Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 69 (2012) 1282 – 1288

Procedia
Social and Behavioral Sciences

International Conference on Education and Educational Psychology (ICEEPSY 2012)

# Exploratory Structural Equation Modelling (ESEM): application to the SET-37 questionnaire for students' evaluation of teaching

Pieter Spooren*[a], Dimitri Mortelmans[a], Francis Van Loon[a]

*University of Antwerp, Sint Jacobstraat 2, Antwerp B-2000, Belgium*

**Abstract**

This study contains a re-validation of the SET37-questionnaire for student evaluations of teaching (Mortelmans & Spooren 2009), using both confirmatory factor analysis and exploratory structural equation modelling (ESEM, Asparouhov & Muthén 2009), that integrates both exploratory and confirmatory factor analyses. In sum, the results provide strong evidence for the relevance of the questionnaire as CFA- and ESEM-models provide a good fit to the data. Although, as expected, item factor loadings and correlations between the twelve factors in the instrument are substantially lower in the ESEM-results. CFA-models which require zero cross-loadings might overvalue factor correlations when testing SET-instruments. The implications for future validation strategies in SET are discussed.



Selection and/or peer-review under responsibility of Dr. Zafer Bekirogullari of Cognitive – Counselling, Research & Conference Services C-crcs.

*Keywords:* students' evaluation of teaching; teacher evaluation; validity studies; higher education; teaching quality

## 1. Introduction

Nowadays, students' evaluations of teaching (SET) are used as a measure of teaching performance in almost every institution for higher education throughout the world. Universities and university colleges have developed more or less complex procedures and instruments to collect, analyze and interpret these data as the dominant and sometimes sole indicator of teaching quality. This widespread use has much to do with the (apparent) ease of collecting the data and presenting and interpreting the results. In most institutions, SET are used for both formative (i.e., students' feedback for the improvement of teaching) and summative (i.e, mapping teaching competence for administrative decision-making and institutional audits) reasons. This double use makes SET very delicate. On the one hand, most teachers are convinced of the usefulness of SET as an instrument for feedback on their teaching (Richardson 2005). As a consequence, many instructors gratefully make use of student evaluations for the improvement of their subsequent teaching. On the other hand, it is argued that nowadays the principal purpose of SET lies in its use as a measure for quality monitoring and administrative policy-making and mapping whether or not teachers reach a certain required standard in their teaching practice (Penny & Coe 2004). It is clear that SET-instruments should be tested and re-tested extensively on their psychometric value, especially when they are used

1877-0428 
Selection and/or peer-review under responsibility of Dr. Zafer Bekirogullari of Cognitive – Counselling, Research & Conference Services C-crcs.
doi:10.1016/j.sbspro.2012.12.063

for administrative decision-making. Still, many instruments remain to be *ad hoc* instruments that were not tested at all (Richardson 2005).

## 2. Objectives

In this contribution, we report on a re-validation procedure of the SET37 questionnaire designed at the University of Antwerp (Belgium). This instrument consists of twelve dimensions that are measured by means of 37 Likert Items and is based on both educational theory and empirical testing (exploratory factor analysis –EFA– and confirmatory factor analyses –CFA) (Spooren, Mortelmans & Denekens 2007). A re-validation procedure using confirmatory factor analysis provided strong evidence for the established dimensional structure (Mortelmans & Spooren 2009). However, Marsh et al. (2009) and Asparouhov & Muthén (2009) argued that, especially in psychological instruments that mostly have a well-defined EFA structure, CFA approaches do not represent such a structure appropriately. These measurement instruments often have small cross loadings that are well motivated by the theory (i.e., an item that measures a teacher's presentation skills might have small loadings on other dimensions concerning effective teaching in a SET-instrument as well), but are fixed to zero in a CFA approach. Model modifications to compensate for these misspecifications when searching for a well-fitting model often lead to distorted factors with overestimated factor correlations (Marsh et al. 2009). The new exploratory structural equation modelling methods (ESEM) therefore integrate the advantages of EFA (using factor loading matrix rotations) and CFA (access to all usual SEM parameters).

This note therefore contains a re-validation of the SET37 questionnaire by using both a CFA-approach and the more recent ESEM-approach (using an oblique rotation) on the same dataset. This allows a comparison of the outcomes of both approaches concerning the dimensional structure in the instrument.

## 3. Results

**Instrument.** For this re-validation study, we used the SET37-questionnaire which represents 12 quasi-balanced scales. All 37 items are measured on a six-point scale and were used in an online SET-procedure (for example items, see Appendix A).

**Participants.** SET were administered during the fall semester of the 2008-2009 academic year. A total of 2837 evaluation forms completed by students enrolled in 75 courses were analysed in the present study. Evaluations were completed in various educational programs by both undergraduate and graduate students.

**Re-validation of the SET37: CFA versus ESEM.** The goodness of fit indices of both analyses (CFA and ESEM) are presented in Table 1. Item loadings and factor correlations of the CFA model are presented in Table 2, the detailed parameters of the ESEM model are provided in Table 3.

Table 1: Summary of Goodness of Fit Statistics for Total Group CFA and ESEM Models

| *Model* | *$X^2$/df* | *NFParm* | *CFI* | *TLI* | *RMSEA* | *CI RMSEA* | *AIC* | *BIC* | *corBIC* |
|---|---|---|---|---|---|---|---|---|---|
| CFA | 4175.581/563 | 177 | .939 | .928 | .048 | .046 - .049 | 267914.065 | 268967.304 | 268404.913 |
| ESEM | 1523.499/288 | 452 | .979 | .952 | .039 | .037 - .041 | 265811.983 | 268501.610 | 267065.448 |

*Note*. CFI = comparative fit index; TLI = Tucker-Lewis Index; NFParm = number of free parameters; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; corBIC = sample-size adjusted BIC; RMSEA = root mean square error of approximation; CI RMSEA = 90 percent confidence interval RMSEA.

Both approaches provide an excellent fit to the data, although the $\chi 2$ tests of exact fit are significant whereas the objective is to achieve a nonsignificant p-value (< .05). However, Hatcher (1994) indicates that a significant $\chi 2$ does not make a confirmatory factor analysis model inadequate. The fit indices of the ESEM model suggest a better fit than the corresponding CFA model (Table 1).

Table 2: ICM-CFA Solution – Twelve ESEM Factors Based on Responses to 37 Items

| *Factors* | *Factor loadings* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Items* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* |
| Clarity of objectives (F1) | | | | | | | | | | | | |
| CO1 | **.83** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| CO2 | **.86** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| CO3 | **.80** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| Value subject-matter (F2) | | | | | | | | | | | | |
| VSM1 | .00 | **.99** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| VSM2 | .00 | **.99** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| VSM3 | .00 | **.89** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| Build-up subject-matter (F3) | | | | | | | | | | | | |
| BSM1 | .00 | .00 | **.75** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| BSM2 | .00 | .00 | **.59** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| BSM3 | .00 | .00 | **.78** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| Presentation skills (F4) | | | | | | | | | | | | |
| PS1 | .00 | .00 | .00 | **.98** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| PS2 | .00 | .00 | .00 | **.92** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| PS3 | .00 | .00 | .00 | **.92** | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| Organization course – learning (F5) | | | | | | | | | | | | |
| OC1 | .00 | .00 | .00 | .00 | **.96** | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| OC2 | .00 | .00 | .00 | .00 | **.90** | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| OC3 | .00 | .00 | .00 | .00 | **.79** | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| Course materials (F6) | | | | | | | | | | | | |
| CM1 | .00 | .00 | .00 | .00 | .00 | **.99** | .00 | .00 | .00 | .00 | .00 | .00 |
| CM2 | .00 | .00 | .00 | .00 | .00 | **.99** | .00 | .00 | .00 | .00 | .00 | .00 |
| CM3 | .00 | .00 | .00 | .00 | .00 | **.86** | .00 | .00 | .00 | .00 | .00 | .00 |
| CM4 | .00 | .00 | .00 | .00 | .00 | **.99** | .00 | .00 | .00 | .00 | .00 | .00 |
| Course difficulty (F7) | | | | | | | | | | | | |
| CD1 | .00 | .00 | .00 | .00 | .00 | .00 | **.87** | .00 | .00 | .00 | .00 | .00 |
| CD2 | .00 | .00 | .00 | .00 | .00 | .00 | **.84** | .00 | .00 | .00 | .00 | .00 |
| CD3 | .00 | .00 | .00 | .00 | .00 | .00 | **.87** | .00 | .00 | .00 | .00 | .00 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Help teacher during learning process (F8) | | | | | | | | | | | | |
| HT1 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.71** | .00 | .00 | .00 | .00 |
| HT2 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.71** | .00 | .00 | .00 | .00 |
| HT3 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.77** | .00 | .00 | .00 | .00 |
| Authenticity of the examination(s) (F9) | | | | | | | | | | | | |
| AE1 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.92** | .00 | .00 | .00 |
| AE2 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.99** | .00 | .00 | .00 |
| AE3 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.99** | .00 | .00 | .00 |
| Linking-up with foreknowledge (F10) | | | | | | | | | | | | |
| FK1 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.80** | .00 | .00 |
| FK2 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.94** | .00 | .00 |
| FK3 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.99** | .00 | .00 |
| Content validity examination (F11) | | | | | | | | | | | | |
| CVE1 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.83** | .00 |
| CVE2 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.89** | .00 |
| CVE3 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.89** | .00 |
| Formative evaluation(s) (F12) | | | | | | | | | | | | |
| FE1 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.96** |
| FE2 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.99** |
| FE3 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | **.97** |

*Factor Correlations*

| | *F1* | *F2* | *F3* | *F4* | *F5* | *F6* | *F7* | *F8* | *F9* | *F10* | *F11* | *F12* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 1.00 | | | | | | | | | | | |
| F2 | .52 | 1.00 | | | | | | | | | | |
| F3 | .65 | .68 | 1.00 | | | | | | | | | |
| F4 | .65 | .55 | .62 | 1.00 | | | | | | | | |
| F5 | .47 | .51 | .52 | .52 | 1.00 | | | | | | | |
| F6 | .66 | .46 | .63 | .71 | .40 | 1.00 | | | | | | |
| F7 | .57 | .50 | .53 | .52 | .35 | .48 | 1.00 | | | | | |
| F8 | .67 | .46 | .61 | .67 | .56 | .51 | .54 | 1.00 | | | | |
| F9 | .31 | .33 | .38 | .30 | .58 | .24 | .11 | .34 | 1.00 | | | |
| F10 | .40 | .48 | .45 | .33 | .32 | .30 | .61 | .31 | .09 | 1.00 | | |
| F11 | .56 | .47 | .54 | .41 | .43 | .41 | .50 | .49 | .36 | .35 | 1.00 | |
| F12 | .55 | .32 | .38 | .37 | .59 | .39 | .29 | .55 | .42 | .22 | .34 | 1.00 |

*Note*. ICM-CFA = independent clusters model-confirmatory factor analysis. This model assumed an independent cluster structure in which each of the SET37-items was allowed to load on only one latent factor and cross-loadings were not allowed. As all nontarget loadings are constrained to zero, only the target loading relating each item to its priori factor is presented. All parameter estimates are completely standardized. N = 2837 evaluations for 75 courses.

The factor loadings in the CFA-solution are very high and significant (Table 2), whereas the nontarget loadings are necessarily set to zero. The factor correlations are high as well, as they exceed .50 several times. This might be interpreted as a threat to the discriminant validity of the various dimensions in the instrument as it looks as they measure the same construct.

Table 3: ESEM Solution – Twelve ESEM Factors Based on Responses to 37 Items

| *Factors* | *Factor Loadings* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Items* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* |
| Clarity of objectives | | | | | | | | | | | | |
| CO1 | **.699** | .042 | .036 | .104 | .037 | -.052 | .037 | .033 | .039 | .036 | .009 | .056 |
| CO2 | **.496** | .070 | .039 | .019 | .118 | .007 | .056 | .031 | .012 | .181 | .052 | .120 |
| CO3 | **.551** | -.002 | .111 | .003 | .044 | .098 | .069 | -.022 | .059 | .055 | .107 | .027 |
| Value subject-matter | | | | | | | | | | | | |
| VSM1 | .079 | **.784** | .053 | .081 | -.004 | .017 | .046 | .025 | .027 | -.012 | .010 | .019 |
| VSM2 | -.020 | **.665** | .133 | .035 | .036 | .025 | .063 | .009 | .044 | .074 | .040 | .011 |
| VSM3 | .030 | **.516** | .046 | .011 | .024 | .124 | .015 | .043 | .124 | .007 | .085 | -.001 |
| Build-up subject-matter | | | | | | | | | | | | |
| BSM1 | .135 | .082 | **.584** | .089 | .028 | .066 | -.005 | .016 | .028 | .004 | .034 | -.007 |
| BSM2 | -.006 | .112 | **.579** | -.029 | .022 | -.066 | .009 | .086 | .005 | .182 | .041 | .050 |
| BSM3 | .052 | .081 | **.600** | .060 | .041 | .110 | .098 | .037 | .099 | .002 | .075 | -.009 |
| Presentation skills | | | | | | | | | | | | |
| PS1 | .080 | .099 | .027 | **.763** | .032 | .034 | .040 | .027 | .019 | .034 | .031 | .021 |
| PS2 | .077 | .012 | .116 | **.648** | .086 | .068 | .098 | .003 | .019 | .033 | .017 | -.010 |
| PS3 | .012 | .054 | .010 | **.575** | .132 | .021 | .011 | .028 | .025 | .297 | .014 | .026 |
| Organization course–learning | | | | | | | | | | | | |
| OC1 | -.003 | .025 | .008 | -.012 | .040 | **.737** | .025 | .076 | .024 | .037 | .029 | .034 |
| OC2 | -.024 | -.007 | .046 | .031 | -.037 | **.636** | .012 | .024 | .042 | .084 | .017 | .126 |
| OC3 | .067 | .186 | .025 | .174 | .046 | **.383** | -.018 | .186 | .011 | -.008 | .015 | .066 |
| Course materials | | | | | | | | | | | | |
| CM1 | .103 | -.024 | .131 | .159 | **.624** | .012 | .012 | -.012 | .012 | -.010 | .008 | .054 |
| CM2 | .057 | .051 | .032 | .048 | **.736** | .008 | .069 | .013 | .041 | .113 | .039 | .036 |
| CM3 | .065 | .020 | .312 | .101 | **.488** | .022 | .009 | -.005 | -.016 | .024 | .005 | .082 |
| CM4 | .117 | .069 | .017 | .129 | **.684** | .067 | .074 | .028 | .014 | .002 | .057 | .001 |
| Course difficulty | | | | | | | | | | | | |
| CD1 | .031 | .036 | .033 | .005 | .068 | -.029 | **.621** | -.012 | .114 | .098 | .040 | .008 |
| CD2 | .025 | .062 | .052 | .076 | -.011 | .040 | **.695** | -.031 | .134 | .021 | .035 | .017 |
| CD3 | .084 | .039 | .015 | .046 | .044 | .042 | **.737** | .003 | .036 | .052 | .075 | .018 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Help teacher during learning process | | | | | | | | | | | | |
| HT1 | .131 | .012 | .089 | .089 | -.080 | .153 | .037 | .008 | .021 | **.431** | .083 | .057 |
| HT2 | .116 | -.001 | .054 | .058 | -.016 | .008 | .065 | .035 | -.005 | **.626** | .037 | .077 |
| HT3 | .023 | .036 | .044 | .083 | .070 | .103 | .066 | .005 | .013 | **.659** | .024 | .039 |
| Authenticity of the examination(s) | | | | | | | | | | | | |
| AE1 | -.048 | .004 | .012 | .006 | .003 | .047 | -.062 | **.638** | -.034 | .034 | -.027 | .110 |
| AE2 | .054 | .044 | .025 | .006 | .003 | .116 | -.013 | **.734** | -.011 | .006 | .048 | .032 |
| AE3 | .025 | .015 | .059 | .031 | .002 | .042 | .023 | **.776** | .013 | .018 | .063 | .010 |
| Linking-up with foreknowledge | | | | | | | | | | | | |
| FK1 | -.007 | .161 | .046 | .003 | .083 | -.046 | .031 | -.003 | **.538** | .071 | .066 | .073 |
| FK2 | .031 | .002 | .055 | .030 | -.033 | .019 | .187 | .006 | **.689** | .000 | -.025 | -.020 |
| FK3 | .043 | .030 | .013 | .009 | .005 | .061 | .032 | -.017 | **.889** | -.009 | .036 | .010 |
| Content validity examination | | | | | | | | | | | | |
| CVE1 | -.049 | -.015 | .072 | .018 | .034 | -.048 | .059 | .040 | .020 | .094 | **.691** | .075 |
| CVE2 | .037 | .038 | .012 | .014 | -.002 | .058 | .012 | -.013 | .021 | -.009 | **.769** | .010 |
| CVE3 | .165 | .074 | .039 | .018 | .022 | .077 | .077 | .094 | .033 | .016 | **.634** | -.030 |
| Formative evaluation(s) | | | | | | | | | | | | |
| FE1 | .193 | .052 | .001 | .064 | .013 | .119 | .010 | .023 | .005 | -.023 | .041 | **.611** |
| FE2 | .013 | -.013 | .029 | -.011 | .012 | .039 | .012 | .030 | .016 | .047 | .012 | **.793** |
| FE3 | .018 | .019 | .008 | -.002 | .034 | .091 | .020 | .079 | .017 | .109 | .018 | **.687** |

*Factor Correlations*

| | *F1* | *F2* | *F3* | *F4* | *F5* | *F6* | *F7* | *F8* | *F9* | *F10* | *F11* | *F12* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 1.000 | | | | | | | | | | | |
| F2 | .221 | 1.000 | | | | | | | | | | |
| F3 | .311 | .347 | 1.000 | | | | | | | | | |
| F4 | .325 | .270 | .280 | 1.000 | | | | | | | | |
| F5 | .311 | .162 | .315 | .383 | 1.000 | | | | | | | |
| F6 | .182 | .235 | .195 | .240 | .106 | 1.000 | | | | | | |
| F7 | .262 | .214 | .212 | .231 | .185 | .123 | 1.000 | | | | | |
| F8 | .106 | .162 | .167 | .116 | .052 | .307 | -.023 | 1.000 | | | | |
| F9 | .166 | .265 | .199 | .115 | .091 | .144 | .374 | -.007 | 1.000 | | | |
| F10 | .305 | .163 | .286 | .339 | .177 | .232 | .260 | .123 | .105 | 1.000 | | |
| F11 | .262 | .204 | .233 | .127 | .131 | .181 | .238 | .157 | .154 | .205 | 1.000 | |
| F12 | .27 | .110 | .135 | .118 | .154 | .306 | .079 | .211 | .077 | .276 | .121 | 1.000 |

*Note*. ESEM = exploratory structural equation modeling; All parameter estimates are completely standardized. N = 2837 evaluations for 75 courses. A priori targets are set in bold..

The correlations between the factors for the ESEM solution are much smaller, with the strongest correlation being .38. Moreover, the factor loadings of the items on their posited indicators are much lower –and more realistic– as well. Their loadings on other factors, which are set free in the ESEM model, appear to be not that high and suggest their unidimensionality (i.e., they belong to only one factor).

In sum, both the CFA model and the ESEM model fit the data adequately and provide further evidence concerning the multidimensional structure of the SET37 questionnaire. Still, the ESEM approach takes into account the typical and complex EFA-structure of many measurement models in psychology and educational sciences, and provides results that are more in alignment with the basic theoretical assumptions of these models. These assumptions allow factor indicators in SET-instruments to have (low) loadings on nontarget factors. It is therefore recommended to use ESEM validation procedures next to the more common CFA-models when mapping the validity of SET-instruments.

## References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modelling. *Structural Equation Modeling*, 16: 397-438.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modelling*. Cary, NC: SAS Institute.

Marsh, H., Muthén, B., Aspahourov, T., Lüdtke, O., Robitzsch, A., Morin, A., & Trautwein, U. (2009). Exploratory structural equation modelling, integrating CFA and EFA: application to students' evaluations of university teaching. *Structural Equation Modeling*, 16: 439-76.

Mortelmans, D., & Spooren, P. (2009). A revalidation of the SET37 questionnaire for student evaluations of teaching. *Educational Studies*, 35: 547–52.

Penny, A., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: a meta-analysis. *Review of Educational Research*, 74: 215-53.

Richardson, J. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment and Evaluation in Higher Education*, 30: 387-415.

Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student evaluation of teaching quality in higher education. Development of an instrument based on 10 Likert scales. *Assessment and Evaluation in Higher Education*, 32: 667-79.