

Teaching confirmatory factor analysis to non-statisticians: a case study for estimating the composite reliability of psychometric instruments

Byron J. Gajewski

University of Kansas Medical Center, USA

Yu Jiang

University of Kansas Medical Center, USA

Hung-Wen Yeh

University of Kansas Medical Center, USA

Kimberly Engelman

University of Kansas Medical Center, USA

Cynthia Teel

University of Kansas Medical Center, USA

Won S. Choi

University of Kansas Medical Center, USA

K. Allen Greiner

University of Kansas Medical Center, USA

Christine Makosky Daley

University of Kansas Medical Center, USA

Texts and software that we are currently using for teaching multivariate analysis to non-statisticians lack in the delivery of confirmatory factor analysis (CFA). The purpose of this paper is to provide educators with a complement to these resources that includes CFA and its computation. We focus on how to use CFA to estimate a “composite reliability” of a psychometric instrument. This paper provides step-by-step guidance for introducing, via a case-study, the non-statistician to CFA. As a complement to our instruction about the more traditional SPSS, we successfully piloted the software R for estimating CFA on nine non-statisticians. This approach can be used with healthcare graduate students taking a multivariate course, as well as modified for community stakeholders of our Center for American Indian Community Health (e.g. community advisory boards, summer interns, & research team members). The placement of CFA at the end of the class is strategic and gives us an opportunity to do some innovative teaching: (1) build ideas for understanding the case study using previous course work (such as ANOVA); (2) incorporate multi-dimensional scaling (that students already learned) into the selection of a factor structure (new concept); (3) use interactive data from the

students (active learning); (4) review matrix algebra and its importance to psychometric evaluation; (5) show students how to do the calculation on their own; and (6) give students access to an actual recent research project.

Keywords: Pile Sorting; Instrument Development; Multivariate Methods; Center for American Indian Community Health

1. Introduction

Using two excellent texts by Johnson (1998) and Johnson and Wichern (2007), we teach a course called “Applied Multivariate Methods” to non-statistician graduate students. Johnson’s *Applied Multivariate Methods for Data Analysts* text is desirable because the author “talks” through the approaches using everyday terminology, although with less mathematical detail. The Johnson and Wichern (2007) text complements the Johnson book, but without overwhelming the non-mathematician. Johnson (1998) mainly uses SAS to deliver computational details, while Johnson and Wichern (2007) do not showcase particular software. Most of our students are in the healthcare behavioral sciences (e.g. PhD nursing, audiology, etc.). They mostly do not have mathematical backgrounds and predictably prefer to use SPSS. Accordingly, we use SPSS in our lectures to complement the SAS presentation in Johnson (1998).

However, in our estimation there is a major shortcoming to the delivery of the current curriculum. While the texts cited above currently deliver PCA and factor analysis at an excellent level, we are dissatisfied with the delivery of *confirmatory* factor analysis (CFA) (e.g. Pett, Lackley, & Sullivan, 2003; Brown, 2006) for testing a particular factor structure. Although CFA is more appropriately called “testing factor analysis” (Wainer, 2011) because it can statistically test parameters and factor structures from a wide variety of models, we will still use its classic name, CFA, in this paper. The lack of CFA in these texts might be due to lack of accessible software. Over the last few years SAS expanded its ability to fit CFAs using “proc calis.” In fact, Christensen (2011) is writing a new edition of *Methods of Multivariate Analysis* by Rencher (2002) that highlights programming in SAS, but this is geared towards statistics majors. SPSS requires the purchase of an extra “add-on” called AMOS in order to fit CFA. Many of our students cannot afford to purchase this extra software (either SAS or AMOS). One option would be to use a free trial version of CFA-focused software called Mplus (<http://www.statmodel.com/>), which students can use as a limited demonstration. This option, however, quickly becomes impractical because it allows only six variables. Therefore, we propose that students use the freeware R (<http://www.r-project.org/>) for performing the CFA calculations. As demonstrated later in the paper, R

has been successfully used by our own students as a supplement to SPSS. Others also have suggested that R be used more often (Burns, 2007) and it has been used successfully in courses with audiences of non-statisticians (e.g. Zhou & Braun, 2010). We think it is a valuable service to our students to continue using SPSS. However, the current gap in the SPSS capabilities to fit CFA models is a perfect opportunity to incorporate R as a supplement for teaching multivariate analyses to non-statisticians.

We are cognizant that complex mathematics can be difficult for non-statistical students. Our philosophy is to combine both graphical techniques (Yu, et al., 2002; Valero-Mora & Ledesma, 2011) and traditional matrix algebra (Johnson & Wichern, 2007) to overcome these difficulties. Our students tend to have minimal background in matrix algebra, so we spend a week (three hours) reviewing matrix algebra (e.g. addition, trace, determinant, etc) using syntax in SPSS. Students initially resist, but they later see a payoff particularly when understanding the connection that Eigen values and eigenvectors have to factor analysis (as well as almost all of multivariate statistics). Our students quickly learn that the benefit of learning to manipulate matrices is worth the cost.

The purpose of this paper is to provide educators of non-statistical graduate students with a complement to texts such as Johnson (1998) and Johnson & Wichern (2007) that includes CFA and its computation in R. We will focus on how to use R’s output to estimate a “composite reliability” of a multivariate psychometric instrument (Alonso, et al., 2010). To accomplish this we will develop a case study using data from two recent clinical trials, supported by the American Heart Association and the National Institutes of Health, respectively, in which we collected data from the 20-item CES-D instrument (measure of depressive symptomatology, e.g. Nguyen, et al., 2004) in caregivers of stroke and Alzheimer’s disease patients.

The reasons underlying the pressing need for CFA in our current curriculum are twofold. First, many of our students are nurse researchers and psychometric

measurement is essential to their work. Second, most of us are members of the Center for American Indian Community Health (CAICH), a National Institute on Minority Health and Health Disparities-funded Center of Excellence, whose goal is to use community-based participatory research (Israel, et al., 2005) methods to integrate community members into all phases of the research process. Part of the mission of CAICH is to create and modify existing methods to make them more applicable to community-based participatory research and more “friendly” for use by community members and community researchers who may not be as familiar with statistics. In true community-based participatory research, community members help with all parts of research, including data analysis. The ability to use freeware and have simple explanations for complex analyses is of paramount importance to conduct this type of research.

One research project being conducted by CAICH involves the development of an American Indian-focused mammography satisfaction psychometric instrument that is culturally sensitive (Engelman, et al., 2010). We will be estimating the reliability of this instrument using CFA. Following the principles of community-based participatory research, we also want to communicate the methodology to non-statisticians, because we are *studying with* a community and *not just studying the* community.

To facilitate the communication of quantitative methods, we have developed a series of CAICH methods core guides. To date, we have developed guides that introduce descriptive statistics in SPSS and R, spatial mapping in SAS and R, data collection via the web (i.e. Comprehensive Research Information Systems, CRIS), pile sorting in SAS, and CFA in R. We include in our guides step-by-step guidance that all stakeholders of CAICH (e.g. community advisory boards, summer interns, & research team members) can use to learn the specific statistical methods we use to support CAICH.

Posing a structure to the CFA can be a challenge, so we offer two approaches. First, because CES-D has been studied in other populations, structures have been proposed and we will use them on the new dataset. Second, in a novel approach, we perform a “single pile sort” (e.g. Trotter & Potter, 1993) of the 20 items from several graduate healthcare students to come up with a proposed factor analytic structure. This pile sort brings together many opinions (six of our graduate students) using multidimensional scaling (MDS). Essentially we ask each student to (i) guess at the number of clusters and (ii) place the items in each of these clusters. From this we

create a ‘dissimilarity matrix’ of the number of times pairs of items are not together. Then a classic MDS is fitted.

Of note is that because CES-D has been studied in other populations, structures have been proposed and we will use them. We will focus on the utility of CFA estimating *reliability* and not focus on goodness-of-fit which can be further examined for example in Kaplan (2000) (i.e. Comparative Fit Index and Root Mean Squared Error Approximation).

In Section 2 we provide the basic materials for instructors to teach CFA from a case study point-of-view. We first define reliability using a classical ANOVA model. This approach has a pedagogical advantage because it reviews materials from a previous course. This ANOVA model transitions into CFA and motivates the multivariate version of composite reliability. We provide an illustrative example for teaching the basics. In Section 2 we also discuss in more detail the data used for the CFA case study before presenting the results of the case study. We close section 2 focusing on classroom implementation. In Section 3 we give a discussion and concluding remarks.

2. Methods and teaching materials for the case-study

2.1. The class

We now provide some details regarding the students we teach, classroom setup, what we the teachers and students actually did, and a setup of how to generally implement the case study. The first author of this paper is the teacher of record for the Applied Multivariate Methods class. By the time the students reached the Spring 2011 course, they all had taken Analysis of Variance (ANOVA), Applied Regression, and had a working knowledge and familiarity with SPSS. The semester’s topics are traditionally: vectors & matrices, multivariate analysis of variance (MANOVA), principal component analysis (PCA), factor analysis, discriminate analysis, canonical correlation, and multidimensional scaling (MDS).

Because a majority of our health research involves psychometric instruments, we motivate PCA and factor analysis from it. In the most recent year we required the nine students in the class to perform the CFA calculations using R. Our class is taught across 16 weeks, each with one three-hour lecture. The lectures also include interactive “drills” as we are strong believers in teaching statistics using active learning principles (more on “drills” later). The presentation of CFA background is done after MDS. The case-study and presentation of R

computations represents one and a half lectures. This is in addition to two full lectures on principal component analysis and factor analysis. All of the steps in this paper are provided to the student in lecture format. Extensions to the lectures serve as an exercise in the form of homework and final exam questions.

The placement of CFA at the end of the class is strategic and gives us an opportunity to do some innovative teaching: (1) build ideas for understanding the case study using previous course work (such as ANOVA); (2) incorporate multi-dimensional scaling (that students already learned) into the selection of a factor structure (new concept); (3) use interactive data from the students (active learning); (4) review matrix algebra and its importance to psychometric evaluation; (5) Show students how to do the calculation on their own; and (6) give students access to an actual recent research project.

2.2. ANOVA-based reliability (previous coursework and knowledge)

As mentioned previously, educating non-statisticians about how to fit CFA for estimating the reliability of a psychometric instrument is the focus of this paper. Because our students have taken ANOVA, it is useful to motivate the idea of a reliable instrument using an ANOVA model. Assuming that we have p different items composing an instrument, a very basic model called the “parallel test” (Lord & Novick, 1968) is defined as follows for the i^{th} subject:

$x_{ij} = \mu_j + f_i + \varepsilon_{ij}$, where $i=1,2,3,\dots,n$ subjects and $j=1,2,3,\dots,p$ items. The f_i is the “true” unknown score for the i^{th} subject and μ_j is the mean for the j^{th} item. It is classically assumed that $f_i \sim N(0, \sigma_f^2)$ are each independently distributed as well as independent of $\varepsilon_{ij} \sim N(0, \sigma^2)$. The σ_f^2 represents the variance of the true scores and σ^2 represents the variance of the measurement error.

A reliable test is one in which the items’ scores (x ’s) are highly correlated with the true scores (f ’s). Specifically the correlation matrix is,

$$\text{corr}(x_{ij}, f_i) = \begin{bmatrix} 1 & \sigma_f^2 / \sqrt{(\sigma_f^2 + \sigma^2)} \sigma_f^2 \\ \text{Sym} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \sigma_f / \sqrt{(\sigma_f^2 + \sigma^2)} \\ \text{Sym} & 1 \end{bmatrix}$$

There liability is defined as the squared correlation $R_T = \sigma_f^2 / (\sigma_f^2 + \sigma^2) = 1 - \sigma^2 / (\sigma_f^2 + \sigma^2)$ and represents the “average item reliability;” if high, one can argue for a

short version to the test (i.e. 1 item). Note that the “T” in R_T stands for “trace” which will be clarified later.

However, as is often the case with instruments, several questions are needed for a reliable instrument. Typically these items are either summed or averaged for each subject. For example, $\bar{x}_{i\bullet} = \mu_{\bullet} + f_i + \bar{\varepsilon}_{i\bullet}$ represents the average of the observed items for subject i . In this case, the correlation between the average of the observed x ’s for each subject and the true scores (f ’s) is

$$\text{corr}(\bar{x}_{i\bullet}, f_i) = \begin{bmatrix} 1 & \sigma_f / \sqrt{(\sigma_f^2 + \sigma^2 / p)} \sigma_f^2 \\ \text{Sym} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \sigma_f / \sqrt{(\sigma_f^2 + \sigma^2 / p)} \\ \text{Sym} & 1 \end{bmatrix}$$

This reliability is known as the “entire reliability” because it uses all of the items. It is the squared correlation $R_A = \sigma_f^2 / (\sigma_f^2 + \sigma^2 / p) = 1 - (\sigma^2 / p) / (\sigma_f^2 + \sigma^2 / p)$. Note that the “A” in R_A stands for “determinant” which will be clarified later and that as long as $\sigma_f^2 > 0$, the number of items (p) gets large and the reliability approaches 1. Shrout has identified the following interpretations for reliability: 0.00-0.10 is virtually none; 0.11-0.40 is slight; 0.41-0.60 is fair; 0.61-0.80 is moderate; and 0.81-1.0 is substantial (Shrout, 1998). As we will later show, a psychometric instrument can have less than moderate average item reliability (i.e. $R_T < .6$) but substantial entire reliability (i.e. $R_A > .8$).

However, some items might be more or less reliable than others; for example, how agreeable a participant is that “green is a favorite color” might be uncorrelated with the total depression score, so clearly this item reduces the reliability if it is kept as a question for measuring the true score. But it might be a perfectly reasonable question for measuring a participant’s desire to eat peas. A more flexible modeling approach is a multivariate estimate of reliability via a confirmatory factor analysis (CFA).

2.3. Confirmatory factor analysis-based reliability (new idea)

Next we form a more flexible model by transitioning from univariate ANOVA to a multivariate CFA model. Transform the responses to the items (x ’s above) to be \mathbf{x} , a vector of observable p variables from a subject. There are q factors. A fairly general factor analytic equation is $\mathbf{x}_{p \times 1} = \boldsymbol{\mu}_{p \times 1} + \boldsymbol{\Lambda}_{p \times q} \mathbf{f}_{q \times 1} + \mathbf{e}_{p \times 1}$ where the j^{th} element of $\boldsymbol{\mu}$ is the mean for that item, λ_{jk} is the “factor loading” for the j^{th} variable on the k^{th} loading, the k^{th} element of \mathbf{f}_k is the k^{th} common factor and \mathbf{e} is the specific error not explained by the common factors. The distributional

assumptions are placed on the factor scores and the specific error:

- \mathbf{f} and \mathbf{e} are independent;
- \mathbf{f} is standardized so it has an average of 0 and a covariance matrix with diagonal of 1's and off-diagonals that represent the correlation among factor scores. In other words, $\mathbf{f} \sim \mathbf{MVN}(\mathbf{0}, \mathbf{\Phi})$;
- \mathbf{e} typically has non-zero diagonal but off-diagonal is 0, so $\mathbf{e} \sim \mathbf{MVN}(\mathbf{0}, \mathbf{\Psi})$.

From a reliability standpoint, this model has two advantages over the ANOVA-based model. First, the model allows for multiple factors. For example, depression can be measured by a two factor analytic model as how you feel and how you think others view you. Second, because the λ 's are not equal to 1, as in the ANOVA model, the reliability of each of the items is allowed to vary.

Just as in the ANOVA model, reliability can be defined using the covariance of \mathbf{x} , the observed responses, and the covariance of \mathbf{f} , the true scores:

$$\mathbf{cov}(\mathbf{x}, \mathbf{f}) = \begin{bmatrix} \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi} & \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' \\ \text{Sym} & \mathbf{\Phi} \end{bmatrix},$$

because it is difficult to take ratios of matrices, we can use the multivariate algebra “trace” or “determinant” to estimate a composite-type (Graham, 2006) reliability (Alonso, et al, 2010). Therefore the multivariate “average reliability” is: $R_T = 1 - \text{tr}(\mathbf{\Psi}) / \text{tr}(\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi})$ and the “entire” reliability is $R_\lambda = 1 - |\mathbf{\Psi}| / |\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}|$. Note that if we set $\mathbf{\Lambda} \equiv \mathbf{\lambda}$ and fix the diagonal of $\mathbf{\Psi}$, we get the ANOVA-based model considered earlier. As noted in Alonso, et al. (2010), the “average reliability” can be written as a weighted average of each individual item’s reliability $R_T = \sum_{j=1}^p v_j R_j$, where

$$v_j = (\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi})_{jj} / \left\{ \sum_{j=1}^p (\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi})_{jj} \right\} \&$$

$$R_j = (\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}')_{jj} / (\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi})_{jj}.$$

The R_j denotes the j^{th} item’s reliability and v_j is the weight associated with the j^{th} item. This final parameter can help

us define the factor based on the item with the highest reliability and decide what items are not properly contributing to the factor’s reliability.

2.4. Deciding what to confirm in CFA: a single pile sort (previous coursework and knowledge applied to new knowledge)

As noted in Johnson & Wichern (2007), in general, the matrix of loadings is non-unique. There are a number of ways that this can be handled. One very practical way is to fit a factor analytic model via principal component analysis (PCA). This fit results in a model that produces factor scores that are uncorrelated but difficult to interpret. An alternative is to fit a PCA via promax rotation. This will give us a better interpretation. A third approach is to *a priori* define items to load onto only one factor. In this case we have uniqueness. However, proposing which items load onto which domain can be a challenge.

Of consideration in this paper involves (1) past literature in which a CFA was fitted using the same instrument with a different population; (2) in the case of a new instrument it might be advantageous to query expert opinion regarding this issue. Novel in this paper is to elicit expert opinion regarding what items belong together. Note that graduate students are treated as the “experts” here. In general research we would carefully select experts, although most of the graduate students have vast clinical experience and serve as a very close set of experts.

More specifically, there are two things that need to be determined from this elicitation process: (1) The number of factors (e.g. domains or subscales) q there should be and (2) Which factor each item belongs to (i.e. defining the structure of $\mathbf{\Lambda}$). Once these are determined a CFA can be fitted. When eliciting expert opinion, each expert reports their own opinion of the number of factors and the structure of $\mathbf{\Lambda}$.

From this we can define a dissimilarity matrix d_{jj} , which represents the number of times an expert placed item j and j' in different factors ($d_{jj}=0$ means all experts placed these items together). From this dissimilarity matrix we will get a composite structure for $\mathbf{\Lambda}$ and an estimate of k . This will be estimated through multidimensional scaling (MDS) and a cluster analysis on the stimulus values (reduced dimension coordinates for each item such that the dissimilarity matrix is preserved optimally) from the MDS. This approach pools experts’ opinion and reviews the statistical methods they have already learned.

2.5. A small illustrative example (new idea)

Before elicitation of expert opinion, it is necessary to provide some background of the matrix Λ and other parameters involved in a CFA. Therefore, in this section we provide a small illustrative example in which to operationally define terms and models for CFA. As a future research problem, our team is interested in developing a culturally tailored psychometric instrument that measures ‘mammography satisfaction’ (Cockburn et al 1991) of Native Americans. This is an important concept to measure in order to understand health disparities and promote health equity. Suppose we have drafted six items ($x_1, x_2, x_3, x_4, x_5,$ and x_6) that represent ‘mammography satisfaction.’ Then after feedback from a single expert, we say that these items represent two factors: one measuring ‘convenience and accessibility (f_1)’ and the other ‘information transfer (f_2).’ From a pile sorting perspective, this provides a dissimilarity matrix with 0’s and 1’s. But we don’t need to perform an MDS on such a matrix because it provides an obvious structure. From a matrix perspective the model looks like $\mathbf{x}_{6 \times 1} = \boldsymbol{\mu}_{6 \times 1} + \Lambda_{6 \times 2} \mathbf{f}_{2 \times 1} + \mathbf{e}_{6 \times 1}$, which can be written as:

$$\begin{aligned} x_1 &= \mu_1 + \lambda_{11}f_1 + e_1 \\ x_2 &= \mu_2 + \lambda_{21}f_1 + e_2 \\ x_3 &= \mu_3 + \lambda_{31}f_1 + e_3 \\ x_4 &= \mu_4 + \lambda_{42}f_2 + e_4 \\ x_5 &= \mu_5 + \lambda_{52}f_2 + e_5 \\ x_6 &= \mu_6 + \lambda_{62}f_2 + e_6. \end{aligned}$$

Notice that the latent variable f_1 is only influencing the first three items and f_2 the last three items. The μ ’s are intercepts and the λ ’s are slopes that are called “loadings” in a factor analysis. This is the same name of a similar term in exploratory factor analysis. The λ ’s represent the covariance between f ’s and x ’s. Suppose for simplicity that the x ’s are standardized (mean=0 and sd=1), then $\mu_1 = \dots = \mu_6 = 0$ and λ ’s are then the correlation between f ’s and x ’s. We emphasize to the students that in this case, $x_1, x_2,$ and x_3 are only correlated with ‘convenience and accessibility’ and $x_4, x_5,$ and x_6 are only correlated with ‘information transfer.’

Therefore, translating the multivariate model we have:

$$\Lambda = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}$$

The standardized model further results in the variance of the f ’s to be 1 and the variances of the error to be $1-\lambda^2$. Specifically,

$$\Phi = \begin{bmatrix} 1 & \phi_{12} \\ Sym & 1 \end{bmatrix} \text{ and}$$

$$\Psi = \begin{bmatrix} 1-\lambda_{11}^2 & 0 & 0 & 0 & 0 & 0 \\ & 1-\lambda_{21}^2 & 0 & 0 & 0 & 0 \\ & & 1-\lambda_{31}^2 & 0 & 0 & 0 \\ & & & 1-\lambda_{42}^2 & 0 & 0 \\ & Sym & & & 1-\lambda_{52}^2 & 0 \\ & & & & & 1-\lambda_{62}^2 \end{bmatrix}$$

For illustrative purposes, suppose $\lambda_{11} = \dots = \lambda_{62} = 0.5$ and $\phi_{12} = 0.25$.

We now wish to illustrate how “average” and “entire” reliabilities are calculated using these parameters. This is an opportunity for students to use SPSS syntax or use R (shown later) to calculate. Some students may want to see an illustration “by hand.”

$$\Lambda\Phi\Lambda' = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0 & 0.5 \\ 0 & 0.5 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.0625 & 0.0625 & 0.0625 \\ & 0.25 & 0.25 & 0.0625 & 0.0625 & 0.0625 \\ & & 0.25 & 0.0625 & 0.0625 & 0.0625 \\ & & & 0.25 & 0.25 & 0.25 \\ & Sym & & & 0.25 & 0.25 \\ & & & & & 0.25 \end{bmatrix}$$

$$\Lambda\Phi\Lambda' + \Psi = \begin{bmatrix} 1 & 0.25 & 0.25 & 0.0625 & 0.0625 & 0.0625 \\ & 1 & 0.25 & 0.0625 & 0.0625 & 0.0625 \\ & & 1 & 0.0625 & 0.0625 & 0.0625 \\ & & & 1 & 0.25 & 0.25 \\ & Sym & & & 1 & 0.25 \\ & & & & & 1 \end{bmatrix}$$

So $tr(\Psi) = 4.5$, $tr(\Lambda\Phi\Lambda' + \Psi) = 6$, and the “average” reliability is $R_T = 1 - 4.5/6 = 0.25$, which is a “slight” reliability. Further, $|\Psi| = 0.1780$, and $|\Lambda\Phi\Lambda' + \Psi| = 0.7008$ making $R_A = 1 - 0.1780/0.7008 = 0.7460$, which is a “moderate” reliability.

2.6. CES-D instrument (actual recent research project)

The Center for Epidemiological Studies Depression Scale (CES-D) was developed to measure depressive symptomatology. The stem for each of the 20 items is “Please tell me how often you have felt this way in the past week.” An example item is “I was bothered by things that usually don’t bother me.” The response options are “rarely,” “some,” “occasionally,” and “most.” The recommended scoring is 0, 1, 2, and 3 for respective response options, except for items 4, 8, 12, and 16 which are reversed scored (3, 2, 1, and 0) because they are on a positive scale (i.e. Item 4. “I felt I was just as good as other people”). One typically sums these scores to gain a score of depression ranging from 0 to 60, with higher scores indicating more “symptomatology.” We will investigate whether this scoring is reliable for our caregiver population or if a more flexible model is necessary. Note that the response options of the items of the CES instruments are ordinal. Polychoric correlations, rather than Pearson correlations, could be used for the analysis. However, when participants respond to items having three or more response options the Pearson and Polychoric estimates are very similar (e.g. Gajewski et al, 2010). Therefore, we report the Pearson version in this paper.

2.7. Incorporate active learning: data from students (active learning)

Each of our lectures is setup so that at the very beginning of a new topic (e.g. CFA), we perform a drill (Gelman, 2005) that serves as a transition piece into the new topic. For one of the drills, after introducing the CES-D scale and the basic CFA model with examples from other applications, we asked the students to each provide us with answers to the following questions: (1) What do you

think the dimensionality of the CES-D is? (2) What items go into each dimension? Six of the students answered these questions that supplied the MDS. Note that this step will be used eventually to help define a structure for the mammography satisfaction instrument, but instead of students we will elicit this information from experts (i.e. executive community advisory board members of CAICH).

2.8. Data from caregivers (actual recent research project)

Study participants included spousal caregivers age 55 and older, whose partners had either received a diagnosis of probable Alzheimer’s disease within the previous 2 years or had experienced a first-ever stroke between 6 months and 3 years before enrollment in the study. All caregivers were currently living with their spouses or spousal equivalent partners, and providing unpaid physical, emotional, social, financial, or decision-making support. Caregivers were referred from memory and stroke clinics and from support groups.

Intervention group participants received a self-care intervention (Self-Care TALK, SCT) that was provided by advanced practice nurses. The six-week SCT program consisted of weekly 30-minute phone calls from the nurse to the spouse caregiver. Topics included healthy lifestyle choices, self-esteem, focusing on the positive, avoiding overload, communicating, and building meaning. The SCT topics were based on theoretical and empirical findings related to self-care and health promotion in aging, particularly in the context of caregiving challenges. Because depression is such a ubiquitous caregiving outcome, the intervention was designed to provide caregivers with self-care strategies to maintain health while managing a new caregiving role.

Alzheimer’s subjects data includes a treatment group (n=27) and comparison usual care group (n=19). The stroke subjects data includes a treatment group (n=21) and comparison usual care group (n=18). Therefore, we have a total of 85 participants from whom we used data from baseline. All participants took the CES-D before randomization (baseline) and at two time points post-treatment. We only use the CES-D at baseline because it is the place not influenced by treatment.

2.9. Software: SPSS and R (performing calculations with R)

For non-statisticians in the social and behavioral sciences, SPSS is by far the most popular statistical software program. Because of this, a university tends to purchase a

license for SPSS that allows most graduate students to gain a copy at minimal expense; in our case, free of charge. This is the software we use for calculating everything in class (e.g. vectors & matrices, MANOVA, PCA, non-CFA factor analysis, discriminate analysis, canonical correlation, and MDS). For CFA, however, we fit a model using R. The R contribution, known as lavaan, is a free, open source R package for latent variable analysis. It can be used for multivariate statistical analysis, such as path analysis, confirmatory factor analysis and structural equation modeling (Rosseel, 2011). Full examples and documentation maybe found at <http://lavaan.ugent.be>.

2.10. CFA structure from the CFA via single pile sort (incorporate MDS)

Before running the CFA, we need specification of the number of factors and the structure of from the student experts. Here are their results using the methodology described in Section 2.7. The stimulus coordinates (from a standard MDS) for the two-dimensional model is presented in Figure 1. It has a stress coefficient=0.1077 which is considered fair (Johnson & Wichern, 2007). However, a three-dimensional model (three-cluster analysis) gives the same clouds of items as shown in Figure 1. These clouds define three different factors because they represent the most similar sets. The cluster analysis also suggested a five factor model which we will fit later. We do not give the specifics of the MDS here since that is a methodology covered earlier in the course.

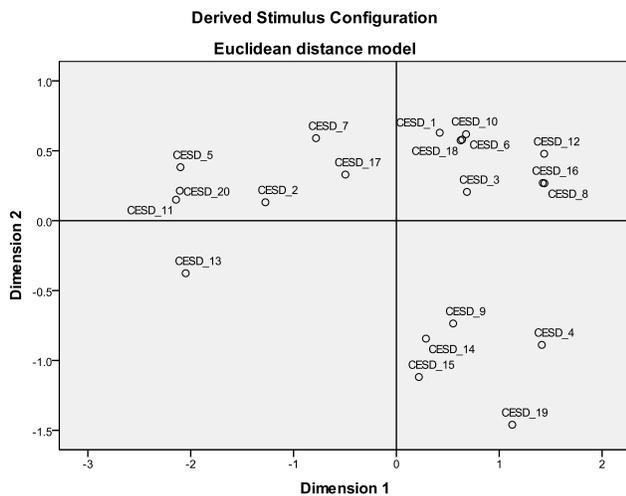


Figure 1. Scatter plot for stimulus coordinates estimates from a standard MDS (stress=.1077, fair) for a two-dimensional model.

Here is the three factor model, dropping λ subscripts:

$$A' = \begin{bmatrix} \lambda & 0 & \lambda & 0 & 0 & \lambda & 0 & \lambda & 0 & \lambda & 0 & \lambda & 0 & 0 & 0 & \lambda & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 & 0 & \lambda & \lambda & 0 & 0 & 0 & \lambda & 0 \\ 0 & \lambda & 0 & 0 & \lambda & 0 & \lambda & 0 & 0 & 0 & \lambda & 0 & \lambda & 0 & 0 & 0 & \lambda & 0 & 0 & \lambda \end{bmatrix}$$

2.11. Example CFA in R (lavaan) with two factors(review matrix algebra)

The tutorial explaining the basic use of the lavaan package and the reference manual can be found at: <http://lavaan.ugent.be/>. The following is a shorter version. Let the x's and y's (e.g. actual items) be variables that we observe and f's be latent variables (e.g. depression).

2.11.1. Model syntax

In the R environment and lavaan package, the set of formula types is summarized in following table:

| Formula type | Operator |
|----------------------------|----------|
| regression | ~ |
| Latent variable definition | =~ |
| (residual) (co) variance | ~~ |
| intercept | ~ 1 |

A regression formula usually has the following form: $y \sim x_1 + x_2 + x_3 + x_4$. We should define latent variables by listing their manifest indicators first, using the operator “=”. For example, we can denote the three latent variables f1 and f2 as: $f_1 = \sim y_1 + y_2 + y_3$ and $f_2 = \sim y_4 + y_5 + y_6$. Variances and covariances can be specified by ‘double tilde’ operator, such as $y_1 \sim y_1$. Intercepts for observed and latent variables are denoted by the number “1” when the regression formula only has an intercept as the predictor.

2.11.2. Fitting latent variables: confirmatory factor analysis

Using the CES-D accompanied data we fit a three factor example (Figure 2) whose structure is defined from the pile sort. The data consist of depression results of caregivers from two different studies (Section 2.6).

A CFA model for these 20 variables consists of three latent variables (or factors):

- (i) a *depressed* factor measured by 8 variables: CESD_10, CESD_18, CESD_6, CESD_1, CESD_3, CESD_8, CESD_16 and CESD_12
- (ii) a *failure* factor measured by 5 variables: CESD_14, CESD_15, CESD_9, CESD_4 and CESD_19

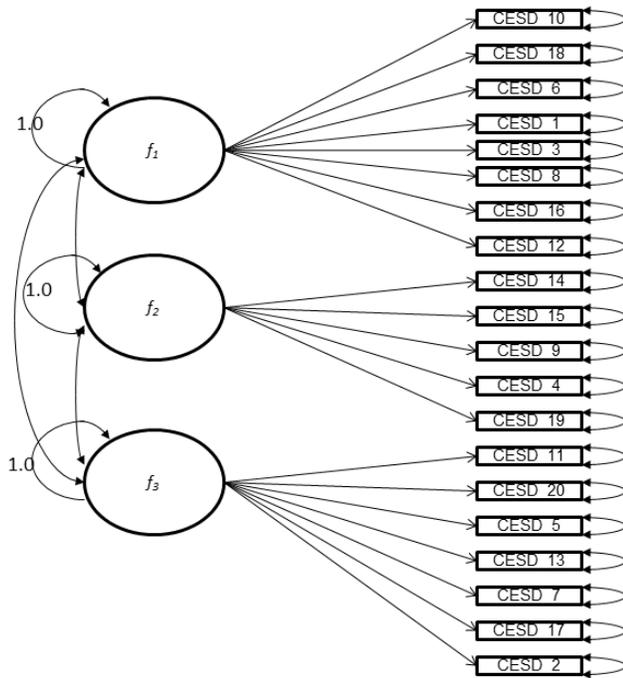


Figure 2. A three factor structure of CES-D CFA.

(iii) an *effort* factor measured by 7 variables: CESD_11, CESD_20, CESD_5, CESD_13, CESD_7, CESD_17 and CESD_2

We now detail the R code in the shaded area and below that give the corresponding output. In this example, the model syntax only contains three 'latent variable definitions.' Each formula has the following format:

latent variable =~ indicator1 + indicator2 + indicator3

The R code is the following. Before running the R code, save the accompanied dataset as CESD1.csv in C drive (we did it from MS Word, 2007).

```
dataset=read.table("C:/CESD1.csv",header=T, sep=",")
head(dataset)
# load the lavaan package (only needed once per session)
library(lavaan)
# specify the model
HS.model2 <- ' f1 =~
CESD_10+CESD_18+CESD_6+CESD_1+CESD_3+
CESD_8+CESD_16+CESD_12
f2 =~
CESD_14+CESD_15+CESD_9+CESD_4+CESD_19
f3 =~
CESD_11+CESD_20+CESD_5+CESD_13+CESD_7
+CESD_17+CESD_2'
# fit the model
fit <- cfa(HS.model2, data=dataset, std.lv=TRUE)
```

```
# display summary output
E<-standardizedSolution(fit)
The output (stored in matrix E) looks like this:
```

| lhs op | rhs | est | est.std | est.std.all |
|--------|--------------------|--------|---------|-------------|
| 1 | F1 =~ CESD_10 | 0.517 | 0.517 | 0.641 |
| 2 | F1 =~ CESD_18 | 0.684 | 0.684 | 0.803 |
| 3 | F1 =~ CESD_6 | 0.733 | 0.733 | 0.829 |
| 4 | F1 =~ CESD_1 | 0.633 | 0.633 | 0.716 |
| 5 | F1 =~ CESD_3 | 0.634 | 0.634 | 0.760 |
| 6 | F1 =~ CESD_8 | -0.408 | -0.408 | -0.422 |
| 7 | F1 =~ CESD_16 | -0.478 | -0.478 | -0.628 |
| 8 | F1 =~ CESD_12 | -0.523 | -0.523 | -0.577 |
| 9 | F2 =~ CESD_14 | 0.588 | 0.588 | 0.773 |
| 10 | F2 =~ CESD_15 | 0.173 | 0.173 | 0.460 |
| 11 | F2 =~ CESD_9 | 0.359 | 0.359 | 0.681 |
| 12 | F2 =~ CESD_4 | -0.316 | -0.316 | -0.462 |
| 13 | F2 =~ CESD_19 | 0.257 | 0.257 | 0.588 |
| 14 | F3 =~ CESD_11 | 0.412 | 0.412 | 0.402 |
| 15 | F3 =~ CESD_20 | 0.612 | 0.612 | 0.681 |
| 16 | F3 =~ CESD_5 | 0.581 | 0.581 | 0.605 |
| 17 | F3 =~ CESD_13 | 0.519 | 0.519 | 0.612 |
| 18 | F3 =~ CESD_7 | 0.641 | 0.641 | 0.789 |
| 19 | F3 =~ CESD_17 | 0.476 | 0.476 | 0.750 |
| 20 | F3 =~ CESD_2 | 0.407 | 0.407 | 0.733 |
| 21 | CESD_10 =~ CESD_10 | 0.383 | 0.383 | 0.589 |
| 22 | CESD_18 =~ CESD_18 | 0.257 | 0.257 | 0.355 |
| 23 | CESD_6 =~ CESD_6 | 0.244 | 0.244 | 0.312 |
| 24 | CESD_1 =~ CESD_1 | 0.381 | 0.381 | 0.488 |
| 25 | CESD_3 =~ CESD_3 | 0.294 | 0.294 | 0.422 |
| 26 | CESD_8 =~ CESD_8 | 0.769 | 0.769 | 0.822 |
| 27 | CESD_16 =~ CESD_16 | 0.351 | 0.351 | 0.606 |
| 28 | CESD_12 =~ CESD_12 | 0.549 | 0.549 | 0.667 |
| 29 | CESD_14 =~ CESD_14 | 0.233 | 0.233 | 0.402 |
| 30 | CESD_15 =~ CESD_15 | 0.112 | 0.112 | 0.788 |
| 31 | CESD_9 =~ CESD_9 | 0.150 | 0.150 | 0.537 |
| 32 | CESD_4 =~ CESD_4 | 0.367 | 0.367 | 0.786 |
| 33 | CESD_19 =~ CESD_19 | 0.126 | 0.126 | 0.654 |
| 34 | CESD_11 =~ CESD_11 | 0.881 | 0.881 | 0.838 |
| 35 | CESD_20 =~ CESD_20 | 0.434 | 0.434 | 0.537 |
| 36 | CESD_5 =~ CESD_5 | 0.584 | 0.584 | 0.634 |
| 37 | CESD_13 =~ CESD_13 | 0.451 | 0.451 | 0.626 |
| 38 | CESD_7 =~ CESD_7 | 0.249 | 0.249 | 0.377 |
| 39 | CESD_17 =~ CESD_17 | 0.177 | 0.177 | 0.438 |
| 40 | CESD_2 =~ CESD_2 | 0.143 | 0.143 | 0.463 |
| 41 | F1 =~ F1 | 1.000 | 1.000 | 1.000 |
| 42 | F2 =~ F2 | 1.000 | 1.000 | 1.000 |
| 43 | F3 =~ F3 | 1.000 | 1.000 | 1.000 |
| 44 | F1 =~ F2 | 0.850 | 0.850 | 0.850 |
| 45 | F1 =~ F3 | 0.982 | 0.982 | 0.982 |
| 46 | F2 =~ F3 | 0.819 | 0.819 | 0.819 |

Storing this output in matrix form makes it convenient for calculating matrix operations such as R_T and R_Λ . First look at this in pieces.

(1) Here is the R language and output for $\Lambda_{20 \times 3}$

```
#Lamda matrix
lamda<-matrix(rep(0,20*3),ncol=3,byrow=TRUE)
lamda[1:8,1]=E$est[1:8]
lamda[9:13,2]=E$est[9:13]
lamda[14:20,3]=E$est[14:20]
lamda
```

| | [,1] | [,2] | [,3] |
|-------|------------|------------|-----------|
| [1,] | 0.5165803 | 0.0000000 | 0.0000000 |
| [2,] | 0.6837385 | 0.0000000 | 0.0000000 |
| [3,] | 0.7329825 | 0.0000000 | 0.0000000 |
| [4,] | 0.6328223 | 0.0000000 | 0.0000000 |
| [5,] | 0.6338626 | 0.0000000 | 0.0000000 |
| [6,] | -0.4083308 | 0.0000000 | 0.0000000 |
| [7,] | -0.4777197 | 0.0000000 | 0.0000000 |
| [8,] | -0.5234052 | 0.0000000 | 0.0000000 |
| [9,] | 0.0000000 | 0.5881401 | 0.0000000 |
| [10,] | 0.0000000 | 0.1733599 | 0.0000000 |
| [11,] | 0.0000000 | 0.3592858 | 0.0000000 |
| [12,] | 0.0000000 | -0.3157676 | 0.0000000 |
| [13,] | 0.0000000 | 0.2574780 | 0.0000000 |
| [14,] | 0.0000000 | 0.0000000 | 0.4118990 |
| [15,] | 0.0000000 | 0.0000000 | 0.6121472 |
| [16,] | 0.0000000 | 0.0000000 | 0.5811397 |
| [17,] | 0.0000000 | 0.0000000 | 0.5193046 |
| [18,] | 0.0000000 | 0.0000000 | 0.6407709 |
| [19,] | 0.0000000 | 0.0000000 | 0.4764386 |
| [20,] | 0.0000000 | 0.0000000 | 0.4068932 |

| | |
|-------|-----------|
| [4,] | 0.3809886 |
| [5,] | 0.2935812 |
| [6,] | 0.7694604 |
| [7,] | 0.3511613 |
| [8,] | 0.5490247 |
| [9,] | 0.2326383 |
| [10,] | 0.1116765 |
| [11,] | 0.1496685 |
| [12,] | 0.3667272 |
| [13,] | 0.1255392 |
| [14,] | 0.8808588 |
| [15,] | 0.4344115 |
| [16,] | 0.5840764 |
| [17,] | 0.4505997 |
| [18,] | 0.2487908 |
| [19,] | 0.1766047 |
| [20,] | 0.1425351 |

(2) Here is the R language and output for $\Phi_{3 \times 3}$

```
#Phi matrix
Phi<-matrix(rep(0,9),ncol=3, byrow=TRUE)
diag(Phi)<-1 #standardized
Phi[1,2:3]=E$est[44:45]
Phi[2,3]=E$est[46]
Phi[2:3,1]=E$est[44:45]
Phi[3,2]=E$est[46]
Phi
```

| | [,1] | [,2] | [,3] |
|------|-----------|-----------|-----------|
| [1,] | 1.0000000 | 0.8495581 | 0.9821504 |
| [2,] | 0.8495581 | 1.0000000 | 0.8187922 |
| [3,] | 0.9821504 | 0.8187922 | 1.0000000 |

(3) Here is the R language and output for Ψ

```
#Psi matrix
Psi<-matrix(rep(0,20^2),ncol=20, byrow=TRUE)
diag(Psi)<-E$est[21:40]
diag(Psi) #off-diagonals are zero
```

| | [,1] |
|------|-----------|
| [1,] | 0.3828351 |
| [2,] | 0.2572087 |
| [3,] | 0.2441909 |

(4) Here is the R language and output for $R_T = 1 - tr(\Psi) / tr(\Lambda\Phi\Lambda' + \Psi)$ and $R_\Lambda = 1 - |\Psi| / |\Lambda\Phi\Lambda' + \Psi|$

```
COV<-lamda%*%Phi%*%(t(lamda))+Psi
RT<-1-(sum(diag(Psi)))/(sum(diag(COV)))
RL<-1-(det(Psi))/(det(COV))
RT
```

| | |
|-----|-----------|
| [1] | 0.4294543 |
|-----|-----------|

RL

| | |
|-----|-----------|
| [1] | 0.9737757 |
|-----|-----------|

(5) Here is the R language and output for item reliabilities $R_j = (\Lambda\Phi\Lambda')_{jj} / (\Lambda\Phi\Lambda' + \Psi)_{jj}$

```
G<-lamda%*%Phi%*%(t(lamda))
V<-lamda%*%Phi%*%(t(lamda))+Psi
R<-matrix(rep(0,20),ncol=1, byrow=TRUE)
R[1:20,1]<-(diag(G)/diag(V))
R
```

| | [,1] |
|-------|-----------|
| [1,] | 0.4107422 |
| [2,] | 0.6450860 |
| [3,] | 0.6875173 |
| [4,] | 0.5124611 |
| [5,] | 0.5778015 |
| [6,] | 0.1780976 |
| [7,] | 0.3938989 |
| [8,] | 0.3328802 |
| [9,] | 0.5978922 |
| [10,] | 0.2120486 |
| [11,] | 0.4630820 |
| [12,] | 0.2137680 |
| [13,] | 0.3455847 |

[14,] 0.1615018
 [15,] 0.4631166
 [16,] 0.3663739
 [17,] 0.3744077
 [18,] 0.6226888
 [19,] 0.5624248
 [20,] 0.5373696

The parameters above are shown in graphical form in Figure 3.

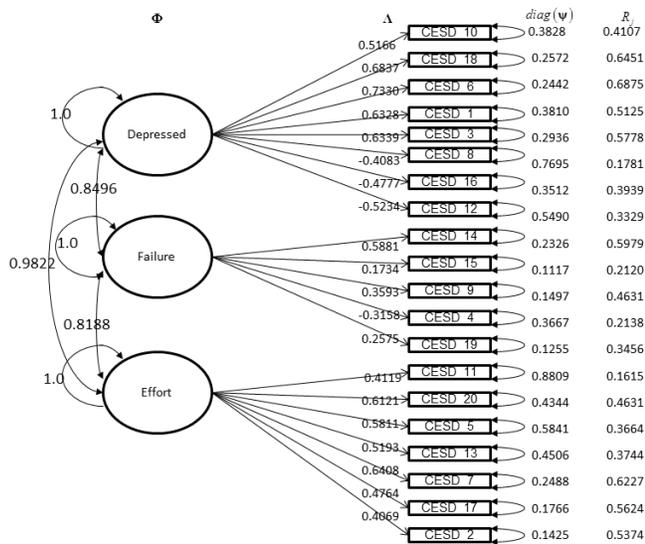


Figure 3. Model 2 of CESD CFA.

2.12. Results of all the models (actual research project)

First, as noted in Section 2, we estimated reliabilities for the “parallel test model”; the estimated reliabilities are $R_T=0.3730$ and $R_A=0.9220$. Note that in this case it is assumed that the reliability is the same for each of the items so R_T is in the classification of “slight.” The “entire” reliability (R_A) is much better because it is in the “substantial” range. Therefore, it is not reliable to perform an analysis (e.g. regression) using the individual items. But the entire items, specifically using the sum (or average) is reliable.

However, using four other models we would like to know if the reliability can be improved (average, entire, and item) and if the reliability of an item in a model and subscale can be used to label a hypothesized factor (or subscale). Model 1 is a single factor model; models 2 and 3 are respectively the three and five factor models motivated from our students’ pile sort, and model 4 is the four factor structure quoted in Nguyen, et al. (2004), which also supplied the item abbreviations shown in the reliability output in Table 1.

The average and the entire reliabilities both increase as the number of factors grow. However, the average reliability never creeps into the “fair” territory. The entire reliability is always in the “substantial” range.

The item reliabilities guide us in naming the factors in each of the models. Clearly the single factor in model 1 can be named depressed. For model 2, the three factors are depressed, lonely, and effort. For model 3, the factors are depressed, enjoyed, lonely, going, and effort. For model 4, the factors are effort, depressed, enjoyed, and dislike. Item 6 is a very reliable item for measuring depression across all models and it makes sense when looking at its specific questioning “In the past week I felt depressed.”

Of note, all of the individual reliabilities are out of the virtually none category, except for item 4. This reverse scored item measures “good” with more specifically “In past week I felt that I was just as good as other people;” perhaps caregivers have a difficult time answering this because they are spending most of their time and effort concerned about a single individual rather than interacting with others.

A conclusion to draw from this actual research project can be summarized. From our analysis of the caregiver CES-D data, the single factor model has substantial entire reliability ($R_A=0.9433$), thus we are justified in using a sum score in the final analysis. This composite approach to reliability justifies the usual sum score scoring practice.

2.13. Implementation: recommendations to the teacher on what to do in the classroom

Implementation of our case-study approach for teaching CFA can be handled in many different ways. It is in our experience that the ideas expressed in Sections 2.2-2.12 have been presented using 4.5 hours of class time and using two exercises, one as a homework example and the other as part of a take-home final exam. We implemented the case study at the end of the course after the last topic (multidimensional scaling, MDS). Implementation this late allows us to incorporate: (1) a review of matrix algebra; (2) a review of factor analysis; (3) a review of MDS; and (4) introduce CFA which includes computation (see Table 2).

Some teachers of multivariate may not want to teach MDS and/or may want to implement CFA much earlier in the course (say after PCA and factor analysis).In this

case we would not implement the MDS portion of the case-study approach.

This is not a problem. One can still use the active learning drills. Instead of combining the opinion of the

class members one can fit a different CFA for each class member. If the teacher feels it's necessary, this would be a good opportunity to incorporate goodness-of-fit and have a race to see which class member has the best model fit. This is a good opportunity for home exercise.

Table 1. Results of four different factor analyses, from a single factor to a five factor model. The R_i represents the item reliability for a specific factor. A superscript means that item is most reliable for that factor and such can be used to help name the factor. Note that using the “parallel test model” the estimated reliabilities are $R_T=0.3730$ and $R_A=0.9220$.

| Item | Abbreviation | M1 | | M2 | | M3 | | M4 | |
|------|--------------|--------|--------------------|-------------------------|--------------------|------------------------|--------------------|---|--------------------|
| | | Single | R_i | Pile Sort Three Factors | R_i | Pile Sort Five Factors | R_i | National (Nguyen et al, 2004, Four Factors) | R_i |
| 1 | Bothered | F1 | 0.51 | F1 | 0.51 | F1 | 0.48 | F1 | 0.51 |
| 2 | Appetite | F1 | 0.54 | F3 | 0.54 | F5 | 0.56 | F1 | 0.55 |
| 3 | Blues | F1 | 0.57 | F1 | 0.58 | F1 | 0.60 | F2 | 0.59 |
| 4r | Good | F1 | 0.15 | F2 | 0.21 | F3 | 0.20 | F3 | 0.09 |
| 5 | Mind | F1 | 0.35 | F3 | 0.37 | F4 | 0.44 | F1 | 0.35 |
| 6 | Depressed | F1 | 0.67 ^{F1} | F1 | 0.69 ^{F1} | F1 | 0.73 ^{F1} | F2 | 0.71 ^{F2} |
| 7 | Effort | F1 | 0.61 | F3 | 0.62 ^{F3} | F5 | 0.65 ^{F5} | F1 | 0.60 ^{F1} |
| 8r | Hopeful | F1 | 0.17 | F1 | 0.18 | F2 | 0.32 | F3 | 0.34 |
| 9 | Failure | F1 | 0.29 | F2 | 0.46 | F3 | 0.46 | F2 | 0.25 |
| 10 | Fearful | F1 | 0.41 | F1 | 0.41 | F1 | 0.43 | F2 | 0.43 |
| 11 | Sleep | F1 | 0.16 | F3 | 0.16 | F4 | 0.19 | F1 | 0.16 |
| 12r | Happy | F1 | 0.33 | F1 | 0.33 | F2 | 0.63 | F3 | 0.65 |
| 13 | Talk | F1 | 0.37 | F3 | 0.37 | F4 | 0.44 | F1 | 0.37 |
| 14 | Lonely | F1 | 0.58 | F2 | 0.60 ^{F2} | F3 | 0.62 ^{F3} | F2 | 0.56 |
| 15 | Unfriendly | F1 | 0.11 | F2 | 0.21 | F3 | 0.21 | F4 | 0.17 |
| 16r | Enjoyed | F1 | 0.40 | F1 | 0.39 | F2 | 0.83 ^{F2} | F3 | 0.78 ^{F3} |
| 17 | Crying | F1 | 0.54 | F3 | 0.56 | F5 | 0.64 | F2 | 0.57 |
| 18 | Sad | F1 | 0.63 | F1 | 0.65 | F1 | 0.64 | F2 | 0.63 |
| 19 | Dislike | F1 | 0.23 | F2 | 0.35 | F3 | 0.32 | F4 | 0.42 ^{F4} |
| 20 | Going | F1 | 0.45 | F3 | 0.46 | F4 | 0.53 ^{F4} | F1 | 0.44 |
| | R_T | 0.4135 | | 0.4295 | | 0.5032 | | 0.4699 | |
| | R_A | 0.9433 | | 0.9738 | | 0.9970 | | 0.9907 | |

“r” stands for “reverse” scoring item.

Table 2. Summarization of how a teacher of statistics could actually implement the ideas expressed in this paper.

| Section(s) | Topic | Implementation |
|------------|--|--|
| 2.2 | ANOVA-based Reliability | In-class lecture |
| 2.3 | Confirmatory Factor Analysis-based Reliability | In-class lecture |
| 2.4 & 2.7 | Deciding What to Confirm in CFA: A Single Pile Sort* | In-class active learning (“Drills”)& Home Exercise |
| 2.5 | A Small Illustrative Example | In-class lecture |
| 2.6 & 2.8 | Data example | In-class lecture & Home Exercise |
| 2.9 | Software: SPSS and R | In-class lecture & Home Exercise |
| 2.10 | CFA structure from the CFA via single pile sort* | In-class lecture & Home Exercise |
| 2.11 | Example CFA in R (laavan) with three Factors | In-class lecture |
| 2.12 | Results of all the Models | Home Exercise |

The focus of this paper has been implementation of the case-study in a graduate course for non-statisticians. However, we mentioned earlier our motivation for implementation in community-based participatory research (CBPR). The best way to effectively communicate the case-study in this setting is to present the small illustrative example, perform a pile sort, and

then present the results for all models. The R calculations should be presented in more detail once a community member is curious about those details or has taken the course work: ANOVA, regression, and are taking multivariate. The latter would mean that we would follow the entire structure laid out in Table 2.

3. Discussion and conclusion

In Section 2 we provided a step-by-step approach for implementing a case-study approach for teaching CFA to non-statistical graduate students. In our own class we used 4.5 hours of class-time that is a mix between lecture and active learning sessions. An instructor of a similar graduate course will be able to use the material immediately for teaching. The case-study approach uses six points to guide the presentation: (1) Previous Coursework and Knowledge; (2) Multi-Dimensional Scaling; (3) Active Learning; (4) Matrix Algebra; (5) Doing Calculations with R and (6) Actual Recent Research Project.

Two attractive features of the case study approach are (1) it uses a relatively new way to think of CFA; and (2) it introduces an exciting technology to better estimate reliability. This new way to think of CFA does not stop at interpreting the parameters of the model. Rather it gathers them together to summarize the entire and average reliability of an instrument.

Our biggest worry in introducing the supplement to SPSS, R, was in how the graduate students would respond. We worried that they would protest until we stopped using a supplemental free software package and abandon doing CFA altogether. Contrary to this worry, the students responded rather well. We piloted the CFA approach using R on our nine graduate students who are non-statisticians. In a homework assignment they followed the instructions on their own for downloading R and successfully fitting a CFA using a dataset that is a part of laavan. As part of a take-home final exam, the students saved the CES-D data in SPSS and all students except one fitted and successfully interpreted a CFA model. This is particularly impressive considering the students were required to work independently. This result is consistent with the success Zhou & Braun (2010) had in teaching R to non-statisticians.

In the same class, we taught matrix manipulations (supporting the reliability calculations) using syntax code in SPSS. However, in the future we will use R to do these matrix manipulations. This will allow our students to engage in R very early in the class so that they may be better prepared for the CFA much later in the course. Furthermore, the SPSS Statistics-R Integration Package is available such that students may perform R under their familiar SPSS environment through syntax (Muenchen, 2008, p. 28 - 32). We do not see R replacing SPSS as we believe it would be a disservice to most of our students; their academic advisors expect them to have proficiencies in SPSS. However, as word gets out about R we will not

be too surprised if academic advisors also want students to complement their SPSS knowledge.

It is important to note on what we did not focus in this paper. Namely, we mentioned very little about the classic goodness of fit measures such as comparative fit index (CFI) and root mean squared error approximation (RMSEA). Both are emphasized in the course and can help disentangle which CFA model is best in terms of number of factors and possible correlated error structures.

This paper provides step by step guidance for instructors to educate non-statisticians on CFA using a case-study approach. The approach is designed and has been implemented with healthcare graduate students taking a multivariate course. It can also be used with community members participating in the research process. Given the pilot's success, we will use this manuscript and the CAICH methods core guides as tools for communicating to all stakeholders of CAICH (e.g. community advisory boards, summer interns, & research team members) the specific reliability analysis of the future mammography satisfaction instrument. All CAICH methods core guides are available for public use by contacting the lead author or going to our website (www.caich.org). Any graduate school educator or community-based participatory researcher can use our methods to train students or other researchers to perform CFA using freely available software.

Acknowledgements: Partial funding for the first and fifth authors comes from grants from the United States of America (USA) NIH, National Institute of Nursing Research (1R21NR009560) and American Heart Association. Partial funding for all the authors, except the fifth, comes from a grant from the USA NIH, National Institute on Minority Health and Health Disparities (5P20MD004805). We thank Lili Garrard for helpful review and comments of an earlier version of the paper.

Correspondence: bgajewski@kumc.edu

REFERENCES

- Alonso, A., Laenen, A., Molenberghs, G., Geys, H., & Vangeneugden, T. 2010. A Unified Approach to Multi-item Reliability. *Biometrics*, 66, 1061-1068.
- Brown T.A. 2006. Confirmatory Factor Analysis for Applied Research. *Guilford Press*.
- Burns, P. 2007. R Relative to Statistical Packages: Comment 1 on Technical Report Number 1 (Version 1.0) Strategically using General Purpose Statistics Packages: A Look at Stata, SAS and SPSS. *UCLA Technical Report Series*, www.burns-stat.com/pages/Tutor/R_relative_statpack.pdf
- Christensen, W. 2011. Personal Communication, October 6, 2011.
- Cockburn J, Hill D, Irwig L, De Luise T, Turnbull D, & Schofield P. 1991. Development and Validation of an Instrument to Measure Satisfaction of Participants at Breast Screening Programmes. *Eur J Cancer*, 27 (7), 831-835.
- Engelman, K.K, Daley, C.M, Gajewski, B.J, Ndikum-Moffor, F, Faseru, B, Braiuca, S, Joseph, S, Ellerbeck E, & Greiner, K.A. 2010. An Assessment of American Indian Women's Mammography Experiences. *BMC Women's Health*, 10(34). PMID: PMC3018433.
- Gajewski, B.J, Boyle, D.K, and Thompson, S. 2010. How a Bayesian Might Estimate the Distribution of Cronbach's alpha from Ordinal-Dynamic Scaled Data: A Case Study Measuring Nursing Home Residents Quality of Life. *Methodology*, 2, 71-82.
- Gelman, A. 2005. A course on teaching statistics at the university level. *The American Statistician*, 59, 4-7.
- Graham, J.M. 2006. Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability: What They Are and How to Use Them. *Educational and Psychological Measurement*, 66, 930-944.
- Israel B.A, Parker E.A, Rowe Z, Salvatore A, Minkler M, López J, et al. 2005. Community-Based Participatory Research: Lessons Learned from the Centers for Children's Environmental Health and Disease Prevention Research. *Environ Health Perspect*, 113, 1463-1471.
- Johnson, D.E. 1998. Applied Multivariate Methods for Data. *Johnson/Duxbury Press*, New York.
- Johnson, R.A & Wichern, D.W. 2007. Applied Multivariate Statistical Analysis, *Pearson*, Upper Saddle River, NJ.
- Kaplan, D. 2000. Structural Equation Modeling: Foundations and Extensions, *Sage Publications*, London.
- Lord, F.M & Novick, M.R. 1968. Statistical Theories of Mental Test Scores, *Addison-Wesley Publishing Company*, Reading, MA.
- Muenchen, R.A. 2008. R for SAS and SPSS Users, *Springer*, New York.
- Nguyen, H.T, Kitner-Triolo, M, Evans, M.K & Zonderman, A.B. 2004. Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. *Psychiatry Research*, 126(2), 177-87.
- Pett, M.A, Lackley, N.R, & Sullivan, J.J. 2003. Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research. *Thousand Oaks, CA: Sage*.
- Rencher A.C. 2002. Methods of Multivariate Analysis. *Wiley*, New York.
- Rosseel, Y. 2011. lavaan: latent variable analysis, <http://lavaan.ugent.be>.
- Shrout, P.E. 1998. Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, 7, 301-317.
- Trotter R.T & Potter J.M. 1993. Pile Sorts, A Cognitive Anthropological Model of Drug and AIDS Risks for Navajo Teenagers: Assessment of a New Evaluation Tool. *Drug & Society: A Journal of Contemporary*. Issues, 7 (3/4), 23 – 39.
- Valero-Mora, P.M & Ledesma, R.D. 2011. Using Interactive Graphics to Teach Multivariate Data Analysis to Psychology Students. *Journal of Statistics Education*, 19(1), www.amstat.org/publications/jse/v19n1/valero-mora.pdf.
- Wainer, H. 2011. The first step toward wisdom. *Chance*, 24(2), 60-61.
- Yu, C.H, Andrews, S, Winograd, D, Jannasch-Pennell, A, & DiGangi, S.A. 2002. Teaching Factor Analysis in Terms of Variable Space and Subject Space Using Multimedia Visualization. *Journal of Statistics Education*, Volume 10, Number 1 (2002)
- Zhou, L & Braun, W.J. 2010. Fun with the R Grid Package. *Journal of Statistics Education*, 18 (3), www.amstat.org/publications/jse/v18n3/zhou.pdf.