CFA MODELS WITH A GENERAL FACTOR AND MULTIPLE SETS OF SECONDARY FACTORS

MINJEONG JEON

UNIVERSITY OF CALIFORNIA, LOS ANGELES

FRANK RIJMEN

AMERICAN INSTITUTES FOR RESEARCH

SOPHIA RABE-HESKETH

UNIVERSITY OF CALIFORNIA, BERKELEY

We propose a class of confirmatory factor analysis models that include multiple sets of secondary or specific factors and a general factor. The general factor accounts for the common variance among manifest variables, whereas multiple sets of secondary factors account for the remaining source-specific dependency among subsets of manifest variables. A special case of the model is further proposed which constrains the specific factor loadings to be proportional to the general factor loadings. This *proportional model* substantially reduces the number of model parameters while preserving the essential structure of the general model. Furthermore, the proportional model allows for the interpretation of latent variables as the expected values of the observed manifest variables, decomposition of the variances, and the inclusion of interactions, similar to generalizability theory. We provide two applications to illustrate the utility of the proposed class of models.

Key words: multifacet models, bifactor models, trifactor models, proportional models, secondary variance sources, general factor, variance decomposition, generalizability theory.

1. Introduction

Confirmatory factor analysis (CFA) is a commonly used tool for the multivariate analysis of psychological data. In single-factor CFA models, the i-th manifest variable for person p is a linear function of a latent variable (θ_p) and a residual term (ϵ_{ip}): $y_{ip} = \beta_i + \alpha_i \theta_p + \epsilon_{ip}$, where β_i is the intercept and α_i is the factor loading of manifest variable y_{ip} on the latent trait variable (or common factor) θ_p. Typically, the residuals (or unique factors) ϵ_{ip} are assumed to be uncorrelated with one another, implying that the manifest measures y_{ip} covary due to the target latent variable θ_p. Violation of conditional independence can corrupt the nature of the target latent variable, distort the true relations between the target variable and other constructs, and lead researchers to false conclusions (Cole et al., 2007; Wainer, 1995; Wainer & Thissen, 1996).

A bifactor model is a useful option if there is an identified source of residual covariation. Suppose a social anxiety scale takes into account three different situation types (e.g., intellectual, physical, separation) in which people can exhibit social anxiety. In this case, extraneous covariation is likely to be created among the manifest variables that share the same situation types. A bifactor model can deal with the extra source of covariation that is not explained by the target factor by

introducing a set of secondary or specific factors that correspond to intellectual, physical, and separation situations, respectively. A manifest variable becomes a function of a target factor, an additional factor for a specific situation that the manifest variable falls into, and the residual term: $y_{ip} = \beta_i + \alpha_i^G \theta_p^G + \alpha_{is}^S \theta_{sp}^S + \epsilon_{ip}$, where α_i^G and α_{is}^S are the loadings for the target factor θ_p^G and for the s-th specific factor θ_{sp}^S ($s = 1, \ldots, T$). The extracted target factor is purified by purging out the influence from the secondary variation source (S) of the situation type.

Suppose the social anxiety scale further considers three behavioral reaction types (e.g., cognitive, physiological, affective) that differ across situations. A bifactor model is no longer appropriate in this case because the extra covariation generated by the shared reaction types is not captured by the model, thus causing a violation of conditional independence. A more elaborate model is needed to take into account both reaction types and the situation types.

In this article, we present a CFA modeling approach that can be useful in such a scenario and can deal with multiple secondary sources of residual covariation. The proposed model is an extension of a standard bifactor model that incorporates more than one set of specific factors in addition to a target factor. One may consider the use of correlated residuals instead of introducing another set of specific factors. However, specification of a factor structure is preferred to the correlated residual approach for its stronger theoretical basis and parsimony (see also, Cole et al., 2007). Analogous to a bifactor model, all factors are assumed to be uncorrelated with each other; that is, each factor captures the variance that is unique to each specified source. Importantly, we propose a special case of the proposed model that constrains the factor loadings of the specific factors to be proportional to the factor loadings of the general factor. This proportional model greatly reduces the number of parameters to be estimated while preserving the essential structure of the general model. Further, this model permits the interpretation of all latent variables as main effects and interactions, decomposition of the variances, and inclusion of interactions, similar to generalizability theory.

The rest of this article is structured as follows: In Sect. 2, we begin with a discussion of several existing methods for dealing with secondary sources of variance. In Sect. 3, we describe the proposed class of models in its general and constrained forms. We also discuss the proposed models' variance decomposition, reliability, and identification issues. In Sect. 4, we provide two empirical examples to illustrate the utility of the proposed models. In Sect. 5, we present a simulation study to demonstrate parameter recovery and scalability of our approach with an increasing number of secondary sources of variance. We end in Sect. 5 with a discussion on findings, limitations, and avenues for future research.

2. Existing Approaches to Secondary Sources of Variance

We begin with a discussion on how secondary variance is currently being handled in the literature. In CFA modeling, *method factors* are typically utilized in multitrait-multimethod (MTMM) designs (Campbell & Fiske, 1959) to take into account an additional source of variance that is attributable to the assessment method (e.g., raters, informants, test-forms). However, the use of method factors tends to produce various practical problems (e.g., model non-convergence and improper parameter estimates), especially when the method factors are allowed to be correlated with each other (e.g., Marsh, 1989). Eid (2000) proposed a correlated trait-correlated method-1 (CTC(M-1)) model that assumes a reference method factor against which the remaining methods are contrasted. CTC(M-1) models are known to show little convergence issues and have conceptually well-defined latent variables. However, CTC(M-1) models have been criticized for having trait factors confounded with a reference method factor and that the fit of the models may not be invariant when a different method is chosen as the reference factor (e.g., Bauer et al., 2013; Pohl & Steyer, 2010). Alternatively, utilizing correlated residuals (e.g., Kenny, 1976) has been suggested

for accommodating method variance in common factor modeling. Although fewer practical problems were reported, the correlated residual approach has been criticized for its weak theoretical foundation and inefficiency when applied to a large number of measurements (Cole et al., 2007).

Our proposed CFA modeling approach is designed for situations where one wishes to evaluate and measure a single construct when multiple sources of secondary variation are present. Note that this scenario does not parallel the typical MTMM design in which usually multiple traits (rather than a single trait) are involved and one secondary variance source (usually methods) is taken into account.

When assessing a single target construct of primary interest in the presence of secondary variance sources, one may consider specifying a higher-order model (Yung et al., 1999). Higher-order models posit that the associations between a set of factors at one level are fully explained by a higher-order factor at the next, higher level. The idea has a long history. For instance, Schmid and Leiman (1957, Table 9, p. 59) presented a hierarchical factor solution that includes one general factor and two sets of nested secondary factors. Such a second-order factor model was also discussed by Jöreskog (1970) and extended to a higher-order factor model by Bentler (1976). Recently, Rijmen et al. (2014) proposed a third-order model in an IRT context with two sets of secondary factors where one set is nested with another set (which is also nested within the third-order factor). Higher-order models assume a fully nested (or hierarchical) structure among the two sets of factors. Such a factor structure is, however, inappropriate when a secondary variance source is not nested but crossed with other secondary sources (e.g., tests and raters).

A model that includes both a general factor and two sets of secondary factors is a trifactor model (Bauer et al., 2013) which was proposed as an extension of a bifactor model in the context of analyzing ratings on a set of items that were assessed by multiple informants. The key idea of the trifactor model is to incorporate in a common factor model two sets of additional latent variables that are expected to capture dependencies within informants as well as within items. Specifically, an observed item response (informant rating) loads on a *perspective (informant) factor* and an *item factor* in addition to a *common factor* that represents the consensus view across informants and across items. All latent variables are assumed to be independent of each other. Some restrictions on the model parameters were discussed which may be needed when interchangeable, rather than structurally different, informants are utilized. Further, a conditional trifactor model was described that includes predictors of different factors. It was indicated that a trifactor model can be formulated as a two-tier model (Cai, 2010) by treating the common factor and the perspective factors as primary factors (the first tier) and the item factors as a set of secondary factors (the second tier).

A trifactor-like structure is not new in the CFA/SEM literature. Mellenbergh et al. (1979) discussed a CFA model with a trifactor-like structure where instruments are constructed based on the Cartesian product of two facets. Recently, Raykov and Marcoulides (2006) presented a similar SEM model for a two-facet crossed design (that involves four raters and two occasions) in the context of illustrating the estimation of the relative generalizability coefficients with a SEM approach.

Our proposed model, in its general form, can be viewed as an extension of trifactor-like models presented in the literature. We go one step further and generalize previous formulations and applications to multiple secondary variance source cases. We additionally provide a simplified version of the general model with proportionality constraints and discuss the possibility of including interactions among secondary sources of variance. Moreover, we examine several theoretical aspects of our proposed models, such as variance decomposition, reliability, and identification.

Generalizability theory (G-theory) (Brennan, 2001; Cronbach et al., 1972) has been utilized to identify and evaluate multiple sources of secondary or "error" variance, called facets, as an extension to classical test theory. Applying G-theory ideas to a MTMM design, Woehr et al. (2012) demonstrated how the total observed variance can be decomposed according to the MTMM

design and how the contribution of each variance component (e.g., trait and method factors) can be evaluated based on the variance decomposition.

We will show how our model with proportionality constraints has clear connections with a G-theory approach and allows for a clearer interpretation of the latent variables as main effects and interactions.

3. Models with a General Factor and Multiple Sets of Secondary Factors

In this section, we explicate a general formulation of the proposed model and present a special case of the model with proportionality constraints. The models will be formulated in a scalar form for conciseness. A matrix formulation is provided in the supplementary material (Appendix A). Variance decomposition, reliability coefficients, and identification of the proposed models will be discussed subsequently.

3.1. Some Notation and Definitions

Let y_{ip} denote the ith continuous/normal response variable ($i = 1, \ldots, I$) for person p ($p = 1, \ldots, N$). We also call y_{ip} a manifest variable, observed variable, indicator variable, or measurement throughout the article. We suppose that intercorrelations among the manifest variables are attributable to a target source (G) as well as K secondary sources ($S_1, S_2, \ldots, S_K$).

The target source produces the common variance that runs through all manifest variables and is represented by a *general factor*. The secondary sources produce covariances among subsets of the manifest variables and are represented by secondary or *specific factors*. The secondary sources typically represent aspects or facets of measurement, such as time of day, and the specific factors represent the individual conditions, such as morning, midday, and evening. We let N_{S_k} denote the number of specific factors from the k-th secondary source S_k.

3.2. General Factor Structure

To have a better understanding of the structure of the proposed model, we first inspect its factor loading patterns. To illustrate, suppose there are two secondary variance sources (S_1 and S_2) and two observations ($n = 2$) per combination of secondary factors from different sources. An example would be two repeated experiments ($n = 2$) for each combination of two situations ($N_{S_1} = 2$) and two raters ($N_{S_2} = 2$). The corresponding factor loading matrix for the eight manifest variables ($I = 8$) can be specified as follows:

$$
\begin{bmatrix}
\alpha_1^G & \alpha_{11}^{S_1} & 0 & \alpha_{11}^{S_2} & 0 \\
\alpha_2^G & \alpha_{21}^{S_1} & 0 & \alpha_{21}^{S_2} & 0 \\
\alpha_3^G & 0 & \alpha_{32}^{S_1} & \alpha_{31}^{S_2} & 0 \\
\alpha_4^G & 0 & \alpha_{42}^{S_1} & \alpha_{41}^{S_2} & 0 \\
\alpha_5^G & \alpha_{51}^{S_1} & 0 & 0 & \alpha_{52}^{S_2} \\
\alpha_6^G & \alpha_{61}^{S_1} & 0 & 0 & \alpha_{62}^{S_2} \\
\alpha_7^G & 0 & \alpha_{72}^{S_1} & 0 & \alpha_{72}^{S_2} \\
\alpha_8^G & 0 & \alpha_{82}^{S_1} & 0 & \alpha_{82}^{S_2}
\end{bmatrix} .
\tag{1}
$$

The rows represent individual manifest variables and the columns represent a general factor for target variance source G (Column 1), two specific factors for secondary source S_1 (Columns 2–3),

and two specific factors for another secondary source S_2 (Columns 4-5). Here α_i^G is the loading of the i-th measurement on the general factor and $\alpha_{is_1}^{S_1}$ ($s_1 = 1, 2$) and $\alpha_{is_2}^{S_2}$ ($s_2 = 1, 2$) are the loadings on the s_1-th specific factor from S_1 and on the s_2-th specific factor from S_2, respectively. Note that cross-loadings within a source are not permitted, meaning that a manifest variable is assumed to be associated with only one specific factor within the source. This is appropriate when the conditions corresponding to the specific factors are mutually exclusive, as they typically are, for example when the source is time of day.

Notice that the two sets of specific factors are fully crossed in the factor loading matrix. Partial cross-classification is feasible in principle, but we will focus on fully crossed cases without loss of generality. We will also not discuss purely nested structures in which case a third-order model (e.g., Rijmen et al., 2014) would be a better modeling choice as discussed in Sect. 2.

For a fully crossed design, the total number of measurements can be determined by the number of secondary factors from each secondary source (S_1 and S_2) and the number of observations (n) per combination of secondary factors from different sources. For the above example, the total number of observations is $I = 8 = N_{S_1} \times N_{S_2} \times n$ ($= 2 \times 2 \times 2$).

Through inspecting the general factor loading matrix, it is clear that the proposed model extends the bifactor structure by including more than one set of specific factors. The factor pattern matrix (1) can indeed be reduced into a bifactor structure by removing, for example, Columns 4–5 (factors from the secondary source S_2).

3.3. General Model Formulation and Assumptions

With K secondary variance sources, a general model can be formulated as follows:

$$y_{ip} = \beta_i + \alpha_i^G \theta_p^G + \sum_{k=1}^{K} \alpha_{is_k}^{S_k} \theta_{s_k p}^{S_k} + \epsilon_{ip}, \tag{2}$$

where β_i is the intercept, θ_p^G is the general factor, and $\theta_{s_k p}^{S_k}$ is the s_k-th specific factor for the k-th secondary source (S_k, $k = 1, \ldots, K$). The parameters α_i^G and $\alpha_{is_k}^{S_k}$ are the factor loadings of the i-th manifest variable for the G factor and the s_k-th secondary factor from Source S_k, respectively. The assignment of measurement i to a specific factor from each source could be made explicit by using subscripts $s_k[i]$, but we leave this implicit for notational simplicity. Finally, ϵ_{ip} is the residual with variance ψ_{ii}. Normality is commonly assumed for all factors and the residuals. Note that when $K = 1$, the general model is reduced to a bifactor model. When $K = 2$, the general model is equivalent to a trifactor or two-facet CFA model. How the trifactor model is different from the bifactor model can more clearly be seen in its matrix form (loading matrix) that is presented in Appendix A of the supplementary material.

The following additional assumptions are required to complete the formulation of the general model: (1) the means of the general and specific factors are zero, (2) the means of the residuals are zero, (3) the variances of the general and specific factors are fixed to 1, (4) all factors are uncorrelated with each other and across different units, (5) the residuals are uncorrelated across different units, and (6) the residuals are uncorrelated with the general factor and the two sets of specific factors.

Assumptions 1 and 2 are imposed so that β_i can be freely estimated. Assumption 3 is imposed to set the scale of the model; all nonzero factor loadings can then be freely estimated and compared in terms of relative magnitude. For instance, a large general factor loading α_i^G indicates that the i-th manifest variable largely reflects the target factor rather than other secondary factors. On the other hand, a large specific factor loading $\alpha_{is_k}^{S_k}$ suggests that variability of the manifest variable

may be driven by the s_k-th specific factor from source S_k. This way, inspection of the relative magnitude of the factor loadings can aid in scale evaluation. Assumptions 4, 5 and 6 are motivated by the bifactor model and correspond to the assumption that all covariation among the manifest variables is due to the additive effects of independent factors.

A simple form of the general model (2) may be obtained by imposing the following constraints: (a) Intercepts that depend only on the conditions of the secondary variance sources, (b) Equal common factor loadings: $\alpha_i^G = C^G$, (c) Equal secondary factor loadings: $\alpha_{is_k}^{S_k} = C^{S_k}$, and (d) Equal residual variances: $\psi_{ii} = \psi$. With these constraints, the general model becomes equivalent to a mixed-effects ANOVA model with fixed main effects and interactions between secondary variance sources (absorbed by the intercepts), random effects $C^G\theta_p$ for subjects, and random interactions $C^{S_k}\theta_{s_kp}^{S_k}$ between subjects and secondary variance sources. Similarly, other kinds of constrains may be imposed to accommodate various specific measurement situations in practice.

3.4. A Model with Proportionality Constraints

Here we propose a set of constraints for the proposed model such that the loadings of secondary factors are proportional to the loadings of the general factor. To illustrate, when $K = 2$ the proposed model with this specific type of constraints can be formulated as follows:

$$y_{ip} = \beta_i + \alpha_i^G \left(\theta_p^G + C_{s_1}^{S_1}\theta_{s_1p}^{S_1} + C_{s_2}^{S_2}\theta_{s_2p}^{S_2} \right) + \epsilon_{ip}. \tag{3}$$

Here α_i^G is the common factor loading shared by the general factor and the two sets of specific factors, and $C_{s_1}^{S_1}$ and $C_{s_2}^{S_2}$ are proportionality constants for the s_1-th specific factor $\theta_{s_1p}^{S_1}$ from source S_1 and the s_2-th specific factor $\theta_{s_2p}^{S_2}$ from source S_2, respectively. Note that this formulation allows the manifest variables that share the same secondary variance condition for a given variance source to have equal factor loadings for the corresponding secondary factor; in other words, essentially tau-equivalent measurement is assumed for a secondary factor (e.g., Raykov, 1997).

The proportionality constants in Eq. (3) reflect the magnitude of impacts that secondary factors have on the manifest variables relative to the general factor. For instance, suppose $C_{s_1}^{S_1} = 0.5$ and $C_{s_2}^{S_2} = 0.3$; this means that the s_1-th S_1 specific factor has a 50% effect on the i-th manifest variable while the s_2-th S_2 specific factor has a 30% effect on the manifest variable relative to the effects of the general factor. We refer to this constrained model as a *proportional model* to highlight its key parameters, the proportionality constants.

For the proportional model the factor loadings for the specific factors can be obtained by multiplying the proportionality constants ($C_{s_1}^{S_1}$ and $C_{s_2}^{S_2}$) by the general factor loading α_i^G so that the factor loading matrix is:

$$\begin{bmatrix} \alpha_1^G & \alpha_1^G C_1^{S_1} & 0 & \alpha_1^G C_1^{S_2} & 0 \\ \alpha_2^G & \alpha_2^G C_1^{S_1} & 0 & \alpha_2^G C_1^{S_2} & 0 \\ \alpha_3^G & 0 & \alpha_3^G C_2^{S_1} & \alpha_3^G C_1^{S_2} & 0 \\ \alpha_4^G & 0 & \alpha_4^G C_2^{S_1} & \alpha_4^G C_1^{S_2} & 0 \\ \alpha_5^G & \alpha_5^G C_1^{S_1} & 0 & 0 & \alpha_5^G C_2^{S_2} \\ \alpha_6^G & \alpha_6^G C_1^{S_1} & 0 & 0 & \alpha_6^G C_2^{S_2} \\ \alpha_7^G & 0 & \alpha_7^G C_2^{S_1} & 0 & \alpha_7^G C_2^{S_2} \\ \alpha_8^G & 0 & \alpha_8^G C_2^{S_1} & 0 & \alpha_8^G C_2^{S_2} \end{bmatrix}. \tag{4}$$

An appealing, practical feature of a proportional model is that the number of parameters to be estimated is substantially reduced compared to the corresponding general model. For instance, with $K = 2$ and $n = 1$, the general model requires $3I$ factor loading parameters, while the proportional model has only $I + N_{S_1} + N_{S_2} < 3I$ factor loadings and proportionality constants. Hence, a proportional model is particularly beneficial when applied to a large data problem, e.g., with $n \gg 1$ or $K \gg 2$.

Further, the proposed model is of theoretical importance. To discuss this point, let us reparameterize (3) as follows:

$$y_{ip} = \beta_i + \alpha_i^G \left(\theta_p^G + \theta_{s_1p}^{S_1}{}^* + \theta_{s_2p}^{S_2}{}^* \right) + \epsilon_{ip}. \tag{5}$$

Equation (5) includes new notation for the unstandardized secondary factors $\theta_{s_1p}^{S_1}{}^* = C_{s_1}^{S_1} \theta_{s_1p}^{S_1}$ and $\theta_{s_2p}^{S_2}{}^* = C_{s_2}^{S_2} \theta_{s_2p}^{S_2}$, which have variances $\mathrm{Var}(\theta_{s_1p}^{S_1}{}^*) = \{C_{s_1}^{S_1}\}^2$ and $\mathrm{Var}(\theta_{s_2p}^{S_2}{}^*) = \{C_{s_2}^{S_2}\}^2$.

Note that the proportionality constants $C_{s_1}^{S_1}$ and $C_{s_2}^{S_2}$ are now scaling constants since they are the standard deviations of the specific factors $\theta_{s_1p}^{S_1}{}^*$ and $\theta_{s_2p}^{S_2}{}^*$. The scaling constants $C_{s_1}^{S_1}$ and $C_{s_2}^{S_2}$ determine the spreads of the associated latent variable distributions, reflecting the magnitude of individual differences due to the s_1-th S_1 factor and the s_2-th S_2 factors relative to the general factor (whose standard deviation is fixed at 1).

Importantly, Eq. (5) shows that the latent variable for person p in conditions s_1 of source S_1 and s_2 of source S_2 is additively decomposed into several sub-factors:

$$\theta_{s_1s_2p}^* = \theta_p^G + \theta_{s_1p}^{S_1}{}^* + \theta_{s_2p}^{S_2}{}^*. \tag{6}$$

We can think of $\theta_{s_1s_2p}^*$ as the true score for a person's trait in that combination of conditions. The measurement model has an item-specific intercept β_i, factor loading α_i^G, and unique factor ϵ_{ip} with item-specific variance. The true score is modeled as a variance-components model with random effects θ_p^G for person p, $\theta_{s_1p}^{S_1}{}^*$ for the combination of condition s_1 and person p, and $\theta_{s_2p}^{S_2}{}^*$ for the combination of condition s_2 and person p.

An important ramification of this latent variable decomposition is that we can now link the proportional model to G-theory, allowing for concrete interpretation of the latent variables of the model. Specifically, θ_p^G can now be interpreted as the "universe score", the expected value of a person's true score across the populations of conditions of the variance sources, or "facets", S_1 and S_2, $\theta_p^G + \theta_{s_1p}^{S_1}{}^*$ as the expected value of a person's score across the population of facet S_2, and $\theta_p^G + \theta_{s_2p}^{S_2}{}^*$ as the expected value of a person's score across the population of facet S_1. The common factor loadings (α_i^G) in Eq. (5) convert these scores to the units with which each manifest variable measures the trait, the unique factors add error, and the intercepts β_i capture all effects of the facets and possibly items that are constant across persons.

In addition, we can also interpret the three types of latent variables as *effects* on the manifest variable as in G-theory. Specifically, θ_p^G can be interpreted as the person main effect, $\theta_{s_1p}^{S_1}{}^*$ as the person by S_1 interaction effect, and $\theta_{s_2p}^{S_2}{}^*$ as the person by S_2 interaction effect. The main effects for S_1 and S_2 are absorbed in β_i, are not random, and are constant across people and therefore do not contribute to the observed variance across people (for a similar argument, see, e.g., Woehr et al., 2012). The three-way interaction effect among S_1, S_2, and persons is contained in the residual term ϵ_{ip}, unless there are multiple manifest variables ($n > 1$) for each combination of S_1 and S_2 sources, which will be discussed in Sect. 3.5. The proportional model becomes a mixed-effects ANOVA model if the factor loadings are set to one, the unique variances are constrained to be

constant across items, and the intercepts are structured in terms of main effects and interactions of the variance sources.

The proportional model can be utilized for its practical and/or theoretical merits that have been discussed above. The use of the proportional model can also be recommended for substantive reasons. Suppose observed variables that share the same conditions of a secondary source are indeed exchangeable (e.g., ratings obtained from particular informants). In this case, investigating specific differences in the factor loadings for the conditions may be of little importance and thus, it may suffice to evaluate the overall impacts of the conditions (i.e., proportionality constants) for the secondary variance sources.

Bauer et al. (2013) considered a different type of parameter constraint in the trifactor model when there are interchangeable informants. Specifically, when informants of one type, such as teachers rating children, are selected randomly from a rater pool so that it is not meaningful to model systematic differences between informants, Bauer et al. suggested to impose equality constraints on all model parameters between the informants. A similar set of constraints can also be applied in the more general context that we consider here; that is, when conditions of a secondary source are interchangeable, equality constraints can be imposed on model parameters across the interchangeable conditions. For the proportional model, for example, we can impose $C_1^{S_1} = C_2^{S_1} \cdots = C_{N_{S_1}}^{S_1}$ if the conditions of Source S_1 are regarded as interchangeable.

The idea of proportionality constraints has been adopted by Bradlow et al. (1999) and Wainer et al. (2007) in the context of capturing conditional dependence within item clusters (or testlets) in item response theory (IRT). Specifically, testlet effects were incorporated in regular IRT models as additional continuous random effects (or latent variables). Model (5) may be seen as a continuous response version of a testlet IRT model that includes two sets of specific factors ($\theta_{s_1 p}^{S_1}{}^*$ and $\theta_{s_2 p}^{S_2}{}^*$). A testlet IRT model is formally equivalent to a second-order IRT model in which the general (second-order) factor has direct effects on the specific (first-order) factors (e.g., Rijmen, 2010; Yung et al., 1999). One may wonder then whether the proportional model can also be parameterized as a second-order factor version of the model in which the target factor has direct effects on the two sets of secondary factors rather than on the manifest variables. Regarding this point, we found that (1) the proportional model is not equivalent to the second-order factor version; (2) the second-order factor model is more complex in terms of parameterization when compared to the proportional model; and (3) the proportional model is not nested within the second-order version, meaning that the proportional model cannot be attained by imposing a set of constraints on the second-order version. A mathematical proof of these points is provided in the supplementary material (Appendix B).

3.5. Interaction Effects

The proportional model (5) is formulated based on an additive decomposition of the latent variables, $\theta_{s_1 s_2 p} = \theta_p^G + \theta_{s_1 p}^{S_1}{}^* + \theta_{s_2 p}^{S_2}{}^*$, implicitly assuming that there is no three-way interaction between the two secondary variance sources and persons. In this case, for a given person, the specific factors represent the effects of conditions from the corresponding source and these effects are assumed to be constant across conditions from other sources.

When multiple measurements ($n > 1$) are present per combination of conditions from different sources, model (5) can incorporate interaction effects represented by *additive* latent variables. The use of additive interaction factors (or 'random interactions') is common in a multi-way mixed-effects analysis of variance (ANOVA).

Suppose there are two secondary variance sources ($K = 2$). An additional interaction factor can be included in model (5) as follows:

$$y_{ip} = \beta_i + \alpha_i^G \left(\theta_p^G + \theta_{s_1 p}^{S_1}{}^* + \theta_{s_2 p}^{S_2}{}^* + \theta_{s_{12} p}^{S_{12}}{}^* \right) + \epsilon_{ip}, \tag{7}$$

where $\theta_{s_{12} p}^{S_{12}}{}^*$ represents the interaction factor between person and the s_1-th and s_2th conditions of sources S_1 and S_2 and has variance $\mathrm{Var}(\theta_{s_{12} p}^{S_{12}}{}^*) = \{C_{s_{12}}^{S_{12}}\}^2$. The newly added interaction factors are assumed to be independent across units and independent of each other, of other factors, and of the residuals. For a given person p, the interaction term can be interpreted as the deviation from the mean effect (across conditions from source S_2) of the s_1-th condition from Source S_1 when combined with s_2-th condition from Source S_2.

The general model can similarly be extended when $n > 1$:

$$y_{ip} = \beta_i + \alpha_i^G \theta_p^G + \alpha_{is_1}^{S_1} \theta_{s_1 p}^{S_1} + \alpha_{is_2}^{S_2} \theta_{s_2 p}^{S_2} + \alpha_{is_{12}}^{S_{12}} \theta_{s_{12} p}^{S_{12}} + \epsilon_{ip}, \tag{8}$$

where $\alpha_{is_{12}}^{S_{12}}$ is the factor loading of the i-th manifest variable on the interaction factor $\theta_{s_{12} p}^{S_{12}}$. Clearly, the general model becomes increasingly complicated due to the extra factor loading parameters introduced for the interaction factors. This complication makes the more parsimonious proportional model more attractive for specifying interaction effects. For example, with $K = 2$, the general model with interaction effects requires $4I$ factor loading parameters to be estimated, whereas the proportional model requires $I + N_{S_1} + N_{S_2} + N_{S_1} \cdot N_{S_2}$ loadings and proportionality constants, which is less than $4I$ since N_{S_1}, N_{S_2}, and $N_{S_1} \cdot N_{S_2}$ are each less than I. In this case, $I/(N_{S_1} * N_{S_2}) = n > 1$ is necessary for the model to be identifiable.

For the sake of simplicity, we will from now on limit our discussions and illustrations to situations where there are two secondary variance sources ($K = 2$) and $n = 1$ manifest variable per combination of conditions from different sources.

3.6. Variance Decomposition

The proposed models assume that all latent variables are uncorrelated across units and with each other (Assumption 4) and the residuals are uncorrelated across units and with all latent variables (Assumptions 5 and 6). On the basis of these independence assumptions, the total variance of the manifest variables can be conveniently decomposed into additive contributions from the common variance source (represented by θ_p^G), secondary variance sources (represented by $\theta_{s_1 p}^{S_1}$ and $\theta_{s_2 p}^{S_2}$), and unexplained variance source (represented by ϵ_{ip}).

For a general model with $K = 2$, the total variance of y_{ip} can be decomposed as follows:

$$\mathrm{Var}(y_{ip}) = \left\{ \alpha_i^G \right\}^2 \cdot \mathrm{Var}\left(\theta_p^G \right) + \left\{ \alpha_{is_1}^{S_1} \right\}^2 \cdot \mathrm{Var}\left(\theta_{s_1 p}^{S_1} \right) + \left\{ \alpha_{is_2}^{S_2} \right\}^2 \cdot \mathrm{Var}\left(\theta_{s_2 p}^{S_2} \right) + \mathrm{Var}(\epsilon_{ip}). \tag{9}$$

Here $\mathrm{Var}(\theta_p^G) = \mathrm{Var}(\theta_{s_1 p}^{S_1}) = \mathrm{Var}(\theta_{s_2 p}^{S_2}) = 1$ and $\mathrm{Var}(\epsilon_{ip}) = \psi_{ii}$. The variance components, $\{\alpha_i^G\}^2$, $\{\alpha_{is_1}^{S_1}\}^2$, and $\{\alpha_{is_2}^{S_2}\}^2$, represent how much a manifest variable varies due to the target factor (G) and the two secondary factors (S_1, S_2), respectively. When $K = 1$, Eq. (9) includes only one secondary source variance, which is the case for the bifactor model.

For a proportional model, the variance decomposition can be obtained conveniently, based on parameterization (5):

$$\mathrm{Var}(y_{ip}) = \left\{ \alpha_i^G \right\}^2 \cdot \underbrace{\left\{ \mathrm{Var}\left(\theta_p^G \right) + \mathrm{Var}\left(\theta_{s_1 p}^{S_1*} \right) + \mathrm{Var}\left(\theta_{s_2 p}^{S_2*} \right) \right\}} + \mathrm{Var}(\epsilon_{ip}). \tag{10}$$

Here $\text{Var}(\theta_p^G) = 1$, $\text{Var}(\theta_{s1p}^{S_1*}) = \{C_{s_1}^{S_1}\}^2$, $\text{Var}(\theta_{s2p}^{S_2*}) = \{C_{s_2}^{S_2}\}^2$, and $\text{Var}(\epsilon_{ip}) = \psi_{ii}$. Note that the variance decomposition shown in the underbraced term of (10) is in line with the latent variable decomposition in (6).

Our model specification is based on viewing the factors as random variables. Some may argue that this treatment is inappropriate when secondary factors are non-interchangeable (or structurally different) and hence should be treated as fixed effects (e.g., Eid et al., 2017, 2008; Steyer et al., 1992). However, as shown above, the secondary factors in our models are the interaction effects between secondary sources (S_1 and S_2) and persons rather than the main effects of secondary sources. Therefore, as long as persons are interchangeable, the interaction terms with persons (hence $\theta_{s_1p}^{S1}{}^*$, $\theta_{s_2p}^{S2}{}^*$) can be considered random.

3.7. Statistical Indices

Building upon the variance decomposition discussed in Sect. 3.6, several statistical indices can be defined. For this purpose, let us first define *systematic variance* as the total observed variance minus the residual variance or the sum of $\{\alpha_i^G\}^2\text{Var}(\theta_p^G)$, $\{\alpha_{is_1}^{S_1}\}^2\text{Var}(\theta_{s1p}^{S_1})$, and $\{\alpha_{is_2}^{S_2}\}^2\text{Var}(\theta_{s2p}^{S_2})$. We can then evaluate the relative contribution of each variance component to the systematic variance of each manifest variable. First, based on the general model, we can define γ_i^G for the general factor (common variance source G) as follows:

$$\gamma_i^G = \frac{\{\alpha_i^G\}^2}{\{\alpha_i^G\}^2 + \left\{\alpha_{is_1}^{S_1}\right\}^2 + \left\{\alpha_{is_2}^{S_2}\right\}^2}. \tag{11}$$

The coefficient γ_i^G indicates the degree to which the common variance source (i.e., the target factor) contributes to the systematic variance of the observed variable y_i. For the two sets of specific factors (secondary variance sources S_1 and S_2), we can define

$$\gamma_i^{S_1} = \frac{\left\{\alpha_{is_1}^{S_1}\right\}^2}{\{\alpha_i^G\}^2 + \left\{\alpha_{is_1}^{S_1}\right\}^2 + \left\{\alpha_{is_2}^{S_2}\right\}^2}, \tag{12}$$

$$\gamma_i^{S_2} = \frac{\left\{\alpha_{is_2}^{S_2}\right\}^2}{\{\alpha_i^G\}^2 + \left\{\alpha_{is_1}^{S_1}\right\}^2 + \left\{\alpha_{is_2}^{S_2}\right\}^2}. \tag{13}$$

The coefficients $\gamma_i^{S_1}$ and $\gamma_i^{S_2}$ represent the degree to which the two secondary variance sources contribute to the systematic variance of the observed variable y_i. These two coefficients enable researchers to identify which manifest variables are more vulnerable to (or greatly influenced by) the secondary variance sources that may be nuisances. Based on this information, researchers may advise test developers to consider removing those variables in future revisions of the scale, or to eliminate a variance source by choosing one condition of the source and holding it constant. Alternatively, researchers may suggest changing the number of variables for a certain source of variance. The decomposition obtained from the estimated model can guide researchers in identifying where increases in the number of observations could yield the greatest gains in reliability to achieve a desired measurement precision.

The proportion of the systematic variance of the observed variable y_i to the total variance (also called *communality*) can be defined as a reliability when the secondary sources are not nuisances

but part of the target of the measurement. The coefficient can be viewed as the 'reliability' of a manifest variable as it indicates the degree to which individual differences in a manifest variable are attributable to the 'reliable' sources of variance induced by the model. A reliability coefficient ρ_i can be computed as follows:

$$\rho_i = \frac{\left\{\alpha_i^G\right\}^2 + \left\{\alpha_{is_1}^{S_1}\right\}^2 + \left\{\alpha_{is_2}^{S_2}\right\}^2}{\left\{\alpha_i^G\right\}^2 + \left\{\alpha_{is_1}^{S_1}\right\}^2 + \left\{\alpha_{is_2}^{S_2}\right\}^2 + \psi_{ii}}. \tag{14}$$

This coefficient can be used to assess the internal quality of a model-based measurement. High reliability coefficients across all manifest variables indicate that the important sources of variation have been identified by the measurement design. One can also compute variants of the reliability coefficients by removing secondary factor variances ($\{\alpha_{is_1}^{S_1}\}^2$ or $\{\alpha_{is_2}^{S_2}\}^2$) from the numerator in (14). This is meaningful if one of the secondary sources is viewed as a nuisance and not considered to be part of the target of the measurement.

For the proportional model, based on variance decomposition (10) the above four coefficients $\gamma_i^G, \gamma_i^{S_1}, \gamma_i^{S_2}$, and ρ_i can be simplified as follows: $\gamma_i^G = \frac{1}{1+\{C_{s_1}^{S_1}\}^2+\{C_{s_2}^{S_2}\}^2}$, $\gamma_i^{S_1} = \frac{\{C_{s_1}^{S_1}\}^2}{1+\{C_{s_1}^{S_1}\}^2+\{C_{s_2}^{S_2}\}^2}$, $\gamma_i^{S_2} = \frac{\{C_{s_2}^{S_2}\}^2}{1+\{C_{s_1}^{S_1}\}^2+\{C_{s_2}^{S_2}\}^2}$, and $\rho_i = \frac{1+\{C_{s_1}^{S_1}\}^2+\{C_{s_2}^{S_2}\}^2}{1+\{C_{s_1}^{S_1}\}^2+\{C_{s_2}^{S_2}\}^2+\psi_{ii}}$. Note that the first three coefficients are invariant across measurements that fall into the same s_1 and s_2 conditions.

The statistical indices discussed above are useful for evaluating the reliability of *individual* manifest variables. Scale-level reliability can be evaluated by extending the indices that have been developed for a bifactor model (Rodriguez et al., 2016). For instance, we can define a variant of the explained common variance (ECV; e.g., Stucky et al., 2013; Reise et al., 2013, 2010) to evaluate the relative strength of the general factor when our proposed models are applied. ECV is computed as the percent of systematic variance explained by the general factor as follows:

$$\text{ECV} = \frac{\sum \left\{\alpha_i^G\right\}^2}{\sum \left\{\alpha_i^G\right\}^2 + \sum \left\{\alpha_{is_1}^{S_1}\right\}^2 + \sum \left\{\alpha_{is_2}^{S_2}\right\}^2}. \tag{15}$$

The ECV index (15) contains in the denominator the sum of the squared factor loadings for multiple sets of secondary factors, which is different from the ECV for the bifactor model.

In addition, we can adapt the coefficient omega hierarchical (omegaH or ω_H; e.g., McDonald, 1999; Reise et al., 2013) for our models. omegaH is defined as the variance in total scores that is due to the general factor divided by the total variance. For the proposed models, omegaH can be computed as follows:

$$\omega_H = \frac{\left\{\sum \alpha_i^G\right\}^2}{\left\{\sum \alpha_i^G\right\}^2 + \left\{\sum \alpha_{is_1}^{S_1}\right\}^2 + \left\{\sum \alpha_{is_2}^{S_2}\right\}^2 + \sum \psi_{ii}}. \tag{16}$$

Hence, the omegaH index can also assess the relative strength of the general factor when the proposed models are applied, similar to the ECV index.

3.8. Identification

Establishing model identification is one of the most important yet neglected steps in CFA modeling (Davis, 1993). Few general rules of identification exist, but they are restricted to simple CFA models (Davis, 1993). In general, a model is said to be *globally* identified if there is no parameter vector $\boldsymbol{\varphi}_b$ for which $\boldsymbol{\vartheta}(\boldsymbol{\varphi}_a) = \boldsymbol{\vartheta}(\boldsymbol{\varphi}_b)$ unless $\boldsymbol{\varphi}_a = \boldsymbol{\varphi}_b$, where $\boldsymbol{\vartheta}$ are the reduced form parameters (means, variances, and non-redundant covariances among the manifest variables). For I manifest variables there are $I + \frac{1}{2}I(I + 1)$ reduced form parameters. Unfortunately, it is generally acknowledged that a proof of such *global identification* of CFA models is nearly impossible except for some special cases (McDonald, 1982). We therefore focus on establishing the *local identification* of the proposed models in the neighborhood of $\boldsymbol{\varphi}_a$.

A CFA model that is theoretically identified can be empirically unidentified for some datasets. Such empirical underidentification can be due to factor variances estimated as zero (so that corresponding loadings cannot be estimated), improper (e.g., out-of-bound) parameter solutions, or model nonconvergence (Kenny, 1979). Therefore, we also evaluate the empirical identification of the model types for which we establish local identification.

To investigate local identification of the proposed models, we first create several model types by manipulating the number of secondary factors, N_{S_1} and N_{S_2}. For each model type, we check whether the necessary condition for local identification is met, namely whether the number of reduced form parameters is equal to or greater than the number of unknown model parameters (q), i.e., $I + \frac{1}{2}I(I + 1) \geq q$. Next, we apply the Wald rank rule (Wald, 1950) for local identification (see also Becker & Cote, 1994; Bollen & Bauldry, 2010; Skrondal & Rabe-Hesketh, 2004) which proceeds as follows: First, form the Jacobian matrix $J(\boldsymbol{\varphi}) = \frac{\partial \boldsymbol{\vartheta}}{\partial \boldsymbol{\varphi}}$, the first derivatives of the reduced form parameters $\boldsymbol{\vartheta}$ with respect to the model parameters $\boldsymbol{\varphi}$. Second, evaluate whether the rank of $J(\boldsymbol{\varphi})$ is equal to the dimension of $\boldsymbol{\varphi}$, which is the necessary and sufficient condition for local identification at $\boldsymbol{\varphi}_a$ if $\boldsymbol{\varphi}_a$ is a regular point of $J(\boldsymbol{\varphi})$. Mathematica (Wolfram Research, Inc, 2010) is used for applying the procedure described above. Example Mathematica code is provided in Appendix C of the supplementary material.

To evaluate the empirical identification of the proposed models, we generate model-implied covariance matrices and vectors of means given several arbitrary sets of parameter values. We then fit the data-generating models to these generated reduced form parameters, specifying an arbitrary sample size, such as $N = 1000$. In practice, the means can be omitted from the reduced form parameters if the β_i are set to 0 because our model imposes no structure on the means. If the data-generating parameter values are accurately recovered, it not only verifies parameter recovery of the proposed models but also corroborates their empirical identification for the sets of parameter values considered.

Table 1 summarizes the results for local and empirical identification of the general and proportional models under the ten conditions that are considered. Additional details on how those conditions are selected for investigation are provided in the supplementary material (Appendix D).

The results show that when the number of S_1 and S_2 factors is increased to 3 and above ($N_{S_1} \geq 3$, $N_{S_2} \geq 3$), the general model is locally and empirically identified. When $N_{S_1} = 2$ and $N_{S_2} = 3$ ($N_{S_1} = 3$ and $N_{S_2} = 2$), however, the general model is not locally identified as confirmed by finding that the model is also empirically underidentified for several different sets of parameter values.

We note that the proportional model is locally and empirically identified whenever the general model is locally and empirically identified ($N_{S_1} \geq 3$, $N_{S_2} \geq 3$). More importantly, the proportional model is identified even when $N_{S_1} = 2$ and $N_{S_2} = 3$ ($N_{S_1} = 3$ and $N_{S_2} = 2$) where the general model is not locally or empirically identified. This is an encouraging finding because

TABLE 1.
Conditions where local and empirical identification are evaluated for the general model and the proportional model.

Condition		N_{S_1}	N_{S_2}	n	$I + \frac{I \times (I+1)}{2} - q$	Local	Empirical
General model[a]							
1	$N_{S_1} = 2$ and $N_{S_2} = 3$ ($N_{S_1} = 3$ and $N_{S_2} = 2$)[b]	2	3	1	0	–	–
2	$N_{S_1} = 3$ and $N_{S_2} = 3$	3	3	1	9	✓	✓
3	$N_{S_1} = 4$ and $N_{S_2} = 3$ ($N_{S_1} = 3$ and $N_{S_2} = 4$)	4	3	1	30	✓	✓
4	$N_{S_1} = 4$ and $N_{S_2} = 4$	4	4	1	72	✓	✓
Proportional model							
1	$N_{S_1} = 2$ and $N_{S_2} = 3$ ($N_{S_1} = 3$ and $N_{S_2} = 2$)	2	3	1	4	✓	✓
2	$N_{S_1} = 3$ and $N_{S_2} = 3$	3	3	1	21	✓	✓
3	$N_{S_1} = 4$ and $N_{S_2} = 3$ ($N_{S_1} = 3$ and $N_{S_2} = 4$)	4	3	1	47	✓	✓
4	$N_{S_1} = 4$ and $N_{S_2} = 4$	4	4	1	96	✓	✓

N_{S_1} and N_{S_2} are the numbers of specific factors based on the secondary sources S_1 and S_2, respectively, n is the number of observations per combination of S_1 and S_2 factors, I is the total number of observed variables, and q is the number of unknown parameters. $I + \frac{I \times (I+1)}{2} - q \geq 0$ is the necessary condition for local identification. ✓ indicates local or empirical identification successfully holds under the specified condition; $-$ indicates local or empirical identification failed under the corresponding condition.
[a] For the general model we have $q = 5I$ and for the proportional model we have $q = 3I + N_{S_1} + N_{S_2}$ where q includes the factor loading parameters, the intercept parameters, and the error variance parameters, assuming the raw data (or the covariance matrix and the mean vector) are utilized for model estimation.
[b] When $N_{S_1} = 2$ and $N_{S_2} = 3$ ($N_{S_1} = 3$ and $N_{S_2} = 2$) with $n = 1$, we impose equality constraints for the general model on their loadings for the specific factors that include only two measurements. For the proportional model, such equality constraints are not appropriate as the proportional model includes only a common set of factor loadings (and specific factors do not have separate loading parameters).

the proportional model can be appropriate even when the general model is unsuitable due to underidentification.

4. Empirical Illustrations

Two real datasets were analyzed to demonstrate the utility of the proposed models. Both datasets were taken from the literature: (1) the data collected by Schneider and Schmitt (1992) for assessment center measurement, and (2) the data analyzed by Mellon and Crano (1977) for assessing students' academic ability as perceived by teachers.

These two examples from different research areas (industrial and educational psychology, respectively) present somewhat distinctive data features and study purposes. The first example utilizes an assessment that is designed to measure job performance, where two fully crossed factors are devised in the assessment to cover various aspects of employment skills. Accordingly, the data involve three sources of variance, associated with general job performance and two types of assessment design factors. A major research goal is to evaluate the internal structure of the assessment by evaluating the relative impacts of the two design factors compared to the general performance factor. Our proposed approach is a reasonable analytical tool in this case, as the structure of the data can be directly specified in the model, allowing us to assess contributions of

multiple variance components. The second example utilizes educational assessment data that were collected repeatedly over a 3-year period. Specifically, a group of students was assessed yearly by teachers on their abilities in three academic subject areas. This data example has the following features: (1) the data were not directly obtained from subjects, and (2) the occasion effect was completely confounded with the teacher effect as a different group of teachers participated in the assessment each year. In this situation, it is impossible to differentiate variability due to grades from teacher effects. However, it is still valid to examine consistency of students' general ability (perceived by teachers) across grades, teachers, and academic subjects. Our models represent the consistent and stable aspect of general ability by the general factor and the effects of grades (or teachers) and subjects by the secondary factors.

We reanalyzed the two data examples by fitting the proposed general model (Model A1) and proportional model (Model A2). In addition, we fit four existing models that researchers might consider as alternatives: We first considered a regular one-factor model (Model B) and a bifactor model (Model C). A regular one-factor model neglects influences from two secondary variance sources, while a bifactor model takes into account one of the sources of secondary variance but neglects possible influences from the other sources of secondary variance. Researchers interested in the measurement of a primary factor only might choose to utilize these models. Additionally, we also considered a correlated trait-correlated method (CTCM) model (e.g., Widaman, 1985) (Model D) and a correlated trait-correlated method-1 (CTC(M-1)) model (e.g., Eid, 2000) (Model E). A CTCM model (Model D) includes two sets of correlated factors without a general factor, while a CTC(M-1) model (Model E) is a variant of the CTCM model in which a reference factor (usually a method factor) is removed from the model. Different from the one-factor and bifactor models, CTCM and CTC(M-1) models do not include a general factor, meaning that researchers interested in the measurement of one of the secondary factors might select these models. Typically, CTCM and CTC(M-1) models are utilized when a set of 'trait' factors (of substantive interest) are measured through a set of 'methods'. Here we arbitrarily choose one set of secondary factors as 'traits' and the other set of factors as 'methods' in order to apply CTCM/CTC(M-1) models to the data.

The software Mplus (Muthén & Muthén, 2008) was used for full-information maximum likelihood estimation of all models. Example Mplus code for fitting the general and proportional models is provided in the supplementary material (Appendix E). For model assessment, we utilized goodness-of-fit statistics such as comparative fit index (CFI), Tucker-Lewis index (TLI) and the root-mean-square error of approximation (RMSEA). CFI values above .95 and TLI values above .95 are generally considered to indicate good fit (Hu & Bentler, 1999) and for RMSEA 0.01, 0.05, and 0.08 are considered as excellent, good, and mediocre fit (MacCallum et al., 1996). We also considered Akaike's information criterion (AIC) and the Bayesian information criterion (BIC), suitable for both nested as well as non-nested model comparisons. A smaller AIC or BIC value indicates that the model gives a better fit to the data when model complexity is taken into account.

4.1. Example 1: Assessment Center Measurement

Assessment center (AC) tools are widely used instruments for both employee selection and personnel development. AC designs typically include *dimensions* that refer to job-related attributes and *exercises* that consist of a series of miniaturized work samples designed to elicit job-related managerial behavior. Earlier debates focused on determining the relative importance of dimensions or exercises in AC assessment (e.g., Lievens et al., 2009). Some recent studies argue that a general performance factor should be included in addition to dimension and exercise factors (see, e.g., Hoffman et al., 2011). Our proposed models can be a useful tool for evaluating the internal AC structure and investigating the significance of a general performance factor in AC measurement.

TABLE 2.
Model fit information for Example 1 (Assessment Center Measurement).

Model	q	AIC	BIC	CFI	TLI	RMSEA
Proposed models						
Model A1: General	60	2611.550	2760.868	0.999	0.998	0.015
Model A2: Proportional	43	2597.267	2704.279	0.995	0.993	0.028
Existing models						
Model B: One-factor	36	2982.857	3072.447	0.416	0.286	0.287
Model C: Bifactor[a]	48	2614.586	2734.040	0.977	0.964	0.065
Model D: CTCM[b]	–	–	–	–	–	–
Model E: CTC(M-1)[c]	51	2619.989	2746.909	0.973	0.955	0.072

q is the number of parameters. Model A1: General model, Model A2: Proportional model, Model B: Bifactor model, Model C: CTCM, Model D: CTC(M-1).

[a] For the bifactor model, the exercise factors were considered as specific factors. When the dimension factors were used as specific factors, the model did not converge.

[b] The CTCM model did not converge.

[c] For the CTC(M-1) model, the first exercise factor was used as the reference factor. When the first exercise factor was chosen as the reference factor, the model was empirically underidentified.

In the proposed models, general job performance can be specified as a general factor while the dimensions and exercises can be modeled with two sets of secondary factors.

The data utilized in Schneider and Schmitt (1992) were originally obtained from the Michigan Department of Education. A total of 56 male and 33 female high school students, of mean age 16.9, received developmental feedback for their participation. About 14% of the students reported having a full-time job (40 h per week) and 84% reported having a part-time job. Twenty-one raters (twelve teachers, six school administrators/ counselors/ curriculum administrators, and three psychology graduate students) measured the students' twelve employability skills that were categorized into three dimensions (S_1): problem-solving, interpersonal skills, and initiative, and into four exercises (S2): grant allocation (competitive group discussion), team manufacturing (cooperative group discussion), customer service (cooperative role play), and team selection (competitive role play). The assessors provided a rating on a 5-point rating scale (1 = low, 5 = high) utilizing behavioral checklist scoring guidelines. The averaged ratings from two independent raters were used as the students' scores. Several steps were taken in the data collection stage to control for potential confounding effects that could arise from using multiple assessors and role players. For more information on the controls and data, see Schneider and Schmitt (1992). The data were available only in the form of a covariance matrix of the average ratings, so other potential complexities of the data were not considered in the current data analysis.

The proposed models are applied by treating the three dimensions as S_1-specific factors ($N_{S_1} = 3$) and the four exercises as S_2-specific factors ($N_{S_2} = 4$). There is one observation per combination of S_1-specific and S_2-specific factors ($n = 1$); therefore, a total of twelve measurements ($I = 12$) are utilized for data analysis. Under this condition, both general and proportional models are identified (see Condition 3 in Table 1). Table 2 lists the model fit information for the proposed models (Models A1 and A2) and four exiting models (Models B, C, D and E).

For the one-factor model, influences from both the dimension and exercise factors were not considered in the measurement of the general factor. We observed that the fit of the one-factor model was substantially poorer compared to the proposed models, suggesting that the general and proportional models provide a more plausible characterization of the underlying structure of the data. For the bifactor model, the dimension factors were ignored and the exercise factors

were treated as specific factors, assuming that the dimensions have little influence on the AC measurement (e.g., Bowler & Woehr, 2006). When the dimensions were considered as specific factors instead, the model did not converge. The bifactor model with the exercise factors, however, showed a good fit in terms of RMSEA even though the dimensions were ignored. This result appears to indicate that the dimension factors are unnecessary given the presence of a general performance factor. However, this initial interpretation is rejected based on the fact that the general model (Model A1) and the proportional model (Model A2) show a better fit than the bifactor model in terms of AIC (for the general model) and both AIC and BIC (for the proportional model). This suggests that the dimension factors do distinctively contribute to explaining the total observed variance when the general and exercise factors are present.

Assuming dimensions and exercises, but no general performance are measured with the AC data, a model with dimensions and exercises can be specified. Such a setup leads to the CTCM model (Model D); however, this model did not converge, unless one of the exercise factor was fixed as a reference factor (when one of the dimension factors was chosen as the reference, the model was empirically underidentified). Such a CTC(M-1) model (Model E) did converge but the fit was worse than the proposed models and the bifactor model. This result corroborates the idea that a general performance factor plays a non-trivial role in explaining the covariance structure of the data.

The general model (Model A1) and the proportional model (Model A2) fit well (and better than all three existing models) according to the CFI, TLI, and RMSEA, and the AIC and BIC values indicate that the proportional model is the best model when model complexity is taken into account. Therefore, we selected the proportional model (Model A2) to further inspect the parameter estimates.[1]

Table 3 lists the factor loading estimates (standard errors) for the general factor (G), constructed loading estimates for the two sets of specific factors (S_1, S_2) [as illustrated in matrix (4)], and the estimated proportionality constants for S_1 and S_2 ($C_{s_1}^{S_1}$, $C_{s_2}^{S_2}$). We also include in Table 3 the general factor loading estimates obtained from the one-factor model for comparison so that we can assess whether and how taking into account two secondary variance sources may lead to a different measurement of the target, general, factor.

Table 3 shows that the estimated factor loadings for the general factor (α_i^G) are non-trivial (ranging from 0.19 to 0.79). This gives support to our earlier conclusion that the inclusion of a general performance factor is indeed meaningful. The general factor loading estimates are larger in the proposed model than the estimated loadings in the one-factor model for the fourth to ninth manifest variables whose factor loading estimates for the exercise factors are smaller for the other six manifest variables. The factor loadings for the exercise factors ($\alpha_{is_2}^{S_2}$) are quite large (ranging from 0.56 to 0.92), confirming that the exercise factors have a strong influence on the AC data. The dimension factors ($\alpha_{is_1}^{S_1}$) have generally smaller factor loadings (ranging from 0.10 to 0.34) than the exercise factors, suggesting that the influences of the dimension factors are somewhat weaker than the exercise factors.

As discussed in Sect. 3.4, with a proportional model the general factor and two sets of specific factors can be interpreted as error components. The proportionality constants of the specific factors (in the bottom of Table 3) represent the effects of specific factors relative to the general factor. We found that the exercise (S_2) effects are approximately four to seven times as large as the dimension (S_1) effects. This result is consistent with a stream of research that supports the importance of exercises compared to dimensions in AC assessment (e.g., Lance et al., 2004). With respect to the dimension factors (S_1), the first factor (problem solving) and third factor (initiative) have

---

[1]In the general model estimation, the estimated factor loading for the 12th manifest variable on the third dimension factor (S_1) was found to be negative and non-significant ($\alpha_{123}^{S3} = -0.283$, SE $= 0.265$, $p = 0.142$), indicating that this manifest variable did not appear to be associated with the specific factor as expected. Accordingly, we set the factor loading for the 12th manifest variable to zero in estimating the proportional model.

TABLE 3.
Estimated model parameters and standard errors (in parenthesis) for the proportional model (Model A2) and the one-factor model (Model B) in Example 1 (Assessment Center Measurement).

i	α_i^{G*}	α_i^{G}	$\alpha_{i1}^{S_1}$	$\alpha_{i2}^{S_1}$	$\alpha_{i3}^{S_1}$	$\alpha_{i1}^{S_2}$	$\alpha_{i2}^{S_2}$	$\alpha_{i3}^{S_2}$	$\alpha_{i4}^{S_2}$	$\sqrt{\psi_{ii}}$	β_i
1	0.348	0.310	0.113			0.914				0.374	3.070
	(.114)	(.109)									(.110)
2	0.218	0.191		0.120		0.563				0.662	2.950
	(.099)	(.070)									(.095)
3	0.399	0.312			0.101	0.920				0.553	2.740
	(.122)	(.111)									(.119)
4	0.295	0.479	0.175				0.737			0.592	3.020
	(.116)	(.110)									(.114)
5	0.328	0.403		0.253			0.620			0.569	2.990
	(.101)	(.093)									(.103)
6	0.306	0.593			0.192		0.913			0.370	3.100
	(.124)	(.133)									(.124)
7	0.565	0.792	0.290					0.820		0.359	3.150
	(.126)	(.152)									(.130)
8	0.535	0.543		0.340				0.562		0.547	3.030
	(.103)	(.110)									(.107)
9	0.430	0.625			0.203			0.647		0.576	2.740
	(.116)	(.127)									(.115)
10	1.062	0.643	0.235						0.906	0.428	3.170
	(.102)	(.139)									(.129)
11	0.915	0.503		0.315					0.708	0.445	3.160
	(.091)	(.110)									(.109)
12	1.082	0.643			a				0.906	0.475	2.820
	(.101)	(.139)			−						(.128)

$C_1^{S_1}$	$C_2^{S_1}$	$C_3^{S_1}$	$C_1^{S_2}$	$C_2^{S_2}$	$C_3^{S_2}$	$C_4^{S_2}$

Proportionality constants

$C_1^{S_1}$	$C_2^{S_1}$	$C_3^{S_1}$	$C_1^{S_2}$	$C_2^{S_2}$	$C_3^{S_2}$	$C_4^{S_2}$
0.366	0.627	0.324	2.950	1.539	1.035	1.409
(.119)	(.143)	(.191)	(1.199)	(.414)	(.324)	(.393)

Dimensions are treated as S_1-specific factors and Exercises as S_2-specific factor. Here α_i^{G*} denotes the general factor loading estimates obtained from the one-factor model. α_i^{G}, $\alpha_{is_1}^{S_1}$ ($s_1 = 1, \ldots, 3$), and $\alpha_{is_2}^{S_2}$ ($s_2 = 1, \ldots, 4$) are the factor loadings obtained from the proportional model for the G, S_1, and S_2 factors, respectively; $\sqrt{\psi_{ii}}$ is the standard deviation of residual ϵ_{ip} and β_i is the intercept of the i-th manifest variable ($i = 1, \ldots, 12$); $C_{s_1}^{S_1}$ ($s_1 = 1, \ldots, 3$) and $C_{s_2}^{S_2}$ ($s_2 = 1, \ldots, 4$) are the proportionality constants for S_1, and S_2 factors, respectively.
Covariance matrix and means were used for analysis.
The loadings for the specific factors ($\alpha_{is_1}^{S_1}$, $\alpha_{is_2}^{S_2}$) were constructed based on the estimated general factor loadings (α_i^{G}) and the estimated proportionality constants ($C_{s_1}^{S_1}$, $C_{s_2}^{S_2}$).
[a] $\alpha_{123}^{S_1}$ was suppressed because the 12th manifest variable was not included in estimating the proportionality constant $C_3^{S_1}$ That is, the influence of the third dimension factor on the 12th manifest variable was suppressed to 0.

much smaller effects than the second factor (interpersonal skills). This result shows that a lower level employment/management skill, such as interpersonal skills, has a relatively larger weight than higher-level skills, such as problem solving and initiative. This is understandable given that

The relative contribution of each factor (γ_i^G (job performance), $\gamma_i^{S_1}$ (dimensions) and $\gamma_i^{S_2}$ (exercises), $i = 1, \ldots, 12$) to the systematic variances, the item-level reliability coefficients (ρ_i), corresponding reliabilities, $\rho_i^{(E)}$ and $\rho_i^{(D)}$, when dimension and exercise variances, respectively, are removed from the numerator, and the scale-level reliability coefficients ω_H and ECV for the proportional model (Model A2) in Example 1 (Assessment Center Measurement).

i	γ_i^G	$\gamma_{i1}^{S_1}$	$\gamma_{i2}^{S_1}$	$\gamma_{i3}^{S_1}$	$\gamma_{i1}^{S_2}$	$\gamma_{i2}^{S_2}$	$\gamma_{i3}^{S_2}$	$\gamma_{i4}^{S_2}$	ρ_i	$\rho_i^{(E)}$	$\rho_i^{(D)}$
1	0.102	0.014			0.885				0.871	0.859	0.100
2	0.099		0.039		0.862				0.457	0.439	0.063
3	0.102			0.011	0.887				0.757	0.749	0.085
4	0.285	0.038				0.676			0.697	0.670	0.225
5	0.266		0.104			0.630			0.654	0.585	0.242
6	0.288			0.030		0.682			0.899	0.872	0.286
7	0.454	0.061					0.486		0.915	0.859	0.470
8	0.406		0.159				0.435		0.708	0.595	0.401
9	0.460			0.048			0.492		0.719	0.684	0.365
10	0.321	0.043						0.636	0.876	0.838	0.318
11	0.296		0.116					0.588	0.812	0.717	0.335
12	0.324			a				0.642	0.850	0.821	0.304
ω_H	0.580										
ECV	0.299										

[a] $\gamma_{123}^{S_1}$ was suppressed because the influence of the third dimension factor on the 12th manifest variable was suppressed to 0.

most participants held an entry-level, part-time position. With respect to the exercise factors (S_2), the effects of competitive group discussion (first factor) is approximately three times as great as the effect of collaborative group discussion (third factor); in contrast, the effect of competitive role plays (second factor) is not much greater than the effect of collaborative role plays (fourth factor). This result implies that the competitive versus collaborative distinction is more effectual in the group discussion scenario than in the role play scenario when applied to participants with entry-level job positions.

Table 4 shows model-implied estimates of the relative contribution of each of the factors to the systematic variance. In addition, we calculated three types of model-based reliability coefficients. (1) the general, exercise, and dimension factors are systematic factors, (2) the dimension factors are considered as a nuisance, and (3) the exercise factors are considered as a nuisance. Comparison of the second and third coefficients to the first one can help us evaluate the impact of the secondary factors in the scale reliability.

These results confirm that the general performance factor does play an important role, explaining about 10–45% of the systematic variance. The scale-level coefficients are ECV = .299 and ω_H = .580. The exercise factors (S_2) explain a large portion of the systematic variance (about 44–88%). As discussed earlier, the first factor (comparative group discussion) shows a larger variance explained compared to the other factors. The contribution of each exercise factor appears generally consistent across the three dimensions. The dimension factors (S_1) explain about 16% of the systematic variance. The second dimension factor (initiative) shows a larger variance explained compared to the other two dimension factors (problem solving and initiative) which explain only little of the variance (less than 6%). All three dimension factors make negligible contributions when combined with the first exercise factor (comparative group discussion). The overall model-based reliabilities (ρ_i) are high with an average of 0.77 when all three factors (general, exercise, and dimension) are considered as systematic factors (only one manifest variable has reliability below 0.6.). Excluding the dimension factors ($\rho_i^{(E)}$) does not change the general reliabilities

substantially; however, excluding the exercise factors ($\rho_i^{(D)}$) seriously impacts the reliability coefficients. This result again confirms a greater importance of exercises compared to dimensions in this AC assessment example.

In summary, our data analysis results provide compelling evidence that the AC assessment has a multifaceted nature. We found that the exercise factors explain greater variance than the dimension factors, although both factors should be taken into account in the AC measurement. More importantly, we confirmed that a general performance factor plays a significant role in explaining the internal AC structure. The estimates for the proportional model also reveal that the weights of individual dimension and exercise factors relative to the general factor reflect the nature of the job that the AC assessment was applied to. An interesting future study would be to analyze additional AC assessment data applied to different kinds of job positions (e.g., managerial positions) and examine how the importance of distinctive job characteristics can be inferred from estimated weights of the AC factors.

4.2. Example 2: Overall Academic Performance

The second example examines elementary school students' academic performance as perceived by teachers. The original data were obtained from a study on homogeneous ability groupings in English and Welsh elementary schools (Barker Lunn, 1970). Children aged 7+ years were tested annually over a 4-year period on a number of academic and attitudinal variables. Teachers also assessed each child in their class on various subjective indicators, such as their perception of the child's reading, arithmetic, and general abilities. Specifically, the arithmetic and reading abilities were measured on 3-point scales with (1) among the top five to seven students, (2) average, and (3) in the bottom five to seven students. General academic ability was measured on a 5-point scale with (1) child was certain of grammar school placement, (2) child was above average, possible grammar school placement, (3) child was average, (4) child was below average, and (5) child was tested at the lowest level.[2] The test–retest reliabilities of the three teacher-rated ability measures ranged from .60 to .80. For the current analysis, we selected the covariance matrix of the teachers' ratings on 4,753 students over a 3-year period (second to fourth grades) which was provided in Mellon and Crano (1977, p. 721, the lower triangle of Table 5).

We apply our proposed models to evaluate consistency on students' overall ability perceived by teachers (as a general factor) while differentiating out the effects of two secondary variance sources, academic subjects (arithmetic, reading, and general; $N_{S_1} = 3$) and grades (second, third, fourth; $N_{S_2} = 3$). Note that as mentioned previously, the grade (year) effect is completely confounded with the teacher effect because a different group of teachers participated in the assessments each year. Hence, in essence, the overall grade effect should be interpreted as a combination of the grade and teacher effects.

In the data, there is one observation per combination of S_1-specific and S_2-specific factors ($n = 1$), giving a total of nine measurements ($I = 9$). In this setting, both general and constrained models are identified (Condition 2 in Table 1). Since the original dataset was unavailable, the covariance matrix published in Mellon and Crano (1977) was utilized for the data analysis. Hence, other data complexities (e.g., two different schools) were not taken into account.

Table 5 lists the model fit information for the proposed models (Models A1 and A2) and four exiting models (Models B, C, D, and E).

A one-factor model was specified which does not take into account the influence of academic subjects or grades in the measurement of the general performance factor. The fit of the one-factor model was considerably worse than the fit of the proposed models, indicating that the two secondary variance sources played an important role in explaining the covariance of the data. The bifactor model was specified with academic subjects as specific factors, assuming that variation

[2]The original wording is slightly revised.

TABLE 5.
Model fit information for Example 2 (Academic performance).

Model	q	AIC	BIC	CFI	TLI	RMSEA
Proposed models						
Model A1: General	36	95788.592	96021.387	0.992	0.966	0.071
Model A2: Proportional	24	96042.911	96198.107	0.981	0.968	0.070
Existing models						
Model B: One-factor	18	99227.326	99343.724	0.858	0.811	0.169
Model C: Bifactor[a]	27	97101.657	97276.253	0.940	0.881	0.134
Model D: CTCM [b]	–	–	–	–	–	–
Model E1: CTC(M-1)[c]	28	95956.272	96137.335	0.985	0.968	0.070
Model E2: CTC(M-1)	28	95957.148	96138.210	0.985	0.968	0.070

q is the number of parameters. Model A1: General model, Model A2: Proportional model, Model B: Bifactor model, Model C: CTCM, Model D: CTC(M-1).

[a] For the bifactor model, the subject factors were considered as specific factors. When the grade factors were used as specific factors, the model was not empirically identified.

[b] The CTCM model did not converge.

[c] Two CTC(M-1) models were fit. Model E1 used the first grade as the reference. Model E2 used the first academic subject as the reference factor.

in the ratings across the grades might be less substantial. This model, however, showed a poor fit in terms of CFI, TLI, and RMSEA. When the grades were included as specific factors (rather than the academic subjects), the bifactor model was not empirically identified. When both academic subject and grade-specific effects were taken into account (Models A1 and A2), the model fit improved substantially. The proportional model (Model A2) showed CFI, TLI, and RMSEA values that were close to the general model (Model A1); however, AIC and BIC values indicate that the general model is still a better fitting model even after model complexity is taken into account. This is in contrast to first example where the proportional model had lower AIC and BIC than the general model.

By assuming that there is no general ability factor, CTCM-type models could be specified. The full CTCM model (Model D) did not converge, however, unless one of the secondary factors was fixed as a reference factor. We considered two CTC(M-1) models (Models E1 and E2) with one subject and one grade as the reference, respectively; both models did converge and showed CFI, TLI, and RMSEA values that were comparable to the general model. In terms of AIC and BIC, however, the general model (Model A1) fit better than the two CTM(M-1) models.

At first glance, the fact that the CTC(M-1) models show a good fit seems to indicate that the role of the general factor may be negligible in explaining the covariance structure of the data. However, a closer inspection of the estimates for the general model (in Table 6) reveals that the general factor loading estimates (α_i^G) are far from negligible (ranging from 0.61 to 0.82). The general factor loading estimates from the general model are similar to those obtained from the one-factor model, which is understandable given that the general factor is a dominant factor in this example.

In Table 6, the factor loading estimates for the academic subject factors (S_1) and the grade factors (S_2) are significantly different from zero; however, the estimated general factor loadings are approximately two to three times as great as the specific factor loadings. This suggests that (1) both academic subjects and grades (and teachers) have statistically significant impacts on the manifest variables, and (2) the influence of the general ability factor is considerably more dominant than that of subjects and grades. Interestingly, teachers' perception of reading ability

TABLE 6.
Parameter estimates and standard errors (in parenthesis) for Example 2 (Academic performance) from the general model (Model A1) and the one-factor model (Model B).

i	α_i^{G*}	α_i^G	$\alpha_{i1}^{S_1}$	$\alpha_{i2}^{S_1}$	$\alpha_{i3}^{S_1}$	$\alpha_{i1}^{S_2}$	$\alpha_{i2}^{S_2}$	$\alpha_{i3}^{S_2}$	$\sqrt{\psi_{ii}}$
1	0.684	0.695	0.275			0.242			0.619
	(.013)	(.014)	(.019)			(.023)			
2	0.648	0.612		0.401		0.479			0.487
	(.013)	(.015)		(.018)		(.038)			
3	0.828	0.756			0.366	0.181			0.509
	(.012)	(.014)			(.017)	(.018)			
4	0.675	0.681	0.382				0.216		0.584
	(.013)	(.015)	(.023)				(.036)		
5	0.661	0.644		0.492			0.322		0.490
	(.013)	(.015)		(.019)			(.049)		
6	0.866	0.820			0.381		0.104		0.414
	(.012)	(.014)			(.017)		(.021)		
7	0.663	0.631	0.377					0.477	0.485
	(.013)	(.015)	(.022)					(.023)	
8	0.644	0.610		0.353				0.322	0.623
	(.013)	(.014)		(.016)				(.018)	
9	0.843	0.743			0.472			0.247	0.407
	(.012)	(.014)			(.016)			(.014)	

Academic subjects are treated as S_1-specific factors and grades as S_2-specific factors. Here α_i^{G*} indicates the general factor loading estimates obtained from the one-factor model. α_i^G, $\alpha_{is_1}^{S_1}$ ($s_1 = 1, \ldots, 3$), and $\alpha_{is_2}^{S_2}$ ($s_2 = 1, \ldots, 3$) are the factor loadings obtained from the general model for the G, S_1, and S_2 factors, respectively; $\sqrt{\psi_{ii}}$ is the standard deviation of residual ϵ_{ip} ($i = 1, \ldots, 9$).
Covariance matrix but no means was available for analysis (therefore no intercepts were estimated).

is most variable in grades 1 and 2, whereas their perception of general ability is most variable in grade 3, followed by math ability.

We computed the relative contribution of the general ability factor as well as the academic subject and grade factors on the individual manifest variables. We also calculated three types of model-based reliability coefficients (1) the general, subject, and grade factors are systematic factors, (2) the grade factors are considered as a nuisance, and (3) the subject factors are considered as a nuisance. Table 7 shows the results.

The general ability factor appears to be the most dominant factor, explaining about 49–81% of the systematic variance of individual manifest variables. The scale-level coefficients are ECV $= .655$ and $\omega_H = .843$. The academic subject factors (S_1) explain approximately 12–32% of the systematic variance for most manifest variables. The grade factors (S_2) explain about 7–30% of the variance of the manifest variables. The proportion of systematic variance explained by the grade specific factors is smallest for general academic ability (about 1–7%) suggesting that the teachers' ratings of students' general ability are more stable over time than their ratings of reading and math. The overall model-based reliability (ρ_i) is high with an average of 0.71 when all three factors (general, subject, and grade) are considered as systematic factors (no manifest variable shows reliability less than 0.6). The reliabilities after removing specific variances from the numerator for grade factors ($\rho_i^{(S)}$) and the subject factors ($\rho_i^{(G)}$) are not much lower, suggesting that the quality of the measurements is still adequate if either source is viewed as a nuisance.

In summary, the current analysis suggests that the proposed model is a useful tool to evaluate consistency of students' general ability perceived by teachers. The factor loading estimates as well

TABLE 7.
Relative contribution of each factor (γ_i^G (academic performance), $\gamma_i^{S_1}$ (subjects) and $\gamma_i^{S_2}$ (grades), $i = 1, \ldots, 9$), item-level reliability coefficients (ρ_i), corresponding reliabilities, $\rho_i^{(S)}$ and $\rho_i^{(G)}$, when grade and subject variances, respectively, are removed from the numerator, and the scale-level reliability coefficients ω_H and ECV for the general model (Model A1) in Example 2 (Academic performance).

i	γ_i^G	$\gamma_{i1}^{S_1}$	$\gamma_{i2}^{S_1}$	$\gamma_{i3}^{S_1}$	$\gamma_{i1}^{S_2}$	$\gamma_{i2}^{S_2}$	$\gamma_{i3}^{S_2}$	ρ_i	$\rho_i^{(S)}$	$\rho_i^{(G)}$
1	0.783	0.123			0.095			0.616	0.558	0.541
2	0.490		0.210		0.300			0.763	0.534	0.602
3	0.774			0.181	0.044			0.740	0.707	0.605
4	0.707	0.222				0.071		0.657	0.611	0.511
5	0.545		0.318			0.136		0.759	0.656	0.518
6	0.812			0.175		0.013		0.828	0.817	0.683
7	0.519	0.185					0.296	0.765	0.538	0.623
8	0.620		0.208				0.173	0.607	0.502	0.481
9	0.660			0.267			0.073	0.834	0.773	0.612
ω_H	0.843									
ECV	0.655									

as the manifest-variable and scale-level coefficients indicate that teachers' perception of students' ability is quite consistent across academic subjects, grades, and teachers. The unique contribution of academic subjects is smaller than that of the overall ability, putting into question the utility of using reading and mathematics subjects for the purpose of obtaining distinct, domain-specific abilities. An interesting future study would be to analyze additional assessment data including different academic subjects and grades and examine whether our findings can be replicated in other settings.

5. Discussion

In this article, we developed a CFA modeling approach that includes more than one set of secondary factors in addition to a target factor. A general model and its special case, referred to as a proportional model, were proposed, which constrains the specific factor loadings to be proportional to the general factor loadings. We discussed the proposed models' variance decomposition, reliability, and identification issues. Additionally, we conducted a simulation study to evaluate parameter recovery and scalability of the proposed models with an increasing number (K) of secondary sources. We found that parameter recovery of the general model as well as the proportional model appears to be satisfactory, with no obvious bias, when the full-information maximum likelihood estimation is used in Mplus. Details of the procedure as well as the results are provided in Appendix F of the supplementary material.

Two empirical studies were provided to illustrate the utilities of the proposed models. An interesting observation from the empirical analysis is that the proportional model was shown to be a better fitting model than the general model in the first example, while the general model was considered to be a better option than the proportional model in the second example. This result suggests that (1) in some cases, the proportional model can indeed be a parsimonious alternative to the general model, and (2) there may be other cases, however, where the general model provides a more accurate picture of the observed data. Although the model fit differences between the general and proportional models appear minor in all examples, further investigations are needed to fully grasp the underlying mechanisms for creating fit differences between proportional and general models.

The empirical study results also provide a practical guideline for applied researchers: in practice, it is beneficial to fit both general and proportional models. One can (1) first fit the general model and inspect whether there are sufficient impacts for secondary factors even when a general factor exists, (2) identify any manifest variables that are minimally associated with secondary factors, (3) take out the identified manifest variables (or set zero (general) factor loadings for those variables), (4) fit the proportional model, and (5) then determine whether the proportional model can be an efficient alternative to the general model, based on parameter estimates, fit statistics, and other theoretical/practical considerations.

The current study includes some limitations that suggest several avenues for future study. Although our proposed approach was presented in a general form (with $K > 2$ and $n > 1$), our illustrations were restricted to two secondary variance sources ($K = 2$) and a single measurement per combination of two secondary factors ($n = 1$). To fully demonstrate the generality of our proposed approach, various empirical studies could consider larger conditions with $K > 2$ and/or $n > 1$. For instance, it would be illuminating to apply the proposed approach to complex-design assessments, such as an educational assessment based on three design factors, e.g., academic subjects $\times$ contents $\times$ cognitive functions or a verbal aggression scale based on behavior mode $\times$ situation types $\times$ behavior types design factors. Applications with $n > 1$ would also be instructive in particular for (1) illustrating specification of interaction effects between secondary factors and (2) further demonstrating efficiency of a proportional model to reduce model complexity when applied to a large data problem.

References

Barker Lunn, J. (1970). *Streaming in the primary school*. Sussex: King, Thorne, & Stace.

Bauer, D. J., Howard, A. L., Baldasaro, R. E., Curran, P. J., Andrea, M. H., Chassin, L., et al. (2013). A trifactor model for integrating ratings across multiple informants. *Psychological Methods*, *18*, 475–493.

Becker, T. E., & Cote, J. A. (1994). Additive and multiplicative method effects in applied psychological research: An empirical assessment of three methods. *Journal of Management*, *20*, 625–641.

Bentler, P. M. (1976). Multistructure statistical model applied to factor analysis. *Multivariate Bahavioral Research*, *11*, 3–25.

Bollen, K. A., & Bauldry, S. (2010). Model identification and computer algebra. *Sociological Methods and Research*, *39*, 127–156.

Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology*, *91*, 1114–1124.

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, *64*, 153–168.

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

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, *75*, 581–612.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Cole, D., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *4*, 381–398.

Cronbach, L., Gleser, G., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Davis, W. (1993). The fc1 rule of identification for confirmatory factor analysis. *Sociological Methods and Research*, *21*, 403–407.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241–261.

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in g-factor models: Explanations and alternatives. *Psychological Methods*, *22*, 541–562.

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, *13*, 230–253.

Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., & Ladd, R. (2011). Exercises and dimensions are the currency of assessment centers. *Personnel Psychology*, *64*, 351–395.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, *57*, 239–251.

Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, *12*, 247–252.

Kenny, D. A. (1979). *Correlation and causality*. New York, NY: Wiley.

Lance, C. E., Lambert, T. A., Gerwin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology*, *89*, 377–385.

Lievens, F., Dilchert, S., & Ones, D. (2009). The importance of exercise and dimension factors in assessment centers: Simultaneous examinations of construct-related and criterion-related validity. *Human Performance*, *22*, 375–390.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, *13*, 335–361.

McDonald, R. P. (1982). A note on the investigation of local and global identifiability. *Psychometrika*, *47*, 101–103.

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

Mellenbergh, G. J., Kelderman, H., Stijlen, J. G., & Zondag, E. (1979). Linear models for the analysis and construction of instruments in a facet design. *Psychological Bulletin*, *86*, 766–776.

Mellon, P., & Crano, W. D. (1977). An extension and application of the multitrait-multimethod matrix technique. *Journal oi Educational Psychology*, *69*, 716–723.

Muthén, L., & Muthén, B. (2008). *Mplus user's guide*. Angeles, CA: Muthen & Muthen.

Pohl, S., & Steyer, R. (2010). Modeling common traits and method effects in multitrait.multimethod analysis. *Multivariate Behavioral Research*, *45*, 45–72.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence for fixed congeneric components. *Multivariate Behavioral Research*, *32*, 329–354.

Raykov, T., & Marcoulides, G. A. (2006). Estimation of generalizability coefficients via a structural equation modeling approach to scale reliability evaluation. *International Journal of Testing*, *6*, 81–95.

Reise, S. P., & Moore, T. M. (2013). Applying unidimensional item response theory models to psychological data. In K. Geisinger (Ed.), *APA handbook of testing and assessment in psychology. Test theory and testing and assessment in industrial and organizational psychology* (Vol. 1, pp. 101–119). Washington, DC: American Psychological Association.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*, 544–559.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013b). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*, 5–26.

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.

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.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*, 137–150.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61.

Schneider, J. R., & Schmitt, N. (1992). An exercise design approach to understanding assessment center dimension and exercise constructs. *Journal of Applied Psychology*, *77*, 32–41.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, *8*, 79–98.

Stucky, B. D., Thissen, D., & Edelen, M. O. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement*, *37*, 41–57.

Wainer, H. (1995). Precision and differential item functioning on a testlet-based test: The 1991 Law School Admissions Test as an example. *Applied Psychological Measurement*, *8*, 157–186.

Wainer, H., Bradlow, E., & Wang, X. (2007). *Testlet response theory and its applications*. New York, NY: Cambridge University Press.

Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*, *15*, 22–29.

Wald, A. (1950). A note on the identification of econometric relations. In T. C. Koopmans (Ed.), *Statistical inference in dynamic economic models* (pp. 238–244). New York: Wiley.

Widaman, F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, *9*, 1–26.

Woehr, D. J., Putka, D. J., & Bowler, M. C. (2012). An examination of G-theory methods for modeling multitrait multimethod data: Clarifying links to construct validity and confirmatory factor analysis. *Organizational Research Methods*, *15*, 134–161.

Wolfram Research, Inc (2010). *Mathematica Edition: Version 8.0*. Champaign, IL: Wolfram Research, Inc.

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.