B. According to the double cancelation axiom,

$$f_{AB}(a_1 \circ b_2) \geq f_{AB}(a_2 \circ b_1) \text{and} f_{AB}(a_2 \circ b_3) \geq f_{AB}(a_3 \circ b_2) \tag{2a}$$

then

$$f_{AB}(a_1 \circ b_3) \geq f_{AB}(a_3 \circ b_1) \tag{2b}$$

1.3. Solvability

Let a_1, a_2, b_1, and b_2 be instances of A and B such that $a_1 b_1 = a_2 b_2$, then for any three of these values, the fourth one must exist in the set of values for A or B and can be solved for. For instance, if a_1, a_2, and b_1 are given, there must exist a value of B, b_2, such that $a_1 b_1 = a_2 b_2$. Likewise, if a_1, b_1, and b_2 are given, there must exist a value of A, a_2, such that $a_1 b_1 = a_2 b_2$. The solvability axiom implies that the values of A and B are "equally spaced (as are the natural numbers) or dense (as are the rational numbers)" (Michell, 1990, p. 73).

1.4. The Arichimedean Principle

For any two values of A or B, there will exist a real number n such that $a_1 = na_2$ ($b_1 = nb_2$). When the Archimedean principle holds, no difference between two values of A is infinite. Likewise, no difference between two values of B is infinite.

Brogden (1977) showed that if test item response probabilities satisfy the Rasch model or the one-parameter logistic (1PL) model, the expected probabilities will satisfy the axioms. However, he also concluded that if the responses follow a two-parameter logistic (2PL) or three-parameter logistic (3PL) model, the axioms will not be satisfied (See also Karabatsos (2001); Kyngdon (2011); Domingue (2014); Perline et al. (1979)). Several researchers have empirically studied the relations between test or questionnaire item responses and the ACM axioms in one of two ways. First, some authors have directly examined violations of the axioms found in observed item response probabilities of a person ability by item difficulty matrix using a two-way factorial design (e.g., Green (1986); Perline et al. (1979)). In these analyses, A refers to person ability scores, B refers to item difficulty, and the concatenated effects of (A, B) are the observed item response probabilities in the persons by items matrix. These authors have used observed item response probabilities to assess whether the attributes of person ability and item difficulty combine additively in a way consistent with the axioms of ACM. When the axioms are met, the data suggest that person ability and item difficulty can be placed on a common interval scale. The axiom results did not lead to clear-cut conclusions in that, for a given data set, the axioms lead to a large number of tests, and the axioms are satisfied in some but not all of the tests.

Rather than examining violations of the axioms directly, Domingue (2014) and Karabatsos (2001) devised methods for fitting data nonparametrically such that the axioms will be satisfied by the expected response probabilities of the model. The models can be used to evaluate the fit of the model to the data and to derive ordinal scales of measurement. In short, two methods for using the axioms with test/questionnaire data have emerged in the literature. The first is to test the axioms directly by counting violations of the axioms in observed item response probabilities of a person ability by item difficulty matrix. The second is to assess the axioms indirectly by fitting an IRT model whose expected response probabilities satisfy the axioms, and then assessing whether observed item response probabilities are closely approximated by those of the model. Our method is more like the second method, fitting a model in which the expected response probabilities will satisfy the axioms, and then evaluating the fit of the model to the data. Unlike the approach that we propose below, all of the models in the studies above are unidimensional.

2. Measurement of Within Person Differences and Equal Units

In discussing the importance of within-person comparisons across traits or abilities, we begin with a real-life example in which the developments of this paper get us only half the way. A high school student took the Preliminary Scholastic Aptitude Test (PSAT). Prior to taking the test he had been considering engineering as a possible career as he liked and did well in both science and math. Upon receiving his PSAT results, he noticed that his Reading and Writing score was higher than his Math score. This led him to question whether engineering was the field for him, and ultimately, he changed his career plans (medicine). The comparison that he made would require Math and Reading/Writing scores expressed in the same units and with the same origin. Is this assumption realistic for two separate scores standardized and equated independently? Our solution can be useful in deriving measurements with the same unit, but not necessarily the same origin. Within-person comparisons such as this one are common in applied psychology. In vocational counseling, clients are encouraged to compare their scores in several vocational interest areas to identify areas of most interest (Erford, 2012). In clinical psychology, the within-person pattern of high and low scores on a test such as the Minnesota Multiphasic Personality Inventory or the NEO Personality Inventory may suggest a diagnosis or intervention (Wiernik et al., 2021). When several scores are reported in a single score report form, it is customary to make comparisons. Thus, it behooves us to calibrate scores in equal units, if possible, so that the comparisons are legitimate.

Beyond applied measurement in clinical, industrial/organizational, and educational psychology, equal units are important for research purposes. Often these research purposes require equal units but not equal origins. An industrial/organizational psychologist may be interested in the question of whether job applicants have a more restricted range due to self-selection on conscientiousness than on extroversion. However, one can only compare the ranges for conscientiousness and extroversion if both are measured in equal units. Equal units are also important for research questions involving coefficients in growth models. Shin et al. (2013) and Peralta et al. (2022) have studied bivariate growth models for reading and mathematics. These researchers were interested in whether the growth rate was the same in reading and math. For comparison of the growth rates to make sense, reading and math would have to be measured in equal units but they need not have equal origins.

There are also times when independent (predictor) variables need to be in equal units. One may predict job performance from personality variables (Booth et al., 2015; Dilchert, 2007; Davison et al., 2014; Shen, 2011) where the question of interest is whether a unit change in one variable (e.g. conscientiousness) is associated with larger improvements in job performance than is a unit change in a second variable (e.g. extroversion). Such a comparison only makes sense when the two predictors are measured in equal units (Davison et al., 2021).

Equal units can also be useful in intervention research. A researcher may be interested in whether a reading curriculum has an equal effect on comprehension of narrative and expository passages. Comparison of effects on the two abilities is meaningful only if they are measured in equal units. Whenever parents, students, administrators, etc receive a list of scores, it is almost impossible not to compare them. Thus, given the potential importance of measuring traits and abilities in equal units, we now introduce our equal units (EU) model.

The EU model is unusual in that it can be parameterized in two different ways. It can be parameterized as a special case of a simple structure model with correlated factors or as a special case of a bi-factor model. If the simple structure form has S dimensions, the bi-factor form will contain S specific dimensions plus a general dimension. Usually, the researcher's goal will be to estimate and report the simple structure parameterization, but the bi-factor model is a tool for estimating the parameters of the simple structure model. It also is useful in evaluating whether the simple structure factors share a common scale.

In the EU model, a simple structure dimension θ_s^* in the simple structure parameterization is directly related to the general dimension θ_g and a corresponding specific dimension θ_s of the bi-factor parameterization: $\theta_s^* = \theta_g + \theta_s$. In what follows, we show that, if the model fits the data θ_g, θ_s, and θ_s^* will be expressed on a common scale. We then extend this result to show that if there are two simple structure dimensions, θ_s^* and $\theta_{s'}^*$ such that $\theta_s^* = \theta_g + \theta_s$ and $\theta_{s'}^* = \theta_g + \theta_{s'}$, the two simple structure dimensions θ_s^* and $\theta_{s'}^*$ must both be in the same unit because both are in the unit of θ_g. Thus, if the data satisfy the EU model, then the model can be used as the basis for deriving measurements of two or more traits or abilities θ_s^* and $\theta_{s'}^*$ that share a common scale.

There are linear and nonlinear factor/IRT forms of our model. We consider the linear factor version of the model in detail. Then we briefly outline the case for other models: two-parameter logistic (2PL), three-parameter logistic (3PL), graded response, and generalized partial credit (GPC) models. Finally, we illustrate the application of our EU model, both the simple structure and the bi-factor parameterization of the two-parameter IRT form, using Quality of Life data and code that can be found in a Dropbox folder (https://www.dropbox.com/sh/bqzmz7o0u2pac9i/AADAiENo284RpxRvNciBBjiya?dl=_{0}).

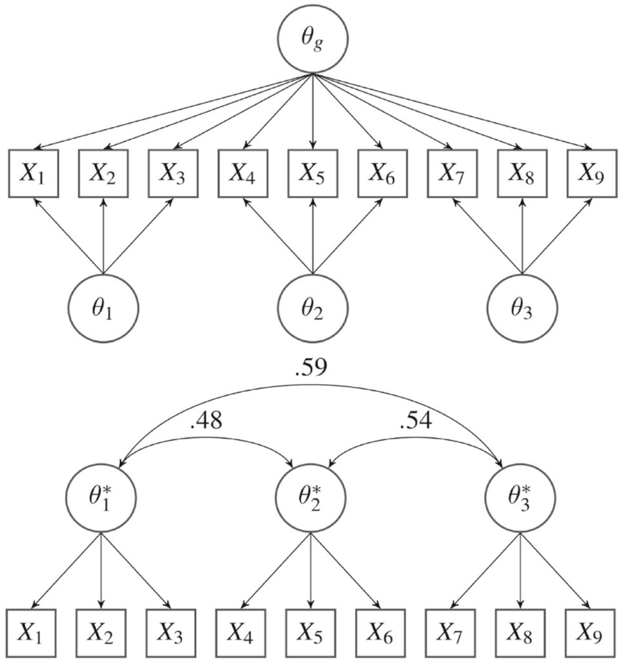

FIGURE 1.
Path diagrams illustrating the latent variable structures of a bi-factor parameterization (top panel) and the corresponding simple structure parameterization (bottom panel) of our model.

We now turn to a description of the linear form for the EU factor model with a discussion of how it satisfies the axioms. The linear form provides a model of continuous variables X. We will describe how the expectations $E(X)$ will satisfy the axioms.

3. The Equal Units Model

In the bi-factor parameterization, our model for the observed variables X_j $(j = 1, \ldots, J)$ is a well-known special case of a bi-factor model in which the factor loading for the general factor is equal to the factor loading for the specific factor (Bradlow et al., 1999; Jeon et al., 2018; Rijmen, 2010; Rijmen et al., 2014; Yung et al., 1999):

$$E\left(X_j\right) = \lambda_{js}\left(\theta_g + \theta_s\right) + \tau_j \tag{3}$$

where λ_{js} is the factor loading for both the general and specific factors and τ_j is the intercept. In a bi-factor model, the general factor and all specific factors are uncorrelated. The top panel of Fig. 1 shows a bi-factor parametrization of our model with a general factor, three specific factors, and nine manifest indicator variables, three indicators for each specific factor.

The simple structure parameterization can be obtained by replacing $\theta_g + \theta_s$ in Eq. 1 with θ_s^* (i.e. $\theta_s^* = \theta_g + \theta_s$).

$$E\left(X_j\right) = \lambda_{js}\left(\theta_s^*\right) + \tau_j \tag{4}$$

The bottom panel of Fig. 1 shows the simple structure counterpart of the bi-factor model in the top panel. The simple structure parameterization will have one factor for each specific factor in the bi-factor parameterization. In our model, the correlations among the simple structure factors are accounted for by a single general factor in the bi-factor parameterization.

As we will show, holding item parameters constant, θ_s^*, θ_g, and θ_s satisfy the axioms of conjoint measurement. This in turn establishes that the three variables are on a common interval scale, or at least the $E(X_j)$ behave as if θ_s^*, θ_g, and θ_s were on a common scale in that incrementing θ_s^*, θ_g, or θ_s by one unit has the same effect on $E(X_j)$for any item j that depends on θ_s^*, θ_g, or θ_s. In this analysis, $A = \theta_g, B = \theta_s, (A \circ B) = \theta_g + \theta_s$, and $f_{AB}(A \circ B) = E(X_j) = \lambda_{js}\left(\theta_s^*\right) + \tau_j = \lambda_{js}\left(\theta_g + \theta_s\right) + \tau_j$ If we define $f_A(A) = \lambda_{js}\left(\theta_g\right)$ and $f_B(B) = \lambda_{js}(\theta_s) + \tau_j$, then

$$f_{AB}(A \circ B) = f_A(A) + f_B(B) = \lambda_{js}\left(\theta_g\right) + \lambda_{js}(\theta_s) + \tau_j \\ = \lambda_{js}\left(\theta_g + \theta_s\right) + \tau_j \tag{5}$$

So long as λ_{js} is positive, all three functions $f_{AB}(A \circ B)$, $f_A(A)$, and $f_B(B)$ are increasing functions of their argument $(A \circ B) = \theta_g + \theta_s$, $A = \theta_g$, and $B = \theta_s$. Typically, items or test scores would be reverse scored, if necessary to make λ_{js} positive.

3.1. *Single Cancelation*

To show single cancelation, we must prove Eq. 1. Translating Eq. 1a into the notation of our EU model, let θ_{g1} and θ_{g2} be two instances of θ_g, and let θ_{s1} and θ_{s2} be two instances of θ_s Then the single cancelation axiom becomes

$$f_{AB}(\theta_{g1} \circ \theta_{s1}) \geq f_{AB}(\theta_{g1} \circ \theta_{s2}) \quad \Leftrightarrow \quad f_{AB}(\theta_{g2} \circ \theta_{s1}) \geq f_{AB}(\theta_{g2} \circ \theta_{s2}) \tag{6}$$

The left side of Eq. 6, $f_{AB}(\theta_{g1} \circ \theta_{s1}) \geq f_{AB}(\theta_{g1} \circ \theta_{s2})$, can be replaced by the factor expression $\lambda_{js}\left(\theta_{g1} + \theta_{s1}\right) + \tau_j \geq \lambda_{js}\left(\theta_{g1} + \theta_{s2}\right) + \tau_j$ to yield

$$f_{AB}\left(\theta_{g1} \circ \theta_{s1}\right) \geq f_{AB}\left(\theta_{g1} \circ \theta_{s2}\right) => \lambda_{js}\left(\theta_{g1} + \theta_{s1}\right) + \tau_j \geq \lambda_{js}\left(\theta_{g1} + \theta_{s2}\right) + \tau_j \tag{7}$$

If we add the quantity $\lambda_{js}(\theta_{g2} - \theta_{g1})$ to both sides of the inequality on the far right of Eq. 7, we get the desired result:

$$\lambda_{js}\left(\theta_{g2} + \theta_{s1}\right) + \tau_j \geq \lambda_{js}\left(\theta_{g2} + \theta_{s2}\right) + \tau_j => f_{AB}(\theta_{g2} \circ \theta_{s1}) \geq f_{AB}(\theta_{g2} \circ \theta_{s2}) \tag{8}$$

This proves the inequality on the right side of Eq. 6.

Single cancelation also requires proving the implication in Eq. 1b. Since the proof of Eq. 1b proceeds exactly the same way as the proof in Eq. 1a, after interchanging the roles of θ_g and θ_s, we omit the proof for Equation 1b.

3.2. *Double Cancelation*

The premises of double cancelation are

$$f_{AB}\left(\theta_{g1} \circ \theta_{s2}\right) \geq f_{AB}\left(\theta_{g2} \circ \theta_{s1}\right) => \lambda_{js}\left(\theta_{g1} + \theta_{s2}\right) + \tau_j \geq \lambda_{js}\left(\theta_{g2} + \theta_{s1}\right) + \tau_j \tag{9}$$

and

$$f_{AB}\left(\theta_{g2}\circ\theta_{s3}\right)\geq f_{AB}\left(\theta_{g3}\circ\theta_{s2}\right) => \lambda_{js}\left(\theta_{g2}+\theta_{s3}\right)+\tau_j\ \geq\lambda_{js}\left(\theta_{g3}+\theta_{s2}\right)+\tau_j \quad (10)$$

We begin by adding the greater quantities on the left sides of the inequalities of Eqs. 9 and 10: $\lambda_{js}\left(\theta_{g1}+\theta_{s2}\right)+\tau_j+\lambda_{js}\left(\theta_{g2}+\theta_{s3}\right)+\tau_j$. Likewise, we add the lesser quantities on the right sides of those same inequalities: $\lambda_{js}\left(\theta_{g2}+\theta_{s1}\right)+\tau_j+\ \lambda_{js}\left(\theta_{g3}+\theta_{s2}\right)+\tau_j$ From Eqs. 9 and 10, we get

$$\begin{aligned}\lambda_{js}\left(\theta_{g1}+\theta_{s2}\right)+\tau_j+\lambda_{js}\left(\theta_{g2}+\theta_{s3}\right)+\tau_j\geq\lambda_{js}\left(\theta_{g2}+\theta_{s1}\right)+\tau_j\\ +\lambda_{js}\left(\theta_{g3}+\theta_{s2}\right)+\tau_j\end{aligned} \quad (11)$$

We now subtract the quantity $\lambda_{js}\left(\theta_{g2}+\theta_{s2}\right)$ from both sides of Eq. 11. This gives the desired result:

$$\lambda_{js}\left(\theta_{g1}+\theta_{s3}\right)+\tau_j\ \geq\lambda_{js}\left(\theta_{g3}+\theta_{s1}\right)+\tau_j => f_{AB}\left(\theta_{g1}\circ\theta_{s3}\right)\geq f_{AB}\left(\theta_{g3}\circ\theta_{s1}\right) \quad (12)$$

3.3. Solvability

As noted by Michell (1990), the solvability axiom is satisfied when the set of values for variable A and the set of values for variable B are dense. Both variables θ_g and θ_s are dense as they include all values on the real number line Hence, the solvability axiom will be satisfied for both variables.

3.4. The Archimedean Principle

Stated in terms of our general and specific factor variables, the Archimedean principle states for every value of θ_{g1} and θ_{g2}, there must be quantity n such that $\theta_{g2}=n\theta_{g1}$ Since θ_g includes every quantity on the real number line, $n=\frac{\theta_{g2}}{\theta_{g1}}$ will exist (except when $\theta_{g1}=0$). Similarly, there must be quantity n such that $\theta_{s2}=n\theta_{s1}$. Since θ_s includes every quantity on the real number line, $n=\frac{\theta_{s2}}{\theta_{s1}}$ (except when $\theta_{s1}=0$). For both the variable θ_g and θ_s, all differences are finite.

The developments above show that the axioms will hold for each dimension in the model. Now consider two of those dimensions s and s'. Because the axioms hold for the first dimension s, θ_g, θ_s, and their sum θ_s^* will be on the same scale. Likewise, for the second dimension, s', the values θ_g, $\theta_{s'}$, and their sum $\theta_{s'}^*$ will be on the same scale. Since θ_s, θ_s^*, $\theta_{s'}$, $\theta_{s'}^*$ will all be on the same scale as θ_g, then θ_s, θ_s^*, $\theta_{s'}$, $\theta_{s'}^*$ must be on the same scale. That is, for any two specific factors, the two specific factor scores, θ_s and $\theta_{s'}$ are on a common scale. Likewise, the simple structure factor scores, θ_s^* and $\theta_{s'}^*$ are on a common scale.

But what does it mean to be on a common scale? First, consider the statement that θ_g and θ_s are on a common scale. It means that for any variable X_j that is a function of $(\theta_g,\ \theta_s)$, a one point increase in θ_g will have the same effect on $E(X_j)$ as will a one point increase in θ_s:

$$E(X_j|\theta_g+1,\theta_s)-E\left(X_j\,|\,\theta_g,\ \theta_s\right)=E\left(X_j\,|\,\theta_g,\ \theta_s+1\right)-E(X_j|\theta_g,\ \theta_s) \quad (13)$$

for all $(\theta_g,\ \theta_s)$.

Now consider the statement that two simple structure factor scores θ_s^* and $\theta_{s'}^*$ are on the same scale. Consider two variables X_j and $X_{j'}$ such that X_j is a function of θ_s^* and $X_{j'}$ is a function of $\theta_{s'}^*$ with equal factor loadings on their respective factors ($\lambda_{j,s1}=\lambda_{j',s2}$) and equal intercepts

$(\tau_j = \tau_{j'})$. If θ_s^* and $\theta_{s'}^*$ are on a common scale, then incrementing θ_s^* by one point will increment $E(X_j)$ by the same amount as incrementing $\theta_{s'}^*$ by one point will increment $E(X_j')$:

$$E(X_j|\theta_s^* + 1) - E\left(X_j \mid \theta_s^*\right) = E\left(X_j' \mid \theta_{s'}^* + 1\right) - E(X_j|\theta_{s'}^*) \tag{14}$$

for all (θ_s^* and $\theta_{s'}^*$).

All of the specific factor scores θ_s and all of the simple structure factor scores θ_s^* have equal units in that all are in the units of the general factor θ_g, and holding item parameters constant, a one unit increase in any of the factors (θ_g, θ_s, or θ_s^*) has the same effect on the expected value of any variable $E(X_j)$.

4. Nonlinear Equal Units Models: Item Response Theory

If the indicator variables are dichotomous or polytomous, nonlinear factor models can be used to model the data. These nonlinear models can be parameterized as either a factor or an IRT model (Takane & de Leeuw, 1987). Since the IRT parameterizations have been more extensively developed in the literature, software, and application, we will focus on the IRT parameterizations. In the IRT parameterization, the multiplicative parameter is an item discrimination parameter, not a factor loading, but we will use the same notation λ_j for the item discrimination parameter as we used previously for the factor loading of variable j. The proofs for the nonlinear models proceed through the same steps as for the linear model above, except that f_{AB} is defined differently for the nonlinear models For each of the models, we will describe the function f_{AB} but we will not repeat the proofs for the four axioms as this would become rather repetitive. We begin with the two-parameter logistic (2PL) model for dichotomous variables X_j.

4.1. Two-parameter Logistic Model

The 2PL model is a model for binary data. The expected response for item j is a probability of the form:

$$\pi_j(X_j = 1) = \frac{\exp[\lambda_j(\theta) + \tau_j]}{1 + \exp[\lambda_j(\theta) + \tau_j]} \tag{15}$$

This can be extended to a multidimensional, simple structure model in which each item j is a function of a single dimension θ_s^*, but the dimension s varies over items:

$$\pi_{js}(X_j = 1) = \frac{\exp[\lambda_j s(\theta_s^*) + \tau_j]}{1 + \exp[\lambda_j s\left(\theta_s^*\right) + \tau_j]} \tag{16}$$

Here, π_{js} is the probability that $X_j = 1$ for item j that is a function of dimension s. The various dimensions in the model are correlated giving rise to the general dimension. The bi-factor parameterization can be obtained by replacing the simple structure person parameter θ_s^* by its bi-factor equivalent $\theta_g + \theta_s$: $\theta_s^* = \theta_g + \theta_s$

$$\pi_{js}(X_j = 1) = \frac{\exp[\lambda_j s(\theta_g + \theta_s) + \tau_j]}{1 + \exp[\lambda_j s\left(\theta_g + \theta_s\right) + \tau_j]} \tag{17}$$

If we take the logit of Eq. 17, we get

$$\ln[\frac{\pi_{js}(X_j=1)}{1-\pi_{js}(X_j=1)}] = \lambda_{js}(\theta_g+\theta_s)+\tau_j \tag{18}$$

and the logit of Eq. 16 is

$$\ln[\frac{\pi_{js}(X_j=1)}{1-\pi_{js}(X_j=1)}] = \lambda_{js}(\theta_s^*)+\tau_j \tag{19}$$

In this model,

$$f_{AB}(A\circ B) = \ln[\frac{\pi_{js}(X_j=1)}{1-\pi_{js}(X_j=1)}] = \lambda_{js}(\theta_g+\theta_s)+\tau_j \tag{20}$$

Recall that in the linear model above,

$$f_{AB}(A\circ B) = E(X_j) = \lambda_{js}(\theta_g+\theta_s)+\tau_j \tag{21}$$

In both the linear and the 2PL model, $f_{AB}(A\circ B)=\lambda_{js}(\theta_g+\theta_s)+\tau_{js}$, and so the steps that establish the axioms for the linear model above will also serve to establish the axioms for the 2PL model.

4.2. Three-parameter Model

For our EU model, the simple structure and bi-factor parameterization of the three-parameter logistic (3PL) model are the same as those for the 2PL model with the addition of a lower asymptote parameter. Analogous to Eq. 16, the simple structure model is

$$\pi_{js}\left(X_j=1\right)=\gamma_j+(1-\gamma_j)\ \frac{\exp[\lambda_{js}(\theta_s^*)+\tau_j]}{1+\exp[\lambda_{js}\left(\theta_s^*\right)+\tau_j]} \tag{22a}$$

$$= \gamma_j+(1-\gamma_j)\ \frac{\exp[\lambda_{js}(\theta_g+\theta_s)+\tau_j]}{1+\exp[\lambda_{js}\left(\theta_g+\theta_s\right)+\tau_j]} \tag{22b}$$

$$= \gamma_j+(1-\gamma_j)\pi_{js}^* \tag{22c}$$

The key element of this equation is the quantity π_{js}^* on the far right of Eq. 22c since it is the only part of the model that depends on the thetas: θ_s^*, θ_g, and θ_s. The 3PL posits that a person can solve the item in one of two ways: by guessing or by reasoning the answer using their ability. π_{js}^* is a conditional probability, the probability that person j solves the item using reasoning given that they do not guess the answer, and π_{js}^* has exactly the same mathematical form as π_{js} in the 2PL model.

Therefore, $f_{AB}(A\circ B)$ for the 3PL model has the same form as in the 2PL if we substitute π_{js}^* into Eq. 20 for π_{js}:

$$f_{AB}(A\circ B) = \ln[\frac{\pi_{js}^*(X_j=1)}{1-\pi_{js}^*(X_j=1)}] = \lambda_{js}(\theta_g+\theta_s)+\tau_j \tag{23}$$

4.3. Graded Response Model

The graded response model (GRM) is a model for polytomous items with K ordered categories ($k = 0,\ldots, K$ - 1). The foundational equation for the graded response model is the representation for $\pi_{js}(X_j \geq k)$, the probability that the response is at or above category k. The simple structure parameterization of the equation is

$$\pi_{js}(X_j \geq k) \quad = \quad \frac{\exp[\lambda_{js}(\theta_s^*) + \tau_{jk}]}{1 + \exp[\lambda_{js}\left(\theta_s^*\right) + \tau_{jk}]} \tag{24a}$$

and the bi-factor parameterization is

$$\pi_{js}\left(X_j \geq k\right) = \frac{\exp[\lambda_{js}(\theta_g + \theta_s) + \tau_{jk}]}{1 + \exp[\lambda_{js}\left(\theta_g + \theta_s\right) + \tau_{jk}]} \tag{24b}$$

The right sides of Eqs. 24a and 24b are the same as the right sides of Eqs. 16 and 17 for the 2PL model except that the intercept is specific to the item category pair (j, k). Therefore, for each (j, k),

$$f_{AB}\,(A \circ B) = \ln[\,\frac{\pi_{js}(X_j \geq k)}{1 - \pi_{js}(X_j \geq k)}\,] = \lambda_{js}(\theta_g + \theta_s) + \tau_{jk}$$

4.4. Generalized Partial Credit Model

Like the GRM, the generalized partial credit (GPC) model is a model for polytomous items with K ordered categories ($k = 0, \ldots, K$ - 1). The foundational equation for the GPC model, however, is the representation for $\pi_{js}(X_j = k)$, the probability that the response is in category k. The simple structure parameterization of the multidimensional equation is

$$\pi_{js}(X_j = k) = \frac{\exp[\lambda_{js}(k\theta_s^*) + \tau_{jk}]}{1 + \sum_{m=1}^{m=K-1} \exp[\lambda_{js}\left(m\theta_s^*\right) + \tau_{jm}]} \tag{25a}$$

and the bi-factor parameterization is

$$\pi_{js}\left(X_j = k\right) = \frac{\exp[\lambda_{js}(k\theta_g + k\theta_s) + \tau_{jk}]}{1 + \sum_{m=1}^{m=K-1} \exp[\lambda_{js}\left(m\theta_g + m\theta_s\right) + \tau_{jm}]} \tag{25b}$$

For Eqs. 25a and 25b, the logit is similar to that for the 2PL model differing only in that the intercept is specific to an item/category combination (rather than to an item) and that there is a multiplier on each person parameter. Taking the logit of Eq. 25b, we get

$$f_{AB}\,(A \circ B) = \ln[\,\frac{\pi_{js}(X_j = k)}{1 - \pi_{js}(X_j = k)}\,] \quad = \quad \lambda_{js}(k\theta_g + k\theta_s) + \tau_{jk} \tag{26}$$

For all models considered here, special cases of the multidimensional linear, 2PL, 3PL, graded response, and GPC models, the axioms will hold. If the model holds, it can serve as a measurement model for which the "true" scores θ_g, θ_s, and θ_s^* are all on a common scale.

TABLE 1.
Item response theory discrimination parameters and fit measures for bi-factor model, simple structure model with equal units, and simple structure model with equal variances of 1.0.

	Bi-factor				Simple structure equal units			Simple structure equal variances		
	G	S1	S2	S3	D1	D2	D3	D1	D2	D3
Fam 1	2.53	2.53	0	0	2.53	0	0	3.81	0	0
Fam 2	1.47	1.47	0	0	1.47	0	0	2.21	0	0
Fam 3	3.02	3.02	0	0	3.02	0	0	4.54	0	0
Fam 4	3.07	3.07	0	0	3.07	0	0	4.61	0	0
Fin 1	2.68	0	2.68	0	0	2.69	0	0	3.70	0
Fin 2	1.60	0	1.60	0	0	1.60	0	0	2.21	0
Fin 3	2.67	0	2.67	0	0	2.67	0	0	3.68	0
Fin 4	2.63	0	2.63	0	0	2.63	0	0	3.63	0
Hlth 1	1.38	0	0	1.38	0	0	1.38	0	0	1.69
Hlth 2	1.59	0	0	1.59	0	0	1.59	0	0	1.95
Hlth 3	1.39	0	0	1.39	0	0	1.39	0	0	1.70
Hlth 4	1.35	0	0	1.35	0	0	1.35	0	0	1.65
Hlth 5	1.95	0	0	1.95	0	0	1.95	0	0	2.39
Hlth 6	1.37	0	0	1.37	0	0	1.37	0	0	1.68
Variance	1.00	1.26	0.90	0.50	2.26	1.90	1.50	1.00	1.00	1.00
-2LL	8337.84				8337.84			8337.84		
AIC	8399.84				NA			8399.84		
BIC	8535.41				NA			8535.42		
RMSEA	0.05				NA			0.05		

G = general factor, S1–S3 = specific factors 1–3, D1–D3 = simple structure dimensions 1–3, Fam 1–4 = Family items 1–4, Fin 1–4 = Financial items 1–4, Hlth 1–6 = Health items 1–6, -2LL = -2 Log Likelihood, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, and RMSEA = Root Mean Square of Approximation. NA = Not Applicable because, by fixing simple structure dimension variances to values estimated from the data, the analysis incorrectly counts parameters making the AIC, BIC, and RMSEA incorrect; correct values in the bi-factor column.

5. Example

In this example, the EU bi-factor model is fitted to 14 items measuring Quality of Life dimensions taken from data files accompanying the IRTPRO 4.2 software (Scientific Software International, 2011). The items used correspond to three dimensions: satisfaction with Family (4 items), Finance (4 items), and Health (6 items). While the data are polytomous, we dichotomized it by recoding categories 0– 3 as 0 and 4 – 6 as 1. Thus, responses indicating that the respondent was more satisfied than dissatisfied were coded as 1. Responses indicating dissatisfaction or equally satisfied and dissatisfied were coded 0. The sample contained 586 chronically mentally ill patients. Originally, all items were rated on a common six-point scale lending some support to the hypothesis that they measure constructs that might be expressed on a common scale.

We first fit a 2PL bi-factor model with one general and three specific dimensions: Family, Finance, and Health. The general and specific discrimination parameters for each item were constrained equal. The variance of the general dimension was set to 1.0, the variances of the specific dimensions were free to vary, and all dimension covariances were set to 0.0. The discrimination parameters are shown in the left panel of Table 1. Note each item has the same discrimination parameter on its general and specific dimension. The variances for the Family, Finance, and Health dimensions were 1.26, .90, and .05. RMSEA was .05 indicating good fit.

In preparation for fitting the simple structure parameterization, we used the bi-factor results to estimate the variance of each simple structure dimension. In a simple structure parameterization corresponding to the bi-factor parameterization, the simple structure theta value θ_s^* is the sum $\theta_s^* = \theta_g + \theta_s$. Since θ_s and θ_g are orthogonal, the variance $\sigma^2\left(\theta_s^*\right)$ is the sum of the general and specific dimension variances: $\sigma^2\left(\theta_s^*\right) = \sigma^2\left(\theta_s\right) + \sigma^2\left(\theta_g\right)$. For Family, the general factor variance is 1.0 and the Family specific variance is 1.26, so the Family simple structure variance is 2.26. Similarly, the variance for Finance is .90 plus 1 = 1.90. Finally, the Health simple structure variance is 1.00 plus the Health specific variance (.50), 1.50.

Next, we fitted the simple structure parameterization with three dimensions (Family, Finance, Health) corresponding to the bi-factor specific dimensions. Results are shown in the middle panel of Table 1. In fitting the simple structure model, we fixed the simple structure dimension variances to values from the bi-factor results: 2.26 for Family, 1.90 for Finance, and 1.50 for Health, and we fixed the dimension covariances to 1.0. Equation 27 shows why the covariance for each pair of simple structure dimensions will be 1.00 given that the variance of the general factor is 1.00 and the covariances of the general and specific factors equal 0.00.

$$\sigma\left(\theta_s^*\theta_{s'}^*\right) = E\left(\theta_s^*\theta_{s'}^*\right) = E\left(\theta_g + \theta_s\right)\left(\theta_g + \theta_{s'}\right) = E\left(\theta_g\theta_g + \theta_s\theta_{s'}\right) = \sigma^2\left(\theta_g\right) + \sigma\left(\theta_s\theta_{s'}\right) = 1.00 + 0.00 = 1.00 \tag{27}$$

Thus, all values in the simple structure variance/covariance matrix were fixed based on the bi-factor variance results and the covariances implied by the EU bi-factor model with general factor variance 1.0. While we freely estimated the discrimination parameters for each item, each item's discrimination parameter estimate in the simple structure model is identical (within rounding error) to its discrimination in the bi-factor model.

The -2 log likelihood (-2LL) fit measure for the simple structure model is shown in the middle panel of Table 1. It is the same as for the bi-factor model (left panel). When the bi-factor model has three specific dimensions, the simple structure and bi-factor model are just reparameterizations of each other. For comparison, the right panel of Table 1 shows parameters and fit measures from a conventional simple structure solution in which all dimension pairs are constrained to have variance 1.00. The correlations of the dimensions are shown in the simple structure diagram in the bottom panel of Fig. 1.

In the Quality of Life domain, individual differences are inequalities of life. Dimension variances indicate the breadth of those inequalities as perceived by respondents. Differences in the variances suggest that family life issues lead to wider perceived inequalities of life than do finance or health issues. This would be surprising since society tends to be very concerned about inequalities, but more in the health and finance areas than in the area of family. However, the dimension variance results here must be interpreted with caution for two reasons First, the sample is clinical: 586 chronically ill patients. The second reason is shown in Fig. 2.

Figure 2 shows the dimension variance estimates with upper and lower error bands equal to 1.96 times their standard errors.[1] Every pair of these error bands overlaps to some degree. This raises the question of whether there are differences in the population variances across the three variables in the EU model. To address this issue, we fitted an EU model that constrained all of the specific dimension variances to be equal. In the corresponding simple structure model, the dimension variances are all equal. This model will be discussed below. In closing this section, we note that dimension variances in the EU model can be used to compare the breadth of perceived inequalities in different aspects of life.

[1] Over samples, standard errors of variance estimates are unlikely to follow a normal distribution, and so the error bands in Fig. 3 should be interpreted descriptively rather than inferentially. Nevertheless, these error bands led us to examine a model with both equal units *and* equal variances. See below.

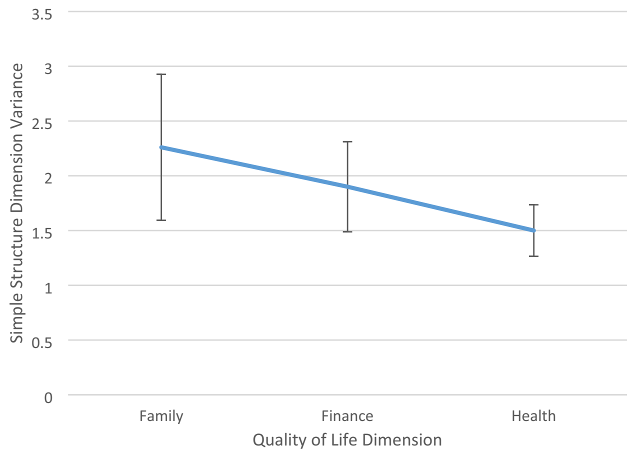

FIGURE 2.
Quality of Life Simple Structure Dimension Variances with 95% Confidence Intervals.

It should be noted that when we estimated person parameters for each simple structure dimension using the *expected a posteriori* (EAP) method, the scores did not have the variances shown in Table 1. Therefore, we standardized them to have those variances (2.26, 1.90, and 1.50) before proceeding to the regression analysis described in the next section.

5.1. *Regression Predicting Overall Quality of Life from Family, Finance, and Health IRT Scores: Equal Units but Unequal Variances IRT Measurement Model*

As noted by Davison et al. (2021), when predictors are measured in equal units, regression weights can be compared to determine if a unit increase in one predictor produces the same expected increase in the criterion as a unit increase in another. In a study predicting weight gain, one may be interested in whether a calorie increase in fats produces the same expected weight gain as a calorie increase in proteins or carbohydrates. Here, we are interested in whether a unit increase of Family life quality is associated with the same expected increase in overall quality of life as is a unit increase of Finance life quality or a unit increase in Health life quality. Item 1 in the data set is a rating of overall Quality of Life. The overall Quality of Life score was a polytomous rating on a scale of 0 – 6. In this regression, we predicted the polytomous overall Quality of Life rating from Family, Finance, and Health quality of life IRT scores measured in the equal units of the general dimension in the bi-factor model.

Here, we fit two regression models. The first was a reduced model where the three predictor weights (Family, Finance, and Health) were constrained equal as described by Davison et al. (2021). In this model, a one unit increase in any predictor is associated with the same expected increase in the criterion irrespective of whether it comes from Family, Finance, or Health. Then we fitted the usual regression model in which the three regression weights were allowed to vary. Results of these two regressions appear in the left panel of Table 2. The reduced model accounted for 31% of the variation while the full model accounted for 33%. While this is a small difference, the hypothesis that the R^2s for the two models were equal was rejected: $F_{2,\,582} = 8.33$, $p < .001$. The variation accounted for by the full model is significantly larger. The adjusted R^2 is also larger for the full model (.33 vs..31) and the AIC and BIC for the full model (unequal weights model) are smaller. We conclude that the full model fits the data better as indicated by the R^2s, adjusted R^2s, AIC, BIC, and F statistic but the improvement is not large.

TABLE 2.
Regression weights and fit measures for equal weights and unequal weights models with the three dimensions from the simple structure model with equal units as predictors.

	Equal units, unequal variances		Equal units, equal variances	
	Equal weights	Unequal weights	Equal weights	Unequal weights
Constant	3.679**	3.679**	3.679**	3.679**
Family	0.269**	0.289**	0.270**	0.323**
Finance	0.269**	0.027	0.270**	0.036
Health	0.269**	0.514**	0.270**	0.449**
R^2	0.311**	0.330**	0.312**	0.330
Adjusted R^2	0.310	0.327	0.311	0.327
AIC	436.07	423.54	434.87	423.34
BIC	444.82	441.03	443.61	440.84

**p <.01, *AIC* Akaike Information Criterion; *BIC* Bayesian Information Criterion.

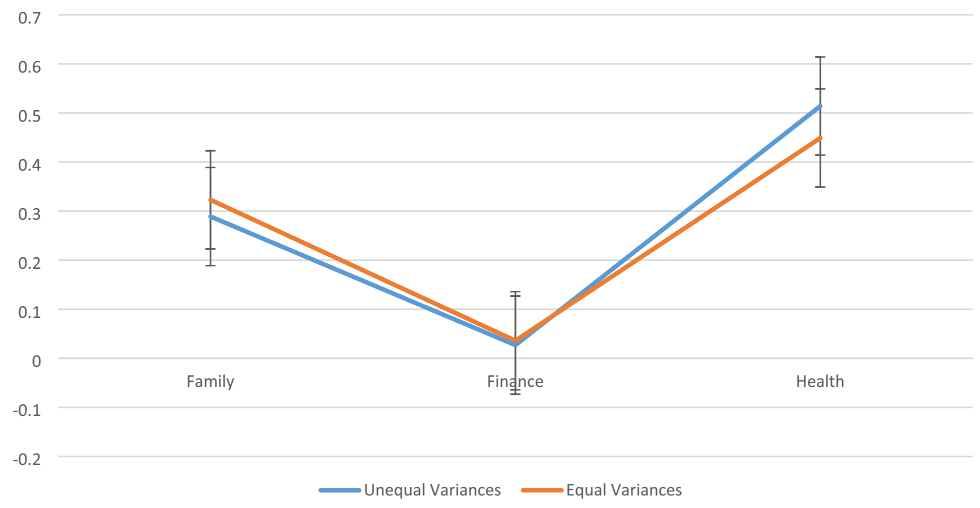

FIGURE 3.
Regression Weights for Simple Structure Quality of Life Variables with Equal Units but Unequal Variances and Equal Units and Equal Variances.

The blue line of Fig. 3 shows the regression weights for the full model along with normal theory 95% confidence intervals. Health has the largest regression weight, but the error band overlaps that for Family. The Finance regression weight is not significant (p >.05), and its error band does not overlap with that of Family or Health. This suggests that a unit increase in perceived Financial quality of life produces a smaller increase in overall Quality of Life than a unit increase in either Family or Health. Thus, a unit of Financial quality is less important than either a unit of Family or Health quality. This may seem surprising, but money may function as a means to an end, not an end itself, whereas both health and good family relationships are ends in themselves. Financial inequalities that do not affect health, family relationships, or other significant outcomes, may contribute less to overall life quality. Once the important outcomes such as health and family relationships are statistically controlled, money becomes less important. Given the nature of this sample (clinical), these results should be considered as illustrative of the method and not as support for substantive conclusions applicable to the general population.

TABLE 3.
Discrimination parameters and fit measures for bi-factor model with equal specific variances, simple structure model with equal units and equal variances, and simple structure model with equal variances of 1.0.

	Bi-factor				Simple Structure Equal Units			Simple Structure Equal Variances		
	G	S1	S2	S3	D1	D2	D3	D1	D2	D3
Fam 1	2.77	2.77	0	0	2.77	0	0	3.81	0	0
Fam 2	1.63	1.63	0	0	1.63	0	0	2.21	0	0
Fam 3	3.33	3.33	0	0	3.32	0	0	4.54	0	0
Fam 4	3.42	3.42	0	0	3.42	0	0	4.61	0	0
Fin 1	2.72	0	2.72	0	0	2.72	0	0	3.70	0
Fin 2	1.62	0	1.62	0	0	1.62	0	0	2.21	0
Fin 3	2.68	0	2.68	0	0	2.68	0	0	3.68	0
Fin 4	2.66	0	2.66	0	0	2.66	0	0	3.63	0
Hlth 1	1.25	0	0	1.25	0	0	1.25	0	0	1.69
Hlth 2	1.42	0	0	1.42	0	0	1.42	0	0	1.95
Hlth 3	1.25	0	0	1.25	0	0	1.25	0	0	1.70
Hlth 4	1.19	0	0	1.19	0	0	1.19	0	0	1.65
Hlth 5	1.76	0	0	1.76	0	0	1.76	0	0	2.39
Hlth 6	1.22	0	0	1.22	0	0	1.22	0	0	1.68
Variance	1.00	0.87	0.87	0.87	1.87	1.87	1.87	1.00	1.00	1.00
-2LL	8341.93				8341.93			8337.84		
AIC	8399.93				NA			8399.84		
BIC	8526.76				NA			8535.42		
RMSEA	0.05				NA			0.05		

G = general factor, S1–S3 = specific factors 1–3, D1–D3 = simple structure dimensions 1 – 3, Fam 1 – 4 = Family items 1–4, Fin 1–4 = Financial items 1–4, Hlth 1–6 = Health items 1–6, -2LL = -2 Log Likelihood, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, and RMSEA = Root Mean Square of Approximation. NA = Not Applicable because, by fixing simple structure dimension variances to values estimated from the data, the analysis incorrectly counts parameters making the AIC, BIC, and RMSEA incorrect; correct values in the bi-factor column.

5.2. Equal Units, Equal Variances IRT Model

There is a special case in which the model with equal dimension units also has dimensions with equal variances. In the EU simple structure parameterization corresponding to a bi-factor model, the variance of a simple structure factor is the sum of the variance for the general and specific factors. Two simple structure dimensions will have equal variances if the corresponding specific dimensions have equal variances. In our example, if the bi-factor Family, Finance, and Health specific variances are equal, then the simple structure Family, Finance, and Health variances will also be equal. As seen in Fig. 2, the error bands for the variances of our earlier simple structure model overlap, suggesting that these variances may be equal.

For reasons explained above, if the general factor in the bi-factor model has variance 1.0, then the covariance between any two pairs of simple structure dimensions will equal 1.0 (the variance of the general factor). Thus, in the special case of equal variances and equal units, the simple structure dimension variance/covariance matrix displays compound symmetry: diagonal elements equal to each other and off-diagonal elements (covariances) equal to each other. All the variances are equal and all of the covariances are 1.00.

To test the equal variance hypothesis, we fit a bi-factor model in which every item had an equal discrimination for its general and specific dimension, and the specific dimension variances were set equal. The variance of the general dimension was set to 1.0. Dimension covariances were set to 0.0.

This solution is shown in the left panel of Table 3. The variance for each specific dimension is.87. We then fit a three-dimensional simple structure model constraining each dimension's variance to 1.87 (sum of the general and specific dimension variances). For every pair of dimensions, the covariance was fixed to 1.0. The simple structure discrimination parameters are shown in the middle panel of Table 3 with corresponding fit measures. The likelihood ratio test comparing the fit of the simple structure, equal units but unequal variance model in Table 1 and the equal unit, equal variance simple structure model of Table 3 was $\chi_2^2 = 4.09,\ p= .13$. We could not reject the hypothesis that the models fit equally well. Both AIC and BIC are smaller for the equal units, equal variance model of Table 3. The error bands in Fig. 3, the likelihood ratio test, the AIC, and the BIC all favor the simple structure model with both equal variances and equal units. There is little evidence to contradict the hypothesis that inequalities in quality of life associated with Family, Finance, and Health are equally broad, at least in this population.

As in the example above, when we estimated person parameters in the simple structure model, those parameter estimates did not have the variances shown in Table 3. Therefore, before performing the regression analysis below, we standardized the person parameters along each dimension to have a variance of 1.87.

5.3. Regression Predicting Overall Quality of Life from Family, Finance, and Health IRT Scores: Equal Units and Equal Variances IRT Measurement Model

We repeated the regression of Table 2 using factors from the equal units, equal variances model (Table 3) as predictors. Results appear in the right panel of Table 2. While the regression weights changed in size, major features of the result were the same. The weight for Finance was not significant. R^2 and adjusted R^2 were larger for the unequal weights model and the F statistic led to rejection of the hypothesis that the two models fit equally well. As in the earlier results, AIC and BIC were smaller for the unequal weights regression model. The Finance variable had the smallest regression weight, and its error band did not overlap with that for Health or Family, but the Health and Family error bands did overlap.

While our presentation has been mainly as an illustration, the overall results provide little or no support for differences in the variation of perceived quality of life IRT scores associated with the Family, Finance, and Health dimensions. The regression analyses, however, suggest that after controlling for the other predictors, a unit increase in Finance quality of life has less effect on overall quality of life than does a unit increase associated with Family or Health issues.

6. Discussion

There are research questions that are meaningless unless variables are measured in equal units. Comparing the variances of two variables requires that they be expressed in equal units. Predictor variables need to be in equal units if we want to identify the most important predictor defined as the variable for which a unit increase yields the largest change in the expected value of the criterion. Comparing rates of growth for two variables requires that those variables be measured in equal units. Comparing the effects of an intervention on two outcome variables requires that the outcome variables be measured in equal units.

Where such comparisons are of interest, we propose a method of calibrating latent variables in multidimensional SEM and IRT simple structure models so that variables in the model share an equal unit. The procedure requires two steps. The first is a pre-processing step to determine if the data fit the EU model and, if so, to estimate the latent variable variances when expressed in an equal unit. This first step involves fitting the EU bi-factor parameterization, a hierarchical model in which (a) all latent variables are uncorrelated, (b) the general factor has variance 1.0, and (c) the variances of all specific factors are freely estimated. By adding variance estimates of the general and specific factors as described above, one can estimate the variances of the latent variables in the corresponding simple structure parameterization with equal units. The second step involves fitting the simple structure parameterization. One first would calibrate the items by fitting the simple structure parameterization fixing the factor variances to the values computed in step 1, rather than fixing them to 1.0. One would also fix the factor covariances to 1.0. Then one would estimate person parameters, standardizing them so they have the variances computed in step 1.

In the special case of an EU model in which the simple structure factors also have equal variances, the bi-factor analysis is unnecessary. In the equal units simple structure parameterization, the covariances among factor pairs are all equal. If their variances are equal as well, then the correlations among pairs of dimensions are also equal. For instance, consider our equal units, equal variances model in which all simple structure variances were set to 1.87 and all covariances were set to 1.00. For any of the simple structure dimension pairs, the correlation is $\rho = \frac{1}{\sqrt{1.87}\sqrt{1.87}}$. If the variances are all equal and the covariances for every pair of dimensions are all equal, the variance-covariance matrix will have a compound symmetric structure. Such a model can be specified without benefit of a bi-factor analysis simply by fixing all simple structure factor variances to the same value — say 1.00 — and then constraining the covariances (correlations if variances equal 1.00) to be equal. The bi-factor analysis is unnecessary only in the special case of equal variances in the equal units model. This equal units, equal variances model may be of particular interest to researchers who want to fit a traditional simple structure model with all variances equal to 1.0, and who are wondering if this equal variances model is also an equal units model. One can fit the traditional simple structure model with all variances equal to 1.0 with and without constraining all factor correlations equal. If the two models fit equally or do not differ significantly, one can retain the equal units, equal variance model for purposes of researching questions that require equal units.

Our results lead to the conclusion that multidimensional 3PL and 2PL models can sometimes satisfy the axioms. At first glance, that might seem to contradict conclusions of Brogden (1977); Karabatsos (2001); Kyngdon (2011); Domingue (2014); Perline et al. (1979) that the probabilities of the one-parameter and Rasch models will satisfy the axioms, but not those of 2PL and 3PL models. Their conclusions apply to unidimensional models only. Our models have no implication for the unidimensional case. Our models include both a general and specific factors/dimensions, and therefore have no unidimensional counterpart. There is no bi-factor model with one factor. In a certain respect, the unidimensional literature contains a conclusion similar to our own: the axioms of ACM are satisfied only in special cases. In the unidimensional situation, the special case is the Rasch model. In the multidimensional, special case, each item has the same loading/discrimination parameter on its general and specific factor.

7. Limitations and Future Research

The approach described above is model based, and the model on which it is based may not fit the data. To date, the evidence on fit of the model is limited and mixed. Jeon et al. (2018) fit the linear form of the model to two datasets. For their first dataset, it was the best fitting model (labeled the proportional model of their Table 2) and for the second, it was not (labeled the proportional model of their Table 5).[2] Rijmen (2010) and Thissen (2013) successfully fit IRT versions of our EU model. If the model does not fit, the approach described above is not applicable. Furthermore,

[2] Jeon et al (2018) fit a model that they call the proportionality model that is a third way to parameterize the EU model. In the proportionality model, all specific factors are constrained to have equal variances. In this bi-factor parameterization, the general and specific factor coefficients for a variable are proportional, rather than equal. See Jeon et al. for details.

there may be cases in which the users have no interest in comparing variables in which case, our procedure is moot as the variables need not be in equal units.

By assuming equal covariances among the simple structure factors, the model may seem quite restrictive. Assuming equal covariances does not mean that the factor correlations are equal as correlations are functions of both the covariances and the variances of the factors. Furthermore, the model is not quite as restrictive as it may seem at first glance. One way to measure the restrictiveness (or parsimony) of a model is to count the number of freely estimated parameters. As an example, consider our linear model compared to a simple structure model with correlated factors, probably the most common multidimensional model in psychological research. Our model has only slightly fewer estimated parameters than the simple structure model with correlated factors. In such a simple structure model with V variables and K factors, the simple structure model has V factor loading parameters, V intercept parameters, and $K(K-1)/2$ factor covariances for a total of $2V+K(K-1)/2$ parameters. Our model has V factor loading parameters, V intercept parameters, and K factor variances for a total of $2V+$ K parameters. When $K=3$, there is no difference in the number of freely estimated parameters. When $K=4$, there is a difference of two parameters; when $K=5$, there is a difference of five; etc. The difference in number of parameters grows as the number of factors increases, but it is not large unless the number of simple structure factors is large. As the number of factors grows, our model becomes more parsimonious compared to a full simple structure model, and parsimony is a scientific virtue. Our model eliminates the need to estimate factor covariances, parameters that are difficult to estimate often resulting in large standard errors. Whether our model is overly restrictive is an issue that should primarily be decided by empirical fit to data, not researcher opinions of the constraints or the number of estimated parameters.

7.1. Multidimensional versus Unidimensional SEM or IRT

In developing measurement instruments, students commonly ask “Why use SEM or IRT?” A related question is “Why use a multidimensional analysis rather than sequential unidimensional ones?” Our discussion suggests an answer to both questions. If an EU model fits the data, then one may be able to derive IRT based scores expressed in equal units. This makes comparisons of standard deviations across variables meaningful. It also makes comparisons of coefficients in linear models meaningful. One can then compare growth over time on two measures. One can also compare the effect of an intervention on two different outcome variables. Measurements in equal units cannot be derived using simple summation of item responses nor sequential, unidimensional IRT/SEM. It requires multidimensional SEM or IRT modeling.

7.2. Importance of Predictors in Multiple Regression

In multiple regression, there is a substantial interest in assessing the relative importance of predictor variables. The most important variables are the ones that account for the most variation in the criterion. One can also compare variables in terms of incremental variation accounted for, the variation accounted for over and above that accounted for by other predictor variables. Dominance analysis Budescu (1993); Azen and Budescu (2006); Johnson and Lebreton (2004) provides another method of comparing predictors based on variation accounted for. A regression coefficient β equals the expected increase in the criterion variable Y given a unit increase in the predictor conditional on remaining predictors. If two predictors are measured in equal units, then the one with the larger regression coefficient is the one that yields a larger expected increase in Y for a unit increase in the predictor. By this criterion, the most important predictor is the one that yields the largest increase in the expected value of Y for a unit increase in the predictor. Davison et al. (2021) discuss the comparison of regression coefficients for predictors expressed in equal units.

7.3. Interventions

The EU hierarchical model provides a way to decompose the simple structure dimensions $\theta*$ into general and specific factors. There may be situations in which the hierarchical decomposition is of interest for reasons that go beyond the issue of equal units. Consider an educational intervention to improve reading. The participants have taken a reading test but also a math and a writing test as pre- and post-measures. Based on a bi-factor analysis, one could derive specific dimension scores in reading, math, and writing plus a general factor. Assuming the intervention has an effect, reading scores would improve pre- to post-, but which component of reading improved? Did the improvement come from improvements in the specific reading factor, the general factor, or both? If there were improvements on the general dimension, one would also expect improvements in math and writing. The bi-factor model may provide a way to study the generalization of learning across academic content domains.

As another example, consider an eating disorders intervention. Participants are given personality scales pre- and post-. Scales include items that seem specific to eating disorders. For instance, the self-esteem scale asks if participants are proud of their body image. In contrast, some items seem more general: are you proud of yourself? The question becomes whether the effects of the intervention influences scores on aspects specific to eating disorders or to more general feelings about one self? Bi-factor decomposition of scores may be used to address questions about the breadth of the intervention effects given that several related measures are administered pre- and post-. For the treatment effects to be compared across the related measures (or across general and specific dimensions), the measures need to be in equal units.

7.4. Individual Differences Theory

To conclude that the several dimensions in a model are measured in equal units, the equal units conclusion must be empirically defensible and theoretically plausible. Such a conclusion is plausible in at least two situations. The first is when there are several sources contributing to the measured attribute. For instance, one's quality of life may be based on contributions to quality coming from several sources: one's family relationships, one's health, and one's financial circumstances. All three types of contributions may be measurable in the same units. Those three variables may behave like separate factors simply because the contributions from the three sources are only moderately correlated across people. To use a nonpsychological example, one may receive caffeine from two sources cola and coffee. The sources are different but their contributions can be measured in the same caffeine units.

A second situation in which dimensions might be expressed in common units is when an attribute has several forms of expression. For instance, creativity may find expression through composing music, writing novels, inventing gadgets, etc. The various types of expression may be measurable in common units. To pick a nonpsychological example, calories can come in different forms: carbohydrates, fats, and proteins. While these forms are all different, the calories of the different forms can be measured in the same units. Thus, several factors measured in common units may be plausible when an attribute has several different sources or several different forms of expression.

8. Conclusions

As in our example, the simple structure model in equal units may also be a simple structure model with equal variances. However, it cannot simply be assumed that the equal variance and equal units models are the same. For a given set of data, the simple structure model with all variances equal to one and all dimensions constrained to correlate equally provides a way to test the assumption before basing any analysis upon that assumption.

The most common way to standardize measures is to calibrate them to have a common mean and variance. The practice of standardizing latent variables to have variance 1.0 may be one of the most unquestioned practices in our field. That practice, however, may limit research questions that can be meaningfully addressed. We propose an alternate standardization using the EU special case of the bi-factor model. This approach, applicable where the EU model holds, yields measures with equal units, but not necessarily equal variances. It involves first fitting the EU model and then using the variances from the EU analysis to set the variances in the corresponding simple structure parameterization. These simple structure dimension scores are the final product of the process calibrated in equal units; the units of the general factor from the EU solution. Calibrating the variables in equal units adds validity to the scores by giving meaning to comparisons among the dimensions. Having calibrated the scores in equal units, one can now meaningfully compare the standard deviations of different measures, linear coefficients for different measures, or change over time on different measures. Meaningful units can enhance the use of measurement in psychological practice and research. In the EU model, the general dimension θ_g serves as a linking variable facilitating calibration of specific dimensions θ_s and simple structure dimensions θ_s^* on common scales.

Declarations

Conflict of interest Professor Nidhi Kohli is a co-author and Section Editor; Applications, Reviews, and Case Studies. The manuscript was submitted to the Theory and Methodology section. Seungwon Chung is an employee of the US Food and Drug Administration (FDA) and has no conflict of interest to report. The views expressed in this document are those of the authors and should not be construed to represent official FDA views or policies. The information and analyses included in this document are provided for academic research purposes only and should not be considered FDA recommended approaches. Mark L. Davison and Ernest C. Davenport, Jr. have no conflict of interest to report.

References

American Educational Research Association. (2014). *Standards for educational and psychological measurement*. Author.

Azen, R., & Budescu, D. V. (2006). Comparing predictors in multivariate regression: An extension of dominance analysis. *Journal of Educational and Behavioral Statistics, 31*(2), 157–180. https://doi.org/10.3102/10769986031002157

Booth, T., Murray, A. L., Overduin, M., Matthews, M., & Furnham, A. (2015). Distinguishing CEOs from top level management: A profile analysis of individual differences, career paths and demographics. *Journal of Business and Psychology, 31*, 205–216.

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*, 153–168. https://doi.org/10.1007/BF02294533

Brogden, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114*, 542–551. https://doi.org/10.1037/0033-2909.114.3.542

Brogden, H. E. (1977). The Rasch model, the law of comparative judgment, and additive conjoint measurement. *Psychometrika, 42*(4), 631–634. https://doi.org/10.1007/BF02295985

Campbell, N. R. (1920). *Physics: The elements* (Vol. 1). Cambridge University Press.

Campbell, N. R. (1928). *An account of the principles of measurement and calculation*. Longmans Green.

Davison, M. L., Jew, G., & Davenport, E. C. Jr. (2014). Patterns of SAT scores, choice of STEM major, and gender. *Measurement and Evaluation in Counseling and Development*, *47*, 118–126. https://doi.org/10.1177%2F0748175614522269

Davison, M. L., Davenport, E. C., Jr., Kohli, N., Kang, Y., & Park, K. (2021). Addressing quantitative and qualitative hypotheses using regression models with equality restrictions and predictors measured in common units. *Multivariate Behavioral Research, 56*(1), 86–100. https://doi.org/10.1080/00273171.2020.1754154

Dilchert, S. (2007). Peaks and valleys: Predicting interests in leadership and managerial positions from personality profiles. *International Journal of Selection and Assessment, 15*(2), 317–334.

Domingue, B. (2014). Evaluating the equal-interval hypothesis with test score scales. *Psychometrika, 79*(1), 1–19. https://doi.org/10.1007/S11336-013-9342-4

Erford, B. T. (2012). *Assessment for counselors* (2nd ed.). Brooks/Cole.

Green, K. E. (1986). Fundamental measurement: A review and application of additive conjoint measurement in educational testing. *The Journal of Experimental Education, 54*(2), 141–147. https://doi.org/10.1080/00220973.1986.10806412

Jeon, J., Rijmen, F., & Rabe-Hasketh, S. (2018). CFA models with a general factor and multiple sets of secondary factors. *Psychometrika, 83*, 785–808. https://doi.org/10.1007/s11336-018-9633-x

Johnson, J. W., & Lebreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods, 7*, 238–257. https://doi.org/10.1177/1094428104266510

Karabatsos, G. (2001). The Rasch model additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement, 2*(3), 389–423.

Krantz, D. H., Luce, R., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Additive and polynomial representations*: Vol. 1. Academic Press.

Kyngdon, A. (2011). Plausible measurement analogies to some psychometric models of test performance. *British Journal of Mathematical and Statistical Psychology, 64*(2), 478–497. https://doi.org/10.1348/2044-8317.002004

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new scale type of fundamental measurement. *Journal of Mathematical Psychology, 1*(1), 1–27. https://doi.org/10.1016/0022-2496(64)90015-X

Michell, J. (1990). *A introduction to the logic of psychological measurement*. Psychology Press.

Peralta, Y., Kohli, N., Lock, E. F., & Davison, M. L. (2022). Bayesian modeling of associations in bivariate linear mixed-effects models. *Psychological Methods, 27*(1), 46–64. https://doi.org/10.1037/met0000358

Perline, R., Wright, B. D., & Wainer, H. (1979). The Rasch model as an additive conjoint measurement. *Applied Psychological Measurement, 3*(2), 237–255. https://doi.org/10.1177/014662167900300213

Rijmen, F. (2010). Formal relations and an empirical comparison between the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement, 47*, 361–372. https://doi.org/10.1111/j.1745-3984.2010.00118.x

Rijmen, F., Jeon, M., von Davier, M., & Rabe-Hesketh, S. (2014). A third-order item response theory model for modeling the effects of domains and subdomains in large-scale educational assessment surveys. *Journal of Educational and Behavioral Statistics, 39*, 235–256. https://doi.org/10.3102/1076998614

Scientific Software International. *IRTPRO Guide*. Author, (2011). http://www.ssicentral.com/index.php/products/irt/irtpro-downloads

Shen, W. (2011). *The application of a person-oriented criterion-related configural approach to the relationship between personality traits and work behaviors* [Doctoral dissertation, University of Minnesota]. https://hdl.handle.net/11299/113566

Shin, T., Davison, M. L., Long, J. D., Chan, C. K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences, 23*(1), 92–100. https://doi.org/10.1016/j.lindif.2012.10.002

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393–408. https://doi.org/10.1007/BF02294363

Thissen, D. (2013). Using the testlet response model as a shortcut to multidimensional IRT subscore computation. In R. E. Millsap et al. (eds.), *New developments in quantitative psychology*. Springer https://doi.org/10.1007/978-1-4614-9348-8_3

Wiernik, B. M., Wilmot, M. P., Davison, M. L., & Ones, D. S. (2021). Meta-analytic criterion profile analysis. *Psychological Methods, 26*(2), 186–209. https://doi.org/10.1037/met0000305

Yung, Y.-F., Thissen, D., & McLeod, L. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*, 113–128. https://doi.org/10.1007/BF02294531

Manuscript Received: 20 DEC 2023
Published Online Date: 8 FEB 2024