JAN JOURNAL OF ADVANCED NURSING

DISCUSSION PAPER

Testing and verifying nursing theory by confirmatory factor analysis

Maria Kääriäinen, Outi Kanste, Satu Elo, Tarja Pölkki, Jouko Miettunen & Helvi Kyngäs

Accepted for publication 27 November 2010

Correspondence to M. Kääriäinen:
e-mail: maria.kaariainen@oulu.fi

Maria Kääriäinen PhD RN
Senior Lecturer
Institute of Health Sciences, University of Oulu, Finland

Outi Kanste PhD RN
Project Manager
Oulu Arc Subregion, Finland

Satu Elo PhD RN
Senior Lecturer
Institute of Health Sciences, University of Oulu, Finland

Tarja Pölkki PhD RN
Adjunct Professor
Institute of Health Sciences, University of Oulu, Finland

Jouko Miettunen PhD
Adjunct Professor
Department of Psychiatry, Institute of Clinical Medicine, University of Oulu, Finland

Helvi Kyngäs PhD RN
Professor
Institute of Health Sciences, University of Oulu, Finland

KÄÄRIÄINEN M., KANSTE O., ELO S., PÖLKKI T., MIETTUNEN J. & KYNGÄS H. (2011) Testing and verifying nursing theory by confirmatory factor analysis. *Journal of Advanced Nursing* 67(5), 1163–1172. **doi: 10.1111/j.1365-2648.2010.05561.x**

Abstract

Aim. This paper presents a discussion of the use of confirmatory factor analysis to test nursing theory.

Background. Theory testing is an important phase in nursing theory development. Testing of theory is intended to give more information about concepts and their usefulness in nursing practice. Confirmatory factor analysis is commonly used in instrument development in nursing science studies, but also in theory testing. However, there has been little discussion of its use in theory testing in nursing science research.

Data sources. Multidisciplinary methodological and research publications from 1990 to 2009 were used.

Discussion. The aim of confirmatory factor analysis is to test nursing theory that has already been established, i.e. researchers have an *a priori* hypothesis based on theoretical knowledge or empirical indications. Analysis is represented as three phases: preparation, model testing and reporting the results. Preparation involves data screening and preliminary analyses. Model testing is divided into model specification, model identification, model estimation, model evaluation and model modification. The results are reported with standardized regression coefficients of the items related to the latent variables, squared multiple correlations (R^2) related to error terms and the model's goodness of fit indexes.

Implications for nursing. Testing of theory is intended to give more valid information about the concepts and their usefulness in nursing practice.

Conclusion. Confirmatory factor analysis is a good method to test the structure of theory, for example to test the concepts built by concept synthesis or analysis. Tested theories are needed to develop nursing science itself.

Keywords: confirmatory factor analysis, nursing theory, statistical analysis, theory testing

Introduction

In nursing science, the focus is on creating theories, but there has been less systematic testing of the theories (McKenna 1997, McEwen 2007c). Empirical theory testing is an important phase in nursing theory development. The empirical testing of theory is a systematic process, where theoretical statements are tested to indicate the structure of theory, for example the relationships between concepts in a real-life context, with the aid of research. Based on the results, statements can be modified and retested, if needed (Chinn & Kramer 1999, Meleis 2005).

Nursing theory defines the nature, structure, concepts and relationships between nursing concepts (Fawcett 1995, Meleis 2005). Nursing theories can be classified based on their type into descriptive, explanatory, predictive and directive (Kim 2000, Peterson & Bredow 2009). Through systematic and empirical testing, a theory can be developed from a descriptive theory to an explanatory, predictive and directive theory (Meleis 2005, McEwen 2007a, 2007b). A hypothetic-deductive approach is generally used in testing nursing theories. Hypotheses are formed about the statements, and the theory is verified empirically by testing the hypotheses (Fawcett & Downs 1996, McKenna 1997).

Empirical testing of theory is intended to give more information about concepts and their usefulness in nursing practice. In order to obtain evidence of usefulness, the empirical indicators for the theoretical concepts and statements describing the relationships between concepts need to be defined (Chinn & Kramer 1999, Hardy 2004, Meleis 2005, McEwen 2007d). In nursing science, theoretical concepts are often multidimensional, and they need to be defined through various forms of action. For example, adherence to health regimens may be defined using concepts such as medical care, responsibility, care planning and implementing care (Kyngäs 2000).

Confirmatory factor analysis (CFA) can be a useful application to test the structure and relationships among nursing theory concepts. It is mainly used for psychometric evaluation of instruments and construct validation, but also for identification of method effects and evaluation of factor invariance (Russell 2002, Brown 2006). According to Watson and Thompson (2006), CFA was rarely used in papers published in the Journal of Advanced Nursing from 1982 to 2004. However, its use in nursing science has increased in the 21st century. Information retrieval from the CINAHL and Ovid Medline(r) databases shows that in nursing journals, CFA was often used as a keyword, but less often in titles. The number of papers published in nursing journals from 1990 to 10th December 2009 with CFA in the title was 84 in CINAHL and 13 in Ovid Medline(r). Although the number of papers as a whole has risen since the 1990s, it is noteworthy that most of those (84%) for which CFA is given as a keyword were published in the 21st century (Table 1). However, there has been little discussion of the use and reporting practices of CFA in nursing science, which makes analysis challenging for researchers. The aim of this paper is to present a discussion of the use of confirmatory factor analysis to test nursing theory.

Table 1 Journal papers in CINAHL and Ovid Medline(r) databases using confirmatory factor analysis as a keyword and title in nursing journals 1990–2009

	CINAHL	Ovid Medline(r)
	N	*N*
In the title	84	13
As a keyword	679	150
Published in January 2000–10 December 2009	580	118
January–10 December 2009	106	13
2008	111	34
2007	69	10
2006	66	14
2005	64	10
2004	44	10
2003	39	9
2002	34	7
2001	30	8
2000	17	3
1990–1999	99	32

Background

Confirmatory factor analysis is understood as a special type or application of Structural Equation Modeling (SEM). SEM combines both factor and regression analyses. The idea of SEM is to allow the study of causal relationships between factors by using regression analysis (Hoyle 1999, MacCallum & Austin 2000, Brown 2006, Meyers *et al.* 2006). For example, factors can be built using factor analysis, and SEM can be used to study the causal relationships between factors.

Confirmatory factor analysis and exploratory factor analysis (EFA) belong to the 'family of factor analyses'. CFA differs from EFA because EFA is used to determine exploratory factor model without an *a priori* assumption of associations between variables. EFA explores the structure of correlation or covariance matrices (Meyers *et al.* 2006). The nature of EFA is data-driven (Stevens 2002, Meyers *et al.* 2006), and no hypothesis of the factor structure of data is needed to use it. This has also been criticized (Gorsuch 2003, Thompson 2004). CFA is based on hypotheses derived from theoretical understanding of a factor model or the structure of factors. That model or structure can be verified or discarded by analysing the empirical data with CFA (Maruyama 1998, Meyers *et al.* 2006).

Confirmatory factor analysis is used to test nursing theory that has already been established. Because of this, researchers have to have an *a priori* hypothesis based on theoretical knowledge or empirical indications. They also need knowledge about relationships between observed measures or indicators (e.g. items, scale scores) and latent variables or factors. In addition, they need to know how many factors the variables are intended to form. Researchers also have to know which variables are loaded on which factors and whether the factors are interrelated (Stevens 2002, Brown 2006). Because of this, in order to develop and test nursing theory by using CFA, exploratory factor analysis has to be conducted first.

Data sources

The data-based and manual literature searches were conducted across the disciplines of nursing, medicine, psychology and statistics. In searches in CINAHL and Medline, the following terms were used: confirmatory factor analysis, theory, testing and nursing. International publications were also searched for in PsycINFO, with the aid of the search terms confirmatory factor analysis, theory, testing. Only English-language publications were included. Methodological and research publications from 1980 to 2009 were used in this paper.

Discussion

Fundamentals of confirmatory factor analysis

The theoretical basis of CFA relates to fundamentals of SEM. SEM is based on correlation and covariance matrixes of variables. Variables that are expected to have relationships with other variables should have a stronger correlation than those that are not expected to have relationships (Byrne 2004, Munro 2005). Models are built by dependent and independent variables whose relationships are expected to be linear. Variables are expected to have normal distributions (Kline 2005, Meyers *et al.* 2006).

The starting points are data and their variables, whose interdependence is illustrated with a figure using residual and latent terms. Models have three typical features: geometrical form or symbol with text, one- or two-way arrows between the symbols, and coefficient attached to arrows. There are different kinds of geometrical forms used: circle or oval, square or rectangle and arrows (Raykov & Marcoulides 2000, Meyers *et al.* 2006) (Figure 1).

Phases of confirmatory factor analysis

Confirmatory factor analysis can be conducted, for example, with the aid of AMOS, LISREL, EQS, MPLUS or Proc Calis software. However, the software used is not always mentioned in publications (Russell 2002). There are differences between software packages, e.g. in the handling of categorical or ordinal (Likert) variables. For example, MPLUS handles

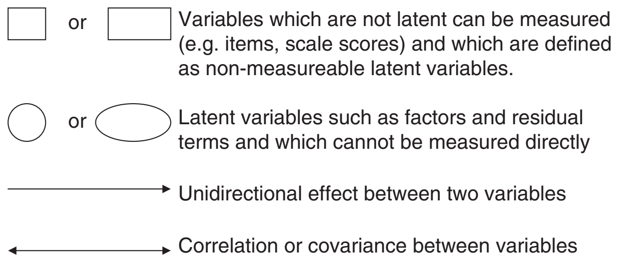

Figure 1 Typical geometrical forms or symbols used to draw a model (e.g. Raykov & Marcoulides 2000, Meyers *et al.* 2006).

them as categorical, calculating correlations accordingly. Analysis is represented as three phases: preparation, model testing and reporting the results (Figure 2).

Phase I: Preparation

The preparation phase starts with *data screening and preliminary analyses*. There are many factors contributing to whether a model is statistically identified or not, for example the characteristics of variables and whether the data are appropriate for CFA. There are different kinds of rules. For example, the sample size required for CFA is not clearly agreed upon in the literature. Sample size has also been pointed out to include inadequacies in reporting results (Schreiber *et al.* 2006). It has also been suggested that the sample size should be more than 3, 12 or as high as 15 times the number of variables (Stevens 2002), or five times the number of parameters. In addition, there should be variation in the data set (Bentler & Chou 1987). It is suggested that a good sample size might be not <200 (Loehlin 2004, Kline 2005). However, it has been indicated that stable factor models can be found with samples of 100 (Fabrigar *et al.* 1999) or with samples of 150 if at least 10 items load at 0·40 or higher (Guadagnoli & Velicer 1988).

Multivariate outliers should also be deleted and the criteria for deleting should be reported. In addition, the missing data should be screened for and the method reported. The most commonly used method is to delete cases with missing data by listwise deletion or available case analysis if data are missing at random (MAR) (Schaefer & Graham 2002, McKnight *et al.* 2007). Because of missing values, the data set can be smaller than was intended at the beginning of the process. On the other hand, missing data can be replaced before the analysis. With AMOS, analysis can be done even with data missing. The software supposes data to be occasionally missing, even if that is generally speaking not true (Arbuckle 2005, Meyers *et al.* 2006). There has been insufficient reporting of missing data in previous studies (Russell 2002, Schreiber *et al.* 2006, Jackson *et al.* 2009).,

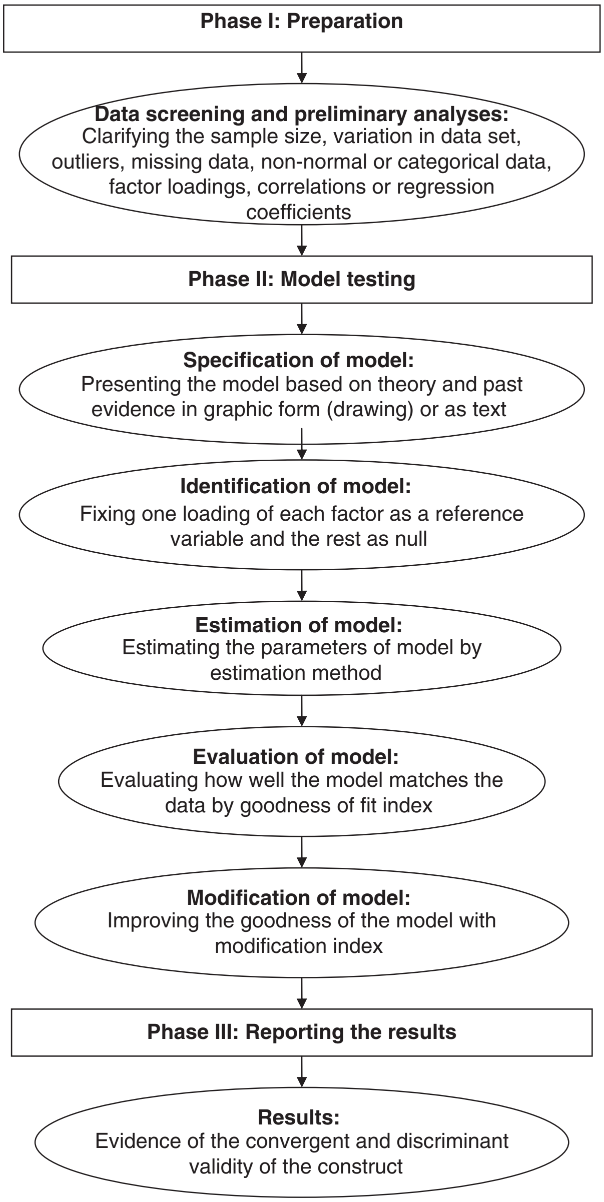

Figure 2 The phases of confirmatory factor analysis.

For example, instead Jenerette and Murdaugh (2008) evaluated all scales with more than 5% missing data with *t*-test to test for statistically significant differences between participants with complete data and those with missing data. They did not find such differences, so they assumed that data were missing at random.

Variables can be dummy or, for example, Likert-scale variables. The phenomenon behind these variables should be normally distributed. If continuous variables are used, normal distribution is not required. In nursing science research, the problem is often that distributions are not normal. For example, Jenerette and Murdaugh (2008) tested normality visually and with the Shapiro–Wilk test, and they excluded variables that did not meet the assumption of normality. However, screening for univariate or multivariate normality has often been said to lead to inadequate reporting of results (DiStefano & Hess 2005, Schreiber *et al.* 2006, Jackson *et al.* 2009).

The correlation or regression coefficient to factors should be examined. Researchers should choose to model only variables with strong loadings or those that have a strong correlation or regression coefficient to the factors. Nevertheless, there is no agreement on the value of high loading. For example, the following criteria for adequate loadings have been used: 0·30, 0·35 or 0·40. If a variable has strong cross-loadings on many factors it can be deleted (Tabachnick & Fidell 2001, Stevens 2002, Meyers *et al.* 2006). However, correlation matrix analyses involve problems in reporting results (MacCallum & Austin 2000, DiStefano & Hess 2005).

Phase II: Model testing

Model testing phase is divided into five main phases: (1) specification, (2) identification, (3) estimation, (4) evaluation and (5) modification of model.

Specification of model

Specification of model means that the model was chosen based on a research hypothesis. The model can be based on theory or a theoretical model developed by using EFA or CFA or the same data as to be used to specify the model. If the same data set is used, it can be divided into two parts: EFA is first performed using one part of the data, after which, based on the result of EFA, the other part of the data is used for CFA. However, it is recommended that the model to be tested is based on separate data (Byrne 2004, Meyers *et al.* 2006). One of the benefits of using separate data is that the results are more generalizable (e.g. see Firat *et al.* 2009).

In the model specification phase the model is presented. It can be drawn in graphic form or presented as text, for example with Amos *(Analysis of Moment Structures)* software. If the model is presented as text, the model or all variables (latent variable, measured variables, latent measuring errors) and all relationships between the variables are written using programme code. In the graphic model, the model or all variables and all relationships between the variables are drawn using software, and geometrical symbols such as rectangles, ovals, circles and one- and two-way arrows are used (Arbuckle 2005) (Figure 3).

The researcher has to know which variables are intended to measure latent variables in order to specify the model. All

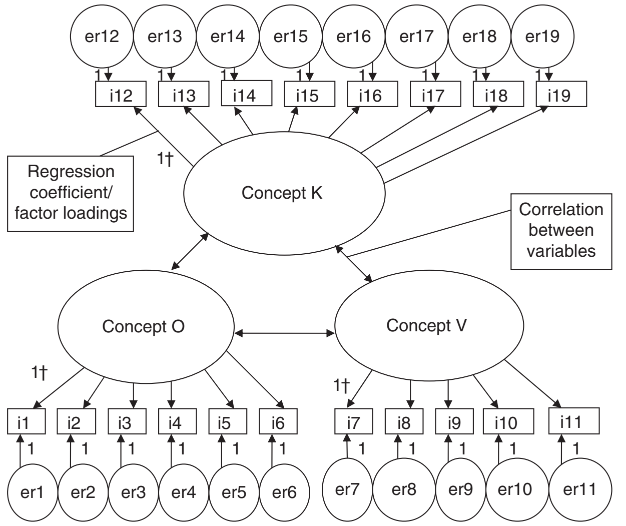

Figure 3 An example of drawing a model (rectangles = measured variables, ovals = latent variables, circles = error terms). †One loading is fixed to the value 1 in order to give the latent factor an interpretable scale.

variables (latent and measured) are dependent or independent variables. Arrows are drawn from independent variables and to dependent variables (Bentler & Weeks 1980). For example, in Figure 3 there are 22 independent variables: three latent variables (factors) and 19 error or unique terms. In addition, there are 19 dependent or measured variables (items). Dependent variables are thought to be indicators of underlying latent variables. In the example, it is also assumed that the model has a measurement error affecting the latent variable.

Identification of the model

After specification, the model has to be identified to estimate statistically the parameters of relationships (Bollen 1989). When identifying a model, one loading is fixed, usually to one, on each factor as a reference variable (Byrne 2004). The variable with the strongest loading on the latent variable can be chosen. The aim is identification of the model. The model is identified if degree of freedom (d.f.) >0. If the degree of freedom is 0, the model is just defined. This means that it fits the data arbitrarily and its appropriateness should be considered. If the degree of freedom is <0, the model is under-identified and not appropriate for analysis (Meyers *et al.* 2006).

Estimation of model

The parameters of a model (factor loadings and residual terms) can be assessed by an estimation method. The most widely used method for estimating CFAs is Maximum Likelihood (ML) (Meyers *et al.* 2006). It can be used when there are enough observations (>200) and when variables have multivariate normality (Loehlin 2004, Kline 2005). If multivariate normality is not met, the chi-square statistic can be an overestimate (Powell & Schafer 2001). On the other hand, minor divergence may be allowed by ML (Fan & Wang 1998). The purpose is to find estimates to unknown parameters of the model that fit with the values of the parameters in the population studied. Estimation of parameters is usually presented as standardized values (Kline 2005, Meyers *et al.* 2006). For example, Kanste *et al.* (2009) used ML and presented an estimation of parameters as standardized regression coefficients (β). However, the estimation method and procedure have said to involve deficiencies in reporting results (Russell 2002, DiStefano & Hess 2005, Schreiber *et al.* 2006, Jackson *et al.* 2009).

Evaluation of model

How well the model matches the data is illustrated by the goodness of fit index (Maruyama 1998, Byrne 2004, Meyers *et al.* 2006). There are more than 20 parameters (Klem 2000) and there is no agreement as to which parameters are the best to use (McDonald & Marsh 1990, Maruyama 1998). According to Jaccard and Wan (1996), at least three parameters should be used. Thompson (2004) and Kline (2005) suggest that at least the chi-square test, Comparative Fit Index (CFI) and RMSEA should be reported.

There is also disagreement on how to categorize the parameters (Jaccard & Wan 1996, Hair *et al.* 1998, Maruyama 1998, Hu & Bentler 1999, Tabachnick & Fidell 2001, Arbuckle 2005). For example, Hu and Bentler (1999) divide fit indexes into two categories: absolute and relative. Table 2 shows absolute and relative fit indexes and their limits.

Absolute parameters indicate how well the hypothetical relationships between the variables match the observed

Table 2 General goodness of fit index values and their limits (limits vary according to different sources)

Absolute		Relative	
Fit indexes	Cutpoint	Fit indexes	Cutpoint
Chi-square	$>0{\cdot}05$	CFI	$>0{\cdot}90/0{\cdot}95$
GFI	$>0{\cdot}90$	NFI	$>0{\cdot}90/0{\cdot}95$
AGFI	$>0{\cdot}90$	IFI	$>0{\cdot}90/0{\cdot}95$
RMR	$<0{\cdot}05/0{\cdot}06$		
RMSEA	$<0{\cdot}06/0{\cdot}08$		

GFI, Goodness-of-fit Index; AGFI, Adjusted Goodness-of-fit Index; RMR, Root Mean Square Residual; RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index; NFI, Normed Fit Index; IFI, Incremental Fit Index.

relationships – in other words, how well the correlations or covariances of the theoretical model match the observed correlations or covariances (Meyers *et al.* 2006). In the following section, the most commonly used parameters are presented.

The chi-square test is used to test the hypothesis. The null hypothesis is intended to estimate how the model fits the data. The *P*-value should not be statistically significant ($>0{\cdot}05$) and the chi-square test statistic divided by the d.f. should be <3. If the sample size is small (n = 75–200), the chi-square test is an adequate measure of fit. However, the chi-square test is very sensitive and can easily reject the model ($P < 0{\cdot}05$) if sample size is large ($n >$ 500–600). In some cases, it can be unclear whether the reason for statistically significant results is sample size or the fact that the model does not fit the data. This has led to development of other parameters to study the model's suitability for the data (Hoyle 1999, Hu & Bentler 1999, Stevens 2002, Byrne 2004, Kline 2005, Meyers *et al.* 2006).

Goodness-of-fit Index (GFI) can be used to test the general ability to compare the theoretical model with the observed model (Hair *et al.* 1998, Hu & Bentler 1999). GFI measures the relative amount of covariance and variance by comparing the theoretical model to the observed model (Meyers *et al.* 2006). This parameter can be assumed to be similar to the square of multiple correlations (R^2) in regression analysis with many variables (Kline 2005). The closer the GFI is to 1·00, the better the model fits the data. GFI also takes into account degree of freedom. GFI and Adjusted Goodness-of-fit Index *(*AGFI) are less sensitive to sample size than the chi-square test. The recommended value of GFI and AGFI is $>0{\cdot}90$ (Meyers *et al.* 2006). However, this value has been criticized as being too stringent for testing developing theories (McRae *et al.* 1996, Vassend & Skrondal 1997, Raykov 1998), and so GFI and AGFI are not always the best indicators of model fit (Floyd & Widaman 1995).

Root Mean Square Residua (RMR) measures residual covariance and variance or the amount of data which the model does not explain. The less difference there is between the theoretical and observed variance, the better the model and the smaller the RMR value. The value depends on the magnitude of measured variables. A good value is $<0{\cdot}05$ or $<0{\cdot}06$. Root Mean Square Error of Approximation (RMSEA) evaluates the model's general adequacy to compare a theoretical model with a perfect (or saturated) model (Meyers *et al.* 2006). It has been found to be one of the most informative parameters (Byrne 2004). Browne and Cudeck (1993) present that values $<0{\cdot}05$ are excellent. It has also been stated that the limit of RMSEA is $<0{\cdot}06$, which is good, but a value $<0{\cdot}08$ (Hu & Bentler 1999, Loehlin 2004, Meyers *et al.* 2006) or $<0{\cdot}10$ can also be acceptable (MacCallum *et al.* 1996). Values above 0·10 indicate poor fit (Browne & Cudeck 1993).

Relative parameters are used to test the adequacy of a model by comparing a theoretical model to a null model. The most common relative parameters are CFI, Normed Fit Index (NFI) and Incremental Fit Index (IFI). CFI quantifies the amount of variation and covariation accounted for by the proposed model by comparing its fit to the fit of a null model of uncorrelated variables (Hu & Bentler 1999). Parameters have values between 0 and 1. The recommended limit value varies. For example, according to Bentler (1990, 1992) and Hu and Bentler (1999), CFI values $>0{\cdot}90$ are acceptable and values $>0{\cdot}95$ are excellent. According to Knight *et al.* (1994), $>0{\cdot}90$ is good, 0·80–0·89 acceptable, 0·60–0·79 poor and $<0{\cdot}60$ very poor. CFI is more reliable than NFI with small sample sizes.

The NFI is used to test the model's general adequacy, comparing a theoretical model to a null model. It can also be used to evaluate the effect of large sample size on the chi-square test result (Hair *et al.* 1998, Hu & Bentler 1999, Byrne 2004). Like NFI, IFI not only tests the general adequacy of a model, but it also detects the degree of freedom of a theoretical model. Values $>0{\cdot}90$ or $>0{\cdot}95$ have been presented as limit values for NFI and IFI (Meyers *et al.* 2006).

Modification of model

If the model is not found to be acceptable, it can be modified (Byrne 2004, Meyers *et al.* 2006). The goodness of the model can be improved with the modification index. This gives information about the relationships of variables in the model which should be added or changed to build a model with better fit to the data. The greater the model's modification index, the greater the reducing effect that the release of a variable for estimation has on the model's χ^2 value. There are different parameters for comparing the models, such as Akaike's Information Criteria (AIC), Consistent AIC (CAIC) and Bayes Information Criteria (BIC). The smaller these parameters, the better the model fits the data.

The model can be specified again by the modification index. In order to be identified, a model can be modified, e.g. it may need more limitations or fewer estimated values. For example, a relationship can be added from one factor to some other factor's indicator. In this case, the indicator reflects the effects of two factors. On the other hand, correlation of the error value of two variables can be allowed, which can improve the model's parameters. The more different dependencies are added to the model, the better, the software fits the data to the model. Adding dependencies gives more freedom to the software to add degrees of freedom. It is obviously crucial that the changes should be minor and in accordance with theory (MacCallum 1995, Byrne 2004, Meyers *et al.* 2006).

Phase III: reporting the results

The results are reported in graphical form with standardized regression coefficients of the items related to the concept O, K and V, squared multiple correlations (R^2) related to error terms and the model's goodness of fit indexes. Figure 4 shows that the chi-square test discarded the unmodified exemplary model as inadequate (*P* value was 0·001), but the NFI evaluating the effect of sample size in the chi-square test indicated the model to be suited to the data, as values 0·96 are considered a measure of good suitability.

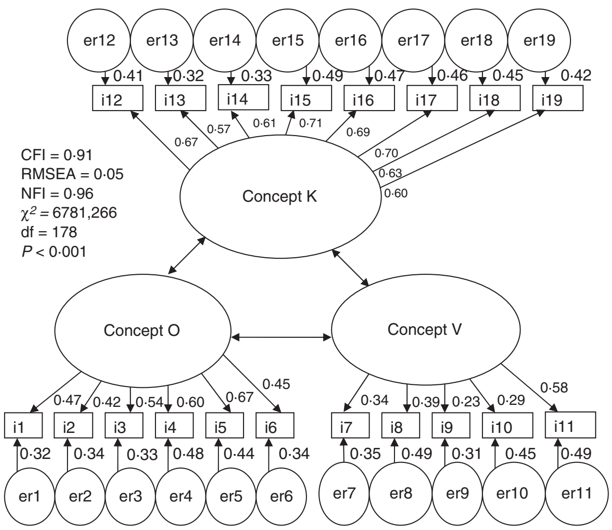

Figure 4 An example of reporting results with standardized regression coefficients of the items related to the concepts O, K and V, squared multiple correlations (R^2) related to error terms and some goodness of fit indexes (CFI, RMSEA, NFI). Rectangles = measured variables, ovals = latent variables, circles = error terms. CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation; NFI, Normed Fit Index.

The goodness of fit indexes may also be presented as a table. The relevant indices of different factor models and the changes after modification can be illustrated in table form (See Table 3).

Conclusion

The use of confirmatory factor analysis is recommended if the aim is to test an earlier theory, validate its constructs or test group differences supported by factorial invariance. CFA studies and strengthens theoretical structures in the data. It is a good method to test the structure of theory, for example to test the concepts built by concept synthesis or analysis. Tested theories are needed to develop nursing science itself. In nursing science, there are middle range theories that have been created by inductive methods and have not been empirically tested. By using inductive methods it is impossible to verify the structure of concepts and indicate the connections between concepts. In this way, descriptive theories are created. However, although there is a need for descriptive theories, exploratory and predictive theories whose usefulness has been tested are needed to explain and predict phenomena.

The use of CFA to test nursing theory has increased in recent years. However, there are some limitations to its use. There is an assumption that the sample size must be large enough. In addition, the variables have to have normal distributions and there has to be variation in the data. Furthermore, measured variables should not be too similar because this can cause correlations of measurement errors. In such a case the model does not work. When choosing a model, the researcher has to know which variables are measuring each latent variable. It is possible that the researcher has a misunderstanding about the model and specifies (selects) a poor model.

With CFA it is possible to modify the model based on parameters describing the fitness of data. However, having a large number of parameters poses problems in choosing parameters. It is recommended that fitness of data be

Table 3 An example of the goodness of fit index values of different factor models

Model	χ^{2*}	d.f.	GFI	AGFI	NFI	CFI	RMR	RMSEA
1-factor model	741	147	0·85	0·80	0·86	0·86	0·17	0·12
2-factor model	647	145	0·87	0·84	0·93	0·93	0·12	0·08
3-factor model	654	147	0·90	0·89	0·94	0·94	0·06	0·07
3-factor modified model	666	148	0·92	0·90	0·96	0·97	0·05	0·06

GFI, Goodness-of-fit Index; AGFI, Adjusted Goodness-of-fit Index; NFI, Normed Fit Index; CFI, Comparative Fit Index; RMR, Root Mean Square Residual; RMSEA, Root Mean Square Error of Approximation.
$^*P < 0{\cdot}001$.

What is already known about this topic

- The accurate use of nursing theory requires that the structure, concepts and associations between concepts of the theory are tested and verified by statistical methods.
- Confirmatory factor analysis is commonly used in instrument development, but not as often in theory testing in nursing science studies.
- There has been little discussion of the use of confirmatory factor analysis in nursing theory testing and there are no unambiguous guidelines for confirmatory factor analysis reporting practices, which makes analysis challenging for researchers.

What this paper adds

- Confirmatory factor analysis is extremely well suited for analysing the structure, concepts and associations between concepts in nursing theories.
- The use of confirmatory factor analysis is recommended if the aim is to test and verify an earlier theory, its concepts and structure and to test group differences supported by factorial invariance to generalize theory.

Implications for practice and/or policy

- Tested theories are needed to develop nursing science itself.
- Testing of theory is intended to give more valid information about the concepts and their usefulness in nursing practice.

evaluated with many different parameters, but there is no agreement as to which parameters should be reported.

The problem with model modification, for example adding parameters, is that it can result in no identified model or in intended relationships that do not have any theoretical basis. If error terms are allowed to correlate freely, the software suggests that data be corrected to fit the model or *vice versa*. Because of this, each correlation of this kind should have a special explanation. Although empirical parameters are important for theory development and model modification, the researcher's knowledge and understanding of the content of theory or concept are equally important. Because of this, attention should be given to the meaning of content when adding statistical numbers.

Modification of a model can also lead to having more than one good model that fits the data well. That is why it is important to bear in mind that finding a model does not mean that it is the one and only optimal model. Confirmatory factor analysis is a very data-specific method. For example, the results for the structure of a concept are possible only in theory without their cross-validation in an independent sample.

Conflict of interest

No conflict of interest has been declared by the authors.

Funding

This study was financially supported by a grant from the Foundation of Nurses' Education.

Author contributions

MK was responsible for study conception and design. MK, OK, SE and JM performed the data collection. MK performed data analysis. MK, OK and HK were responsible for the drafting of the manuscript. OK, SE, TP, JM and HK made critical revisions to the paper for important intellectual content. JM provided statistical expertise. MK and HK obtained funding. HK supervised the study.

References

Arbuckle J.L. (2005) *Amos™ 6.0 User's Guide*. Amos Development Corporation, Spring House, PA.

Bentler P.M. (1990) Comparative fit indices in structural models. *Psychological Bulletin* **107**, 238–246.

Bentler P.M. (1992) On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin* **112**(3), 400–404.

Bentler P.M. & Chou C.P. (1987) Practical issues in structural modeling. *Sociological Methods & Research* **16**, 78–117.

Bentler P.M. & Weeks D.G. (1980) Linear structural equations with latent variables. *Psychometrika* **45**, 289–308.

Bollen K.A. (1989) *Structural Equation with Latent Variables*. John Wiley & Sons, New York.

Brown T.A. (2006) *Confirmatory Factor Analysis for Applied Research*. Guilford, New York.

Browne M.W. & Cudeck R. (1993) Alternative ways of assessing model fit. In *Testing Structural Equation Models* (Bollen K.A. & Long J.S., eds), Sage Publications, Thousand Oaks, CA, pp. 136–162.

Byrne B.M. (2004) *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Erlbaum, Mahwah, NJ.

Chinn P.L. & Kramer M.K. (1999) *Theory and Nursing: A Systematic Approach*. Mosby Year Book, St. Louis.

DiStefano C. & Hess B. (2005) Using confirmatory factor analysis for construct validation: an empirical review. *Journal of Psychoeducational Assessment* **23**, 225–241.

Fabrigar L.R., Wegener D.T., MacCallum R.C. & Strahan E.J. (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* **4**, 272–299.

Fan X. & Wang L. (1998) Effects of potential confounding factors on fit indices and parameter estimates for true and misspecified SEM models. *Educational and Psychological Measurement* **58**, 701–735.

Fawcett J. (1995) *Analysis and Evaluation of Nursing Theories*. Davis, Philadelphia.

Fawcett J. & Downs F.S. (1996) *The Relationship of Theory and Research*. Davis, Philadelphia.

Firat M.F., zen Kulakac Ö., Öncel S. & Akcan A. (2009) Menstrual Attitude Questionnaire: confirmatory and exploratory factor analysis with Turkish samples. *Journal of Advanced Nursing* **65**(3), 652–662.

Floyd F.J. & Widaman K.F. (1995) Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment* **7**, 286–299.

Gorsuch R.L. (2003) Factor analysis. In *Handbook of Psychology: Vol 2. Research Methods in Psychology* (Schinka J.A. & Velicer W.F., eds), Wiley, Hoboken, NJ, pp. 143–164.

Guadagnoli E. & Velicer W.F. (1988) Relation to sample size to the stability of component patterns. *Psychological Bulletin* **103**, 265–275.

Hair J.F., Anderson R.E., Tatham R.L. & Black W.C. (1998) *Multivariate Data Analysis*. Prentice Hall, Englewood Cliffs, NJ.

Hardy M.E. (2004) Theories: components, development, evaluation. In *Perspectives on Nursing Theory*, 4th edn (Reed P.G., Shearer N.C. & Nicoll L.H., eds), Williams & Wilkins, Lippincott, pp. 75–86.

Hoyle R.H. (1999) The structural equation modeling approach: basic concepts and fundamental issues. In *Structural Equation Modeling: Concepts, Issues and Applications* (Hoyle R.H., ed), Sage, Thousand Oaks, CA, pp. 1–15.

Hu L.T. & Bentler P.M. (1999) Evaluating model fit. In *Structural Equation Modeling: Concepts, Issues and Applications* (Hoyle R.H., ed.), Sage, Thousand Oaks, CA, pp. 76–99.

Jaccard J. & Wan C.K. (1996) *LISREL Approaches to Interaction Effects in Multiple Regression*. Sage, Thousand Oaks.

Jackson D.L., Gillaspy J.A. & Purc-Stephenson R. (2009) Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological Methods* **14**(1), 6–23.

Jenerette C.M. & Murdaugh C. (2008) Testing the theory of self-care management for sickle cell disease. *Research in Nursing & Health* **31**, 355–369.

Kanste O., Kääriäinen M. & Kyngäs H. (2009) Statistical testing of the full-range leadership theory in nursing. *Scandinavian Journal of Caring Sciences* **23**(4), 775–782.

Kim H.S. (2000) *The Nature of Theoretical Thinking in Nursing*. Springer Publishing Company, New York.

Klem L. (2000) Structural equation modeling. In *Reading and Understanding More Multivariate Statistics* (Grimm L.G. & Yarnold P.R., eds), American Psychological Association, Washington, DC, pp. 227–260.

Kline R.B. (2005) *Principles and Practice of Structural Equation Modeling*, 2nd edn. Guilford Press, New York.

Knight G.P., Virdin L.M., Ocampo K.A. & Roosa M. (1994) An examination of the cross-ethnic equivalence of measures of negative life events and the mental health among Hispanic and Anglo American children. *American Journal of Community Psychology* **22**, 767–783.

Kyngäs H. (2000) The theoretical model of adherence of young diabetics. *Journal of Clinical Nursing* **8**, 73–80.

Loehlin J.C. (2004) *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*, 4th edn. Erlbaum, Mahwah, NJ.

MacCallum R.C. (1995) Model specification: procedures, strategies and related issues. In *Structural Equation Modeling: Concepts, Issues and Applications* (Hoyle R.H., ed.), Sage, Thousand Oaks, CA, pp. 16–36.

MacCallum R.C. & Austin J.T. (2000) Applications of structural equation modeling in psychological research. *Annual Review of Psychology* **51**, 201–226.

MacCallum R.C., Roznowski M. & Necowitz L.B. (1992) Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin* **111**, 490–504.

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.

Maruyama G.M. (1998) *Basics of Structural Equation Modeling*. Sage Publications, Thousand Oaks.

McDonald R.P. & Marsh H.W. (1990) Choosing a multivariate model: noncentrality and goodness of fit. *Psychological Bulletin* **107**, 247–255.

McEwen M. (2007a) Overview of theory in nursing. In *Theoretical Basis for Nursing* (McEwen M. & Wills E.M., eds), Lippincott Williams & Wilkins, Philadelphia, pp. 24–49.

McEwen M. (2007b) Theory development: structuring conceptual relationship in nursing. In *Theoretical Basis for Nursing* (McEwen M. & Wills E.M., eds), Lippincott Williams & Wilkins, Philadelphia, pp. 73–94.

McEwen M. (2007c) Application of theory in nursing research. In *Theoretical Basis for Nursing* (McEwen M. & Wills E.M., eds), Lippincott Williams & Wilkins, Philadelphia, pp. 434–456.

McEwen M. (2007d) Theory analysis and evaluation. In *Theoretical Basis for Nursing* (McEwen M. & Wills E.M., eds), Lippincott Williams & Wilkins, Philadelphia, pp. 95–114.

McKenna H.P. (1997) Theory and research: a linkage to benefit practice. *International Journal of Nursing Studies* **34**, 431–437.

McKnight P.E., McKnight K.M., Sidani S. & Figueredo A.J. (2007) *Missing Data: A Gentle Introduction*. Guilford Press, New York.

McRae R.R., Zonderman A.B., Costa P.T., Bond M.H. & Paunonen S.V. (1996) Evaluating replicability of factors in the revised NEO personality inventory: confirmatory factor analysis versus procrustean rotation. *Journal of Personality and Social Psychology* **70**, 552–566.

Meleis A.I. (2005) *Theoretical Nursing: Development & Progress*, 3rd edn. J.B. Lippincott Company, Philadelphia.

Meyers L.S., Gamst G. & Guarino A.J. (2006) *Applied Multivariate Research. Design and Interpretation*. Sage Publications, Thousand Oaks, CA.

Munro B.H. (2005) *Statistical Methods for Health Care Research*. Lippincott Williams & Wilkins Company, Philadelphia.

Peterson S.J. & Bredow T.S. (2009) *Middle-Range Theories. Application to Nursing Research*. Lippincott Williams & Wilkins, Philadelphia.

Powell D.A. & Schafer W.D. (2001) The robustness of the likelihood ratio chi-square test for structural equation models: a meta-analysis. *Journal of Educational and Behavioral Statistics* **26**, 105–132.

Raykov T. (1998) On the use of confirmatory factor analysis in personality research. *Personality and Individual Differences* **24**, 291–293.

Raykov T. & Marcoulides G.A. (2000) *A First Course in Structural Equation Modeling*. Erlbaum, Mahwah, NJ.

Russell D.W. (2002) In search of underlying dimensions: the use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin* **28**(12), 1629–1646.

Schaefer J.L. & Graham J.W. (2002) Missing data: our view of the state of the art. *Psychological Methods* **7**, 147–177.

Schreiber J.B., Stage F.K., King J., Nora A. & Barlow E.A. (2006) Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Educational Research* **99**, 323–337.

Stevens J. (2002) *Applied Multivariate Statistics for the Social Sciences*, 4th edn. Lawrence Erlbaum Associates, Erlbaum, Mahwah, NJ.

Tabachnick B.G. & Fidell L.S. (2001) *Using Multivariate Statistics*, 4th edn. Allyn & Bacon, Needham Heights, MA.

Thompson B. (2004) *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. American Psychological Association, Washington, DC.

Vassend O. & Skrondal A. (1997) Validation of the NEO personality inventory and the three-factor model: can findings from exploratory and confirmatory factor analysis be reconciled? *European Journal of Personality* **11**, 147–166.

Watson R. & Thompson D.R. (2006) Use of factor analysis in Journal of Advanced Nursing: literature review. *Journal of Advanced Nursing* **55**(3), 330–341.