Studies in Educational Evaluation 64 (2020) 100816

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

University students' learning approaches: An adaptation of the revised two-factor study process questionnaire to Norwegian

Yusuf F. Zakariya[a,*], K. Bjørkestøl[a], H.K. Nilsen[a], S. Goodchild[a], M. Lorås[b]

[a] Department of Mathematical Science, University of Agder, Kristiansand, Norway
[b] Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway

ARTICLE INFO

Keywords:
Confirmatory factor analysis
R-SPQ-2F
Validity
Reliability
Learning approaches

ABSTRACT

This paper reports a Norwegian validation study of a widely used instrument to measure students' approaches to learning, namely, Bigg's revised two-factor study process questionnaire (R-SPQ-2F). Its cultural sensitivity and psychometry evaluations have provoked rigorous discussion among educators in different languages. A survey design was adopted involving 253 undergraduate engineering students across two universities. Confirmatory factor analyses were used to test six models hypothesized to reflect the factor structures of R-SPQ-2F and unidimensionality of its subscales. The results showed appropriate fits of a two-factor first-order model with 10 items measuring deep approach and 9 items measuring surface approach subscales. The reliability was found to be high with coefficients of .81, .72 and .63 on deep subscale, surface subscale and the whole instrument respectively. Findings may be interpreted as evidence of cultural sensitivity of the instrument and more validation studies were recommended.

1. Introduction

The increase in number and diversity of higher education students coupled with huge investment on the parts of government, parents, educational stakeholders and students have prompted enormous research into undergraduate students' learning experience. An important aspect of students learning that has attracted attention of education researchers over the last decades is their learning approaches (e.g., Fryer & Vermunt, 2018; Maciejewski & Merchant, 2016). Approaches to learning in higher education (HE) connotes predispositions adopted by an individual when presented with learning materials and strategies used to process the learning contents (Baeten, Kyndt, Struyven, & Dochy, 2010). A long-standing categorization of learning approaches into notions of "deep" and "surface" was introduced by Ference Marton and colleagues over 40 years ago.

Marton and Säljö developed the students' approaches to learning (SAL) theory from their qualitative clinical experimental series of studies (Marton & Säljö, 1976a, 1976b) on Swedish undergraduate students' approaches to reading, understanding and answering questions based on some presented passages of prose and newspaper articles. The experiments were aimed at exploring qualitative differences in the presented materials and describing practical differences in learning processes. In these experiments, they utilized the term "approaches to learning" to connote the *processes* adopted by the students, prior to the experiments which directly influence their learning outcome. The series of experiments resulted in a categorization of students' learning processes into deep and surface approaches.

A deep approach learner processes information with the intent of discovering the meaning of intended content of the material while a surface approach learner is preoccupied with the discourse or the text itself with little or no attention to the intended meanings. More recently, Biggs (2012) while describing surface and deep approaches to learning posited that the surface approach to learning "refers to activities of an inappropriately low cognitive level, which yields fragmented outcomes that do not convey the meaning of the encounter" and the deep approach to learning "refers to activities that are appropriate to handling the task so that an appropriate outcome is achieved." (p.42).

Measurement of students' approaches to learning is an aspect of instruction in HE that has attracted attention for the past 45 years. Questions like what should be measured in SAL?, how should it be measured?, and how many subcategories should SAL measuring instrument contain?, etc., have been investigated extensively (e.g., Kember, 1990). John Bigg's revised two-factor study process questionnaire R-SPQ-2F has been identified among the most widely studied instruments for measuring approaches to learning (e.g., Lake, Boyd, & Boyd, 2017; López-Aguado & Gutiérrez-Provecho, 2018). Similar instruments are the approaches and study skills inventory for students (ASSIST) and revised approaches to studying inventory (RASI) that

* Corresponding author.
E-mail address: yusuf.zakariya@uia.no (Y.F. Zakariya).

https://doi.org/10.1016/j.stueduc.2019.100816
Received 22 May 2019; Received in revised form 6 September 2019; Accepted 7 September 2019
Available online 09 October 2019

were developed and validated in different languages (e.g., Diseth, 2001; Tait, Entwistle, & McCune, 1998; Valadas, Gonçalves, & Faísca, 2010). Meanwhile, R-SPQ-2F has advantage over ASSIST with regards to its concise length and it is more readily interpretable than the RASI because of its lower number of primary latent factors.

However, cultural sensitivity of R-SPQ-2F when adapted to different languages has generated heated debates among researchers (López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014). In most instances, only two latent factors as opposed to four hypothesized by Biggs, Kember, and Leung (2001) have been reported to be the best explanation for the factor structure of the instrument (e.g., López-Aguado & Gutiérrez-Provecho, 2018). Contrary to Biggs et al. (2001), a handful of studies also recommended deletion of some items from the original instrument in order to achieve model fits (e.g., Socha & Sigler, 2014). These contrasting findings have created knowledge gaps for more studies on the cultural sensitivity of the instrument. It is therefore necessary to validate the Norwegian version of R-SPQ-2F before applying it to our university students. The main purpose of this study is to confirm the underlying factor structure of R-SPQ-2F and establish its reliability estimates using appropriate psychometric analysis.

2. Literature review

2.1. Factor structures of R-SPQ-2F

Psychometric properties such as validity and reliability of R-SPQ-2F have been studied extensively and the results well documented (Biggs et al., 2001; Chan & Sheung Chan, 2010; Weller et al., 2013). In Biggs et al. (2001), validity, reliability and dimensionalities of R-SPQ-2F were investigated involving 495 university students across various departments in a university in Hong Kong. The unidimensionality of each substructure – deep motive (DM), deep strategy (DS), surface motive (SM) and surface strategy (SS) – was investigated by conducting confirmatory factor analysis (CFA) which established the homogeneity of each 5-item subscale. Two models were hypothesized and tested using CFA to explain the factor structures of R-SPQ-2F. The first model (see, Fig. 1A) was a first-order four-factor model – DM, DS, SM and SS – with partial covariance and five indicators on each latent variable. The results showed a good fit with standardized root mean square residual (SRMR) = .058, comparative fit index (CFI) = .904 and correlations of .93, .70 and -.18 between DM and DS, SS and SM, and DM and SM respectively. The second model (see, Fig. 1B) was as well a first-order two-factor model – deep and surface – with two indicators each DM and DS, SM and SS respectively got by summing items corresponding to the subscales. The results also showed a good fit with SRMR = .015, CFI = .992 and correlation −.23 between deep and surface factors.

A reliability check was conducted and Cronbach's alpha coefficients of .62, .63, .72 and .57 were reported for DM, DS, SM and SS respectively. Further, acceptable Cronbach's alpha coefficients of .73 and .64 were also reported for the 10-item deep approach (DA) and Surface Approach (SA) factors respectively (Biggs et al., 2001). In a similar corroborative empirical study involving 404 students of higher diplomas and associate degrees in Hong Kong, Chan and Sheung Chan (2010) reported much higher Cronbach's alpha coefficients of .70, .74, .70, .65, .85 and .80 for DM, DS, SM, SS, DA, and SA respectively. More so, Weller et al. (2013) conducted an exploratory factor analysis (EFA) using maximum likelihood (ML) coupled with CFA after some changes in the wordings of R-SPQ-2F to suit their research field. The results made a perfect match of the two-factor extracted as in the original instrument with a considerable internal consistency and Cronbach's alpha values of .74 and .83 for DA and SA respectively.

2.2. Cultural sensitivity of R-SPQ-2F

The cultural sensitivity of R-SPQ-2F has stirred up debates among educationists in recent time especially when adapted into Spanish (Justicia, Pichardo, Cano, Berbén, & De la Fuente, 2008), Turkish (Önder & Besoluk, 2010), Japanese (Fryer, Ginns, Walker, & Nakao,

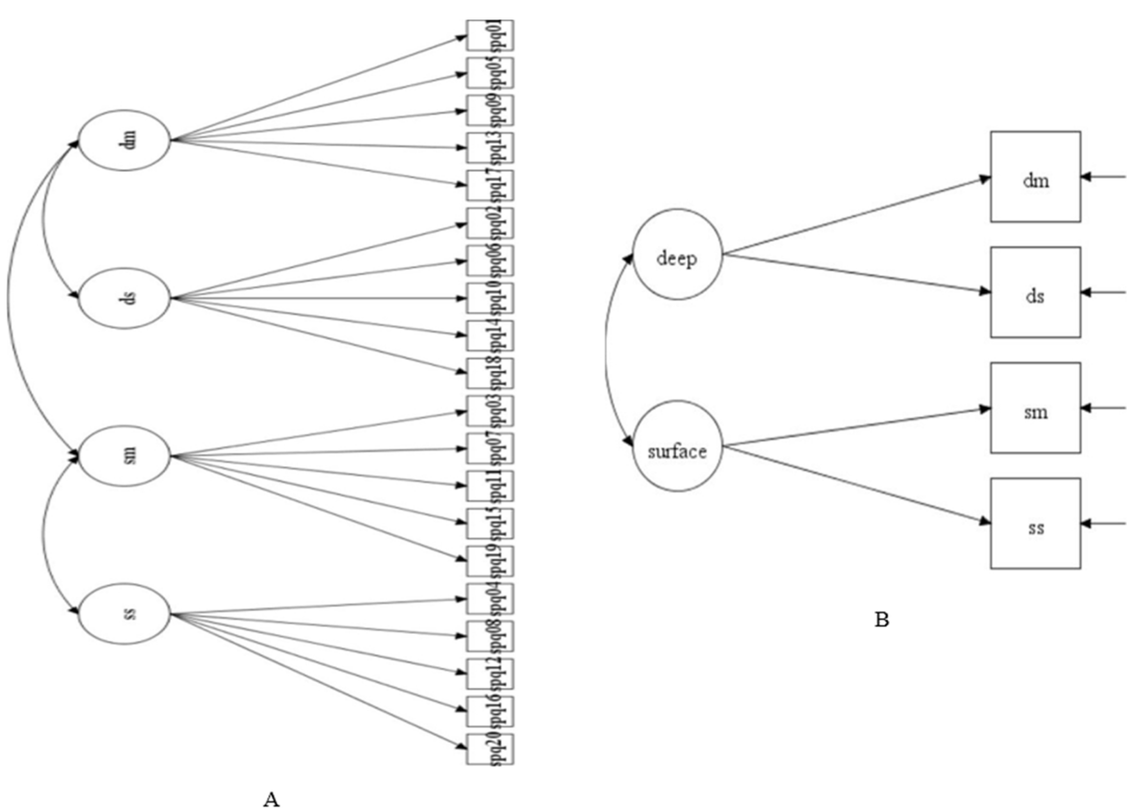

Fig. 1. Models 1 and 2 as hypothesized by Biggs et al. (2001).

2012), Dutch (Stes, De Maeyer, & Van Petegem, 2013), Chinese (Xie, 2014), and Arabic (Shaik et al., 2017). The results of these studies have ended in different conclusions with most studies proposing a two-factor R-SPQ-2F without any further subcategories into motive and strategy. In an attempt to investigate this phenomena, Leung, Ginns, and Kember (2008) conducted an empirical study on two independent samples of 1146 university students in Australia and 1266 students in Hong Kong. Their results showed no significant difference in the description of students approaches to learning in both countries, the range of developing approaches from surface to deep is common and the Cronbach's alpha coefficients ranged from 0.64 (SS) to 0.74 (SM) and 0.70 (SS) to 0.77 (DM) for both Hong Kong and Australian samples respectively. The hypothesized two-factor structural model was confirmed using CFA which gave a good fit for both samples (Leung et al., 2008). This is one of the few studies that have confirmed no cultural sensitivity of R-SPQ-2F across different cultural settings.

On the other hand, Justicia et al. (2008) were among the earlier researchers to provoke the discussion on cultural sensitivity of R-SPQ-2F. In their empirical study, data were collected from two independent samples of 314 and 522 university students. The first sample composed of mainly year one education students (used for EFA) and second sample composed of 274 and 248 final year students of education and psychology respectively (used for CFA). The R-SPQ-2F was translated to Spanish employing back-translation coupled with some modifications to cater for cultural differences. Their analysis was rigorous including EFA (both PCA and PFA – Principal Factor Analysis), CFA, item polychoric correlations to cater for multivariate normality and comparing other models. The final results confirmed two-factor structures for R-SPQ-2F, and no empirical evidence was found for differentiating between motive and strategy subscales. A corroborative result for best fit of two underlying factor structures for R-SPQ-2F was as well reported in the Turkish version of the instrument (Önder & Besoluk, 2010). Evidence of reliability was also provided with Cronbach's alpha coefficients of .78 and .74 for deep and surface dimensions respectively. The Japanese (Fryer et al., 2012), Dutch (Stes et al., 2013) and Arabic (Shaik et al., 2017) versions also reported two underlying factor structures for the R-SPQ-2F in their respective studies with little conceptual variations in deep and surface approaches.

A study that stood out almost completely was the report of Immekus and Imbrie (2010) involving two cohorts (A = 1490 and B = 1533) of university students in the United States of America. The reliability estimates were .81 and .80 (cohort A) and .81 and .78 (cohort B) for deep and surface approach subscales respectively. The interesting part was the factor analysis results. There was no empirical evidence for neither the two-factor nor for the four-factor structures of the R-SPQ-2F in the cohort A. However, a four-factor model was found fit after deleting 5 items. This was later confirmed using CFA on cohort B and found to have a good fit with acceptable statistics and the final four-factor items considerable overlapped with Bigg's et al. 2001 initial substructures (Immekus & Imbrie, 2010). In an attempt to reconcile between these variant reports on latent structures of R-SPQ-2F, Socha and Sigler (2014) conducted an empirical study involving 868 university students and compared 8 statistical models. Rather than solving the problem, they also came up with a two-factor best description of R-SPQ-2F at a cost of deleting two items (Socha & Sigler, 2014).

3. Methods

3.1. Participants

A total of 253 year-one university engineering and computer science students participated in this study. This comprised 168 males and 72 females distributed across two universities in Norway and age range of 20–23 years. 13 students did not indicate their gender. An effective sample of size of 253 was realized after subtracting ten missing cases in the main data. Despite the sample size was smaller than envisaged due to general attitudes of undergraduate students towards responding to questionnaires, it conforms with the recommendations of Monte Carlo simulation studies reported in (Gagne & Hancock, 2006; Wolf, Harrington, Clark, & Miller, 2013). This was based on the many elements such number of factors ($\le$ 4), expected factor loading ($\le$ .8), number of indicators per factor ($\le$ 10), expected power ($\ge$ .8), expected ratio of χ^2 -value to df ($\le$ 4), etc.

3.2. Materials

R-SPQ-2F was translated independently by two Norwegian first language associate professors of mathematics education. Comparison of translated versions was done, and agreements were reached on the appropriate choices of words. A back-translation to English was conducted by an English professor of mathematics education who has spent about 15 years in Norway. The back translation was compared with the original English version and minor corrections were made to cater for cultural language differences. The instrument was then converted to electronic form using SurveyXact and paper version was printed for back-up.

3.3. Procedure for data collection

Electronic version of consent forms was sent to the students via their university emails followed by a class visit for a presentation on the project. In the presentation, we gave a brief description of our project to the students and stressed the importance of their involvement in the research. At this instance, some students filled-out the paper version of the consent forms. A week after, we paid another visit with paper version of the translated R-SPQ-2F, gave a 5-minute presentation on the questionnaires and some students as well completed the paper version. This was preceded by distribution of R-SPQ-2F electronic version via emails. We gave a time frame of about three weeks to receive responses accompanied with occasional reminders. The response rate was about 35% of the total population. The low response rate could be ascribed to the general attitudes of undergraduate students towards completing questionnaires as well the busy schedules of most of the students at the time.

3.4. Procedure for data analysis

The collected data from both paper and electronic versions of R-SPQ-2F were merged, screened, relabeled, coded and saved in ASCII format. Confirmatory factor analysis was used to test six models and the results were reported in the current article. The first CFA was used to confirm model 1 proposed by Biggs et al. (2001) using weighted least square mean and variance adjusted (WLSMV) estimator in Mplus version 8.3 (Muthén & Muthén, 1998-2017; Muthén and Muthén, 1998). WLSMV was utilized as it is robust enough to perform well on analysis of ordinal data (in which basic assumptions of normality, absence of kurtosis and skewness are violated), presence of missing data and small sample size as compared to ML and others (Brown, 2015; Suh, 2015). The second CFA was used to test model 2 proposed by Biggs et al. (2001). The default ML estimator was used for this model because summing the indicators scores has inflated the categories which make it too cumbersome of WLSMV to handle. Model 3 was a modification of model 2 containing four first-order factors – DM, DS, SM and SS – measured by five indicators each and two second-order factors – deep and surface – hierarchical model.

Model 4 was a proposed modified version of model 3 containing two first-order factors – deep and surface – model measured by ten and nine indicators respectively. Models 5 and 6 were single-factor models used to check the unidimensionality of items in deep and surface subscales. WLSMV estimator was used in the analysis of models 3-6. Cronbach alpha coefficient estimate for the reliability of the instrument was not used because it depends on Pearson correlations which requires normality assumption for accurate estimates. Rather, the internal consistency of R-SPQ-2F and its subscales was checked using Raykov and Marcoulides' formula which have been confirmed to performed more efficiently than Cronbach alpha under violations of multiple assumptions (Raykov & Marcoulides, 2016). For instructional purposes, the data used for this study are available upon request and Mplus syntax codes as well the English final version of R-SPQ-2F are enclosed in the appendices. The Norwegian version of R-SPQ-2F is available upon request from the corresponding author.

3.5. Criteria for assessing a model fit

Apart from a non-significant χ^2 -value, there are a number goodness of fits (GOF) indices proposed to assess the optimality of approximate prediction of sample matrix by a CFA model. Popularly reported indices in educational studies are: TLI-Tucker-Lewis index (Tucker & Lewis, 1973), RMSEA-root mean square error of approximation (Steiger and Lind, 1980 in Steiger, 2016), SRMR (Jöreskog & Sörbom, 1988), and CFI (Bentler, 1990). For both CFI and TLI a value 1.00 indicates a perfect model fit while values close to or greater than 0.90 indicate a good fit (Bentler, 1990; Hu & Bentler, 1999). A cut-off RMSEA value of less than or equal to 0.06 was proposed by Hu and Bentler (1999) for a good model fit. Other experts (e.g., Browne & Cudeck, 1992) have proposed RMSEA values between 0.00 to 0.05 and 0.05 to 0.08 as depicting a good and an adequate model fits respectively. A model with RMSEA value between 0.08 to 0.10 was characterized as having a "mediocre fit" while models with value greater than 0.10 should be rejected (MacCallum, Browne, & Sugawara, 1996). In the case of SRMR, a value less than or equal to .08 was suggested by Hu and Bentler (1999) as an indicator of a good fit.

In practice, methodologists and researchers do not take the cut-off values of GOF indices as a rule of thumb. In fact, a close look at the work of Hu and Bentler (1999) revealed that their cut-off criteria are not generalizable especially when other estimators e.g. WLSMV apart from ML are used and more than five indicators per factors are involved in the instrument (Marsh, Hau, & Wen, 2004). Further, Hu and Bentler (1998), 1999) criteria have been considered unrealistic for most social sciences research especially when the data involved ordinal scales with multiple violations of assumptions (Marsh et al., 2004). It is therefore helpful, and of course the criteria adopted in the current study, to utilize a combination of the indices with some relaxation in cut-off values coupled with significant level of indicator factor loadings and interpretability of other parameter estimates.

4. Results

4.1. Analysis of hypothesized model 1 (Biggs et al., 2001)

The first-order four-factor model of Biggs et al. (2001) was subjected to CFA and the results are presented in Tables 1 and 2.

The results in Table 1 show a poor fit of model 1. This is evident with a significantly high χ^2 -value (167, N = 253) = 609.79, $p < .05$, and none of the fit indices is within the recommended acceptable range. In fact, the problem of this model is worse than the out of range fit indices. The latent variable covariance matrix is not positive definite (see, Table 2) which renders the model non-admissible. This was because of the presence of Heywood cases in form of standardized correlations great than one (1.018 and 1.048) between latent variables DM and DS, SM and SS respectively. This is interpreted to be an evidence of over-factoring in the model and gross misspecifications that are suggestive of redundant latent factors with high multicollinearity (Brown, 2015; Byrne, 2012). One may argue that the nonpositive definite matrix was due to pairwise estimations (one by one correlation) involved in the computation of polychoric correlation matrix used in the analysis of ordinal data. In order to clear this doubt, the estimator was changed to ML and the analysis was run again. The result is no different from the one reported in Tables 1 and 2. Hence, it can be deduced that the hypothesized model 1 is not descriptive enough of the data and therefore rejected.

Table 1
Mplus output of model 1: Selected GOF statistics.

Tests of model fits	
Chi-Square Test of Model Fit	
Value	609.786
Degrees of freedom	167
p-value	0.0000
RMSEA (Root Mean Square Error of Approximation)	
Estimate	0.102
90 Percent C.I.	0.094 0.111
Probability RMSEA < = .05	0.000
CFI/TLI	
CFI	0.710
TLI	0.670
Number of Free Parameters	103
SRMR (Standardized Root Mean Square Residual)	
Value	0.100

Table 2
Estimated correlation matrix for the latent variables.

	DM	DS	SM	SS
DM	1.000			
DS	1.018	1.000		
SM	−0.602	0.000	1.000	
SS	0.000	0.000	1.048	1.000

This finding, though contrary to the hypothesized model proposed by Biggs et al. (2001) and those who confirmed it (e.g., Merino & Kumar, 2013; Xie, 2014) it does conform to the results of non-admissible solutions reported in many studies (e.g., López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014; Stes et al., 2013). Moreover, some of the studies that confirmed admissible solutions for the four-model (e.g., Xie, 2014) also found high correlation coefficients between DM and DS, SM and SS which are suggestive of an over-factored model. They therefore concluded their studies with a two-factor explanation of the instrument (e.g., Merino & Kumar, 2013; Xie, 2014).

4.2. Analysis of hypothesized model 2 (Biggs et al., 2001)

The first-order two-factor model of Biggs et al. (2001) was subjected to CFA and the results are presented in Table 3 and Fig. 2.

The results in Table 3 appear to show a good model fit from the perspective of GOF indices. The χ^2 -value (1, N = 253) = 3.269, $p > .05$ was not significant and all the fit indices are within recommended acceptable range except RMSEA. However, the main

Table 3
Mplus output of model 2: Selected GOF statistics.

Tests of model fits	
Chi-Square Test of Model Fit	
Value	3.269
Degrees of Freedom	1
p-value	0.0706
CFI/TLI	
CFI	0.991
TLI	0.944
Number of Free Parameters	13
RMSEA (Root Mean Square Error of Approximation)	
Estimate	0.095
90 Percent C.I.	0.000 0.217
Probability RMSEA < = .05	0.160
SRMR (Standardized Root Mean Square Residual)	
Value	0.017

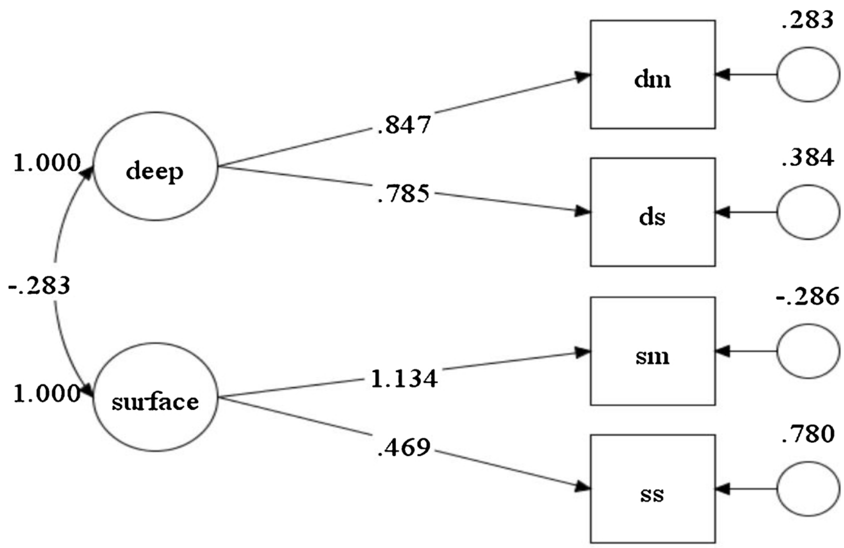

Fig. 2. Model 2 diagram with standard estimated parameters.

problem of this model is that the residual covariance matrix was not positive definite which renders the model non-admissible. This was because of the presence of Heywood case in form of negative unique variance of SM, see Fig. 2. This is suggestive of model gross misspecifications as positive definite variance/covariance matrix is a necessary condition for an admissible model (Brown, 2015; Kolenikov & Bollen, 2012). Hence, it can be deduced that the hypothesized model 2 as well is not descriptive enough of our data and therefore rejected.

This finding as well, though contrary to hypothesized model proposed by Biggs et al. (2001) it does conform to the results of non-admissible solutions reported in many studies (e.g., López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014; Stes et al., 2013). The negative unique variance found in the SM indicator completely overlapped with the finding of Socha and Sigler (2014) who also found negative error disturbance in both SM and SS indicators.

4.3. Analysis of proposed hierarchical model 3

Observed methodological issues in terms of adding up scores on component items to make indicators in model 2 coupled with its within range accompanied GOF indices prompted the test of a hierarchical four-factor model. It consists two first-order and two second-order factors tested at item levels contrary to aggregating scores used by Biggs et al. (2001). This proposed hierarchical model also relied on previous studies which have tested similar models and found admissible solutions (e.g., Justicia et al., 2008). The results are presented in Table 4 and Fig. 3.

Table 4
Mplus output of model 3: Selected GOF statistics.

Tests of model fits	
Chi-Square Test of Model Fit	
Value	521.114
Degrees of Freedom	168
p-value	0.0000
CFI/TLI	
CFI	0.769
TLI	0.739
Number of Free Parameters	102
RMSEA (Root Mean Square Error of Approximation)	
Estimate	0.091
90 Percent C.I.	0.082 0.100
Probability RMSEA < = .05	0.000
SRMR (Standardized Root Mean Square Residual)	
Value	0.081

The results in Table 4 show a poor fit of the proposed hierarchical model. This is evident with a significantly high χ^2 -value (168, N = 253) = 521.11, p < .05, and none of the fit indices is within the recommended acceptable range. Here again, the latent variable covariance matrix is not positive definite which renders the model non-admissible. This was because of presence of Heywood cases in form of negative unique variance in latent variables DM and DS, SM and SS. This is interpreted to be an evidence of over-factoring in the model and gross misspecifications that are suggestive of redundant latent factors (Brown, 2015; Byrne, 2012). Therefore, it can be deduced that the hypothesized model 3 is not descriptive enough of the data and rejected. This finding also corroborated previous studies (e.g., López-Aguado & Gutiérrez-Provecho, 2018; Merino & Kumar, 2013; Socha & Sigler, 2014) who have also reported poor fit as well as non-admissible solutions of this model.

4.4. Analysis of proposed model 4

The over-factoring observed in model 3 was corrected by collapsing latent variables DM with DS and SM with SS to form a hypothesized two-factor model and tested at item level. The results were presented in Table 5 and Fig. 4.

The results in Table 5 (with item 8) show a poor fit of the proposed two-factor model. This is evident with a significantly high χ^2 -value (169, N = 253) = 522.18, p < .05, and none of the fit indices is within the recommended acceptable range. In fact, all the GOF indices are the same with ones obtained in model 3 except a slight change in χ^2 -value. However, the model solution was admissible with a positive definite variance/covariance matrix. We investigated the estimated standardized factor loadings and found that item 8 has an extremely small nonsignificant loading (.06, p > .05) on surface approach. This item was removed as its contribution is negligible to the instrument.

The analysis was repeated and the obtained are results in Table 5 (without item 8). This showed an admissible solution with reduced χ^2 -value (151, N = 253) = 377.68, significant p < .05 with $\chi^2/df < 3$. All factor loadings are significant (see, Fig. 4), SRMR (≤ .08), CFI/TLI (closed to .90) and RMSEA (closed to 0.60) are within an acceptable range. The combined GOF indices qualified the model for an appropriate fit of the data (Marsh et al., 2004). This finding is consistent with most reported literature on the validation of R-SPQ-2F (e.g., Socha & Sigler, 2014). The negative correlation (r = -.52, p < .05) found between deep and surface subscales is an indication of discriminant

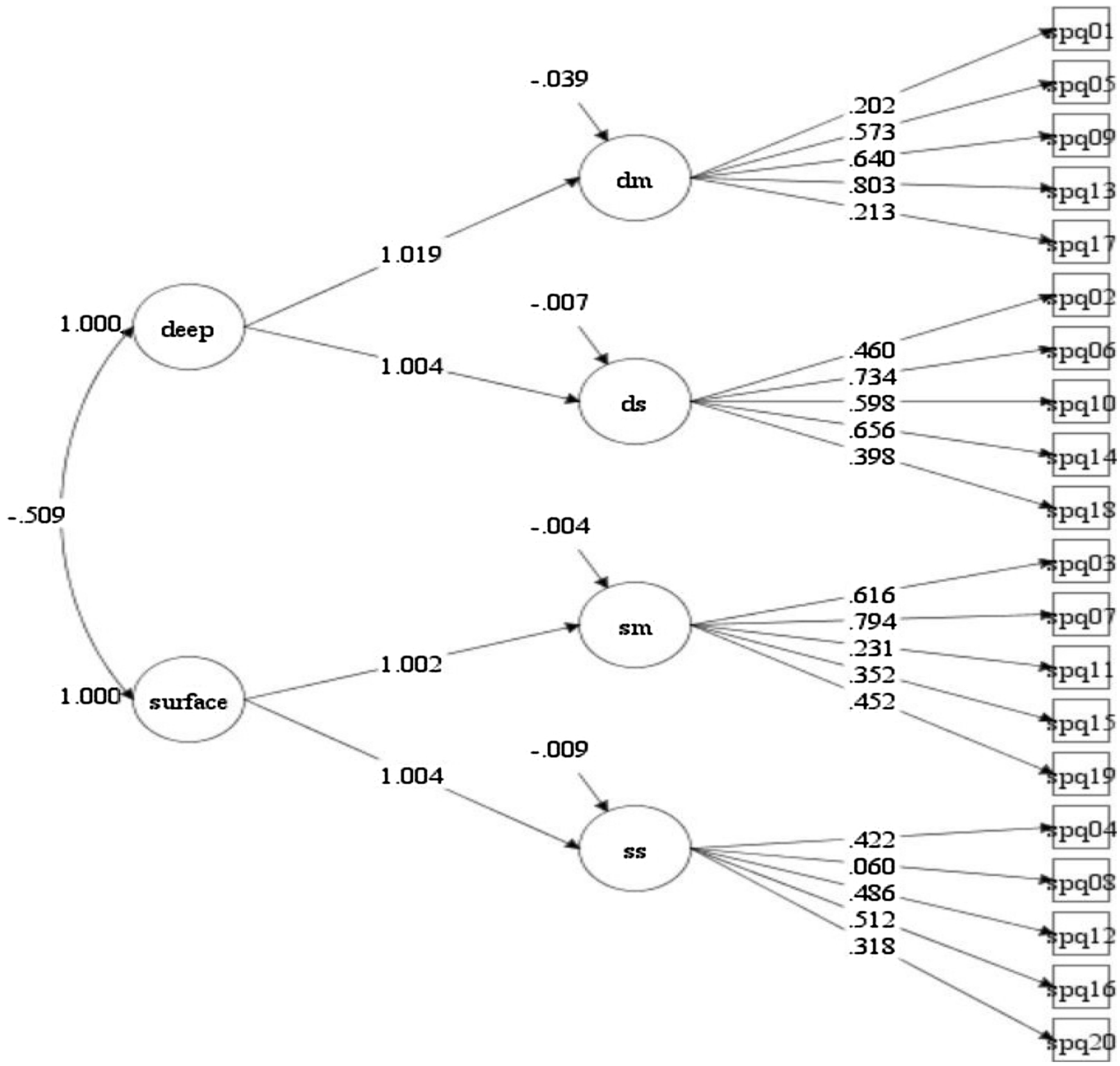

Fig. 3. Model 3 diagram with standard estimated parameters.

Table 5
Mplus output of model 4: Selected GOF statistics.

Tests of model fits		
Chi-Square Test of Model Fit	with item 8	without item 8
Value	522.179	377.676
Degrees of Freedom	169	151
p-value	0.0000	0.000
CFI/TLI		
CFI	0.769	0.844
TLI	0.740	0.824
Number of Free Parameters	101	96
RMSEA (Root Mean Square Error of Approximation)		
Estimate	0.091	0.077
90 Percent C.I.	0.082 0.100	0.067 0.087
Probability RMSEA < = .05	0.000	0.000
SRMR (Standardized Root Mean Square Residual)		
Value	0.081	0.072

validity between the subscales. This is expected and it is consistent with the literature (e.g., López-Aguado & Gutiérrez-Provecho, 2018).

4.5. Analysis of models 5 and 6

The one-factor models 5 and 6 containing 10 items on deep approach subscale and 9 items on surface approach subscale of R-SPQ-2F were fitted and reported in Table 6. The results show good model fits which confirmed the unidimensionality of each subscale. These results partly agreed with Biggs et al. (2001) and some other literature that have reported unidimensionality of items on each of deep and surface subscales (e.g., López-Aguado & Gutiérrez-Provecho, 2018).

4.6. Reliability of R-SPQ-2F

Reliability estimate of the whole R-SPQ-2F was checked as well as its subscales using latent variable approach suggested in (Raykov & Marcoulides, 2016). This approach has been proven to perform better than the conventional Cronbach's alpha estimates under violations of multiple assumptions like normality, skewness, etc. Simplified formulae adapted for the current research involving unidimensional and two-factor multidimensional scale are presented in Eqs. (1) and (2).

$$r_{RM} = \frac{(\sum_{i=1}^{n} L_i)^2}{(\sum_{i=1}^{n} L_i)^2 + \sum_{i=1}^{n} V_i} \tag{1}$$

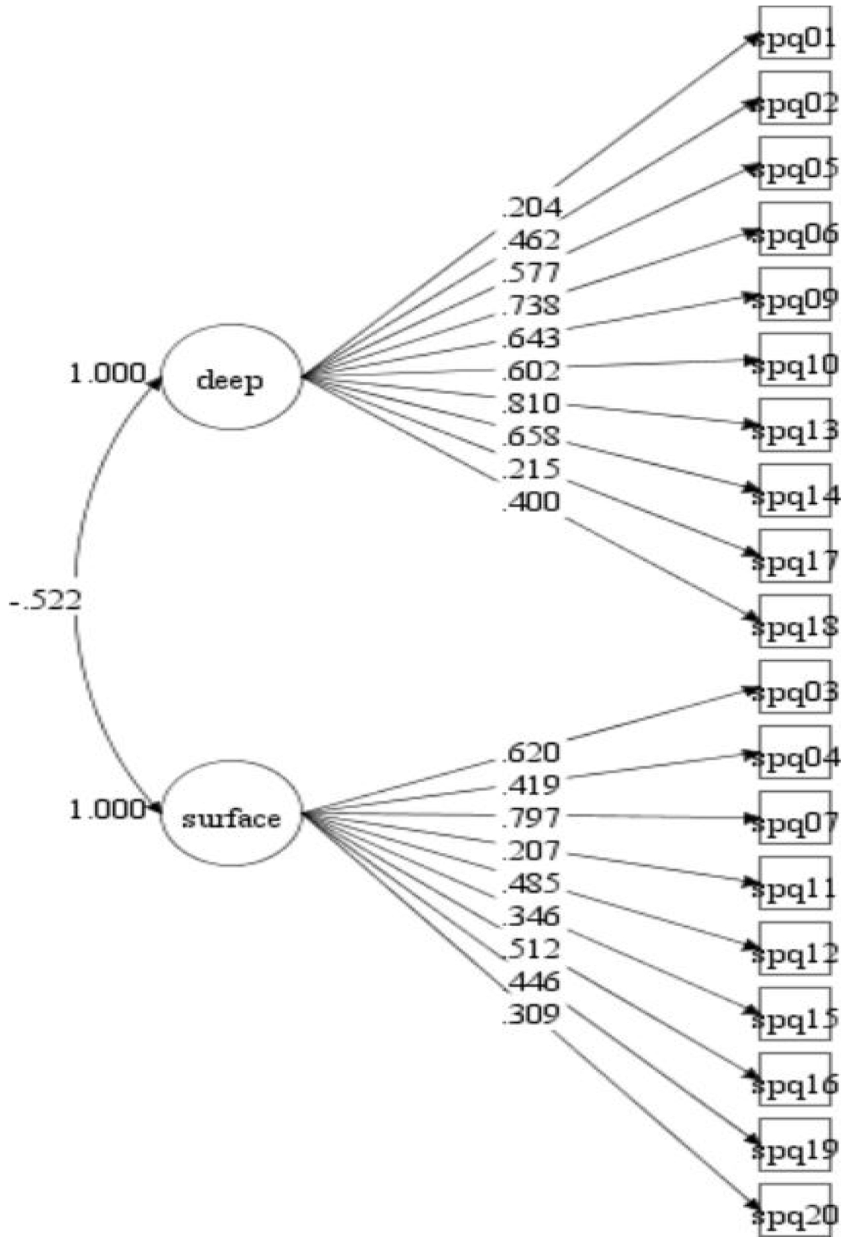

Fig. 4. Model 4 diagram with standard estimated parameters.

Table 6
Mplus output of models 5 and 6: Selected GOF statistics.

Tests of model fits		
Chi-Square Test of Model Fit	Model 5 (Deep)	Model 6 (Surface)
Value	92.884	65.58
Degrees of Freedom	35	46
p-value	0.0000	0.000
CFI/TLI		
CFI	0.943	0.919
TLI	0.926	0.887
Number of Free Parameters	50	96
RMSEA (Root Mean Square Error of Approximation)		
Estimate	0.081	0.078
90 Percent C.I.	0.061 0.101	0.054 0.101
Probability RMSEA < = .05	0.006	0.027
SRMR (Standardized Root Mean Square Residual)		
Value	0.047	0.050

$$r_{RM} = \frac{(\sum_{i=1}^{n} L_i)^2 + (\sum_{j=1}^{m} L_j)^2 + 2 * F_{12} *(\sum_{i=1}^{n} L_i)*(\sum_{j=1}^{m} L_j)}{(\sum_{i=1}^{n} L_i)^2 + (\sum_{j=1}^{m} L_j)^2 + 2 * F_{12} *(\sum_{i=1}^{n} L_i)*(\sum_{j=1}^{m} L_j) + \sum_{i=1}^{n} V_i + \sum_{j=1}^{m} V_j} \quad (2)$$

In Eqs. (1) and (2), r_{RM} is the Raykov and Marcoulides' correlation coefficient with a value ranges from 0 to 1 and interpreted like the Cronbach alfa coefficients with 0 to 1 indicative of item internal consistency from weakest to strongest (Raykov & Marcoulides, 2016). L_i' s and L_j' s are standardized factor loadings of subscale indicators, V_i' s and V_j' s are standardized unique variance computed by subtracting respective squared factor loading from 1 of each subscale indicator and F_{12} is the standardized covariance between factors 1 and 2. Using Eqs. (1) and (2) it was found that deep and surface subscales as well as the whole instrument have reliability coefficients of .81, .72 and .63 respectively. These are indicative of high reliability of the instrument. They are higher than the ones reported in (Biggs et al., 2001; López-

Aguado & Gutiérrez-Provecho, 2018) and within the ranges reported in other literature (e.g., Justicia et al., 2008; Socha & Sigler, 2014).

5. Conclusion

The cultural sensitivity of R-SPQ-2F has been attracting attention of educationists over a decade ago. Perhaps, this increased attention was prompted by the global quest for breeding university students towards deep approach to learning. This study was aimed at addressing issues related to the construct validation of this instrument as applied to the Norwegian context. In order to explain the factor structure of this instrument a series of confirmatory factor analyses were conducted, and several hypothesized models were evaluated. The best explanation found was a two-factor structure of the instrument measuring deep approach with 10 items (as theorized in the English version) and surface approach with 9 items (excluding item 8). The two-factor solution of the R-SPQ-2F found in the current study is in-line with a handful of adaptations of the instrument to Turkish (Önder & Besoluk, 2010), Spanish (Merino & Kumar, 2013), Chinese (Xie, 2014), etc.

There are several reasons that justify the removal of item 8 ("I learn some things by rote, going over and over them until I know them by heart even if I do not understand them") from the instrument. First, apart from its nonsignificant factor loading as revealed by CFA, a close look at the item itself raises some concerns. It includes some terms like "rote", "going over and over them" and "learning by heart" which seems confounding and could pose some levels of confusion to students (López-Aguado & Gutiérrez-Provecho, 2018). More so, a local misfit of this item as well as its nonsignificant factor loading have been reported in literature and its removal from the scale was recommended to obtain a valid measure (e.g., Immekus & Imbrie, 2010; Socha & Sigler, 2014).

The results, though partly contrary to the hypothesized models of Biggs et al. (2001) were similar to the ones in related studies (e.g., López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014). The findings of the current study being the first of its kinds in Norway to the best of our knowledge have provided insights into the cultural sensitivity of the R-SPQ-2F. We acknowledge the study of Diseth (2001) on approaches to learning in the Norwegian context and the contributions made in relating approaches to learning with other constructs e.g. performance (Diseth, Pallesen, Brunborg, & Larsen, 2010). Some of our findings such as classification of learning approaches into deep and surface partly overlapped. However, their studies have relied on an old instrument, ASSIST, which was considered too lengthy and almost outdated as a revised version had been invoked. This current study made a significant shift form this old trend by considering a concise and easily interpretable measure of approaches to learning in the Norwegian context. Further, the studies of Diseth and colleagues (e.g., Diseth et al., 2010) have concentrated on Psychology students in contrary to engineering students which were the focus of this current study.

The approach adopted in this study has relied on recent development in structural equation modeling for psychometric studies and very selective in the choice of statistical tools. However, the results presented here are representative of the data which might not be generalizable to other cultural backgrounds. It is therefore recommended to make further explorations of this instrument before adapting it to another cultural context. A limitation acknowledged in this current study is the inability to investigate the measurement invariance of the proposed model across different groups. It is hoped that more validation studies on the hypothesized model in an independent sample and comparison of it with other models will be conducted. The instrument as attached in Appendix A will be indispensable to university teachers within Norway and the Mplus syntax codes provided in Appendix B could be modified for further related studies in this area. It is recommended that scoring should be done as proposed by Biggs et al. (2001) and scaled after summing by dividing scores on deep approach (1 + 2 + 5 + 6 + 8 + 9 + 12 + 13 + 16 + 17) by 10 and scores on surface approach (3 + 4 + 7 + 10 + 11 + 14 + 15 + 18 + 19) by 9. This is conjectured to enhance the interpretation of the scores.

Acknowledgements

We acknowledge the support of faculty of engineering and science, University of Agder, Kristiansand, Norway as well as MatRIC, Centre for Research, Innovation and Coordination of Mathematics Teaching for supporting this research. The anonymous reviewers are also acknowledged for their comments and recommendations towards the improvement of the article.

Appendix A. Revised Study Process Questionnaire (R-SPQ-2F)

This questionnaire has a number of questions about your attitudes towards mathematics and your usual way of studying it.

There is no *right* way of studying. It depends on what suits your own style and the course you are studying. It is accordingly important that you answer each question as honestly as you can.

Place the mark (✓) at the appropriate option to each statement. The letters alongside each number stand for the following response.

A—this item is *never* or *only rarely* true of me

B—this item is *sometimes* true of me

C—this item is true of me about *half the time*

D—this item is *frequently* true of me

E—this item is *always* or *almost always* true of me

	Statement on approaches to learning mathematics	A	B	C	D	E
1	I find that at times studying gives me a feeling of deep personal satisfaction.					
2	I find that I have to do enough work on a topic so that I can form my own conclusions before I am satisfied.					
3	My aim is to pass the course while doing as little work as possible.					
4	I only study seriously what's given out in class or in the course outlines.					
5	I feel that virtually any topic can be highly interesting once I get into it.					
6	I find most new topics interesting and often spend extra time trying to obtain more information about them					
7	I do not find my course very interesting so I keep my work to the minimum.					
8	I find that studying academic topics can at times be as exciting as a good novel or movie.					
9	I test myself on important topics until I understand them completely.					
10	I find I can get by in most assessments by memorising key sections rather than trying to understand them.					
11	I generally restrict my study to what is specifically set as I think it is unnecessary to do anything extra.					
12	I work hard at my studies because I find the material interesting.					
13	I spend a lot of my free time finding out more about interesting topics which have been discussed in different classes.					
14	I find it is not helpful to study topics in depth. It confuses and wastes time, when all you need is a passing acquaintance with topics.					
15	I believe that lecturers shouldn't expect students to spend significant amounts of time studying material everyone knows won't be examined.					

16	I come to most classes with questions in mind that I want answering
17	I make a point of looking at most of the suggested readings that go with the lectures.
18	I see no point in learning material which is not likely to be in the examination.
19	I find the best way to pass examinations is to try to remember answers to likely questions.

Appendix B. Mplus syntax codes used for the analysis

```
TITLE: CFA MODEL 1 BIGGS et al., 2001.
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   CATEGORICAL ARE SPQ01-SPQ20;
   USEVARIABLES ARE SPQ01-SPQ20;
   MISSING ARE ALL (-1);
ANALYSIS:
ESTIMATOR IS WLSMV;
MODEL:
DM by SPQ01 SPQ05 SPQ09 SPQ13 SPQ17;
DS by SPQ02 SPQ06 SPQ10 SPQ14 SPQ18;
SM by SPQ03 SPQ07 SPQ11 SPQ15 SPQ19;
SS by SPQ04 SPQ08 SPQ12 SPQ16 SPQ20;
DM WITH SS@0;
DS WITH SS@0;
DS WITH SM@0;

OUTPUT: MODINDICES STANDARDIZED TECH4;
```

```
TITLE: CFA MODEL 2 BIGGS et al. 2001
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   USEVARIABLES ARE DM DS SM SS;
   MISSING ARE ALL (-1);
ANALYSIS:
ESTIMATOR IS ML;
MODEL:
Deep by DM DS;
Surface by SM SS;

OUTPUT: MODINDICES STANDARDIZED TECH4;
```

```
TITLE: CFA HIERARCHICAL MODEL 3
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   CATEGORICAL ARE SPQ01-SPQ20;
   USEVARIABLES ARE SPQ01-SPQ20;
   MISSING ARE ALL (-1);
ANALYSIS:
ESTIMATOR IS WLSMV;
MODEL:
DM by SPQ01 SPQ05 SPQ09 SPQ13 SPQ17;
DS by SPQ02 SPQ06 SPQ10 SPQ14 SPQ18;
SM by SPQ03 SPQ07 SPQ11 SPQ15 SPQ19;
SS by SPQ04 SPQ08 SPQ12 SPQ16 SPQ20;
Deep by DM* DS;
Surface by SM* SS;
DM DS(1);
SM SS(1);
Deep@1;
Surface@1;

OUTPUT: STANDARDIZED TECH4;
```

```
TITLE: CFA 2-factor model without item 8
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   USEVARIABLES ARE SPQ01-SPQ06 SPQ07 SPQ09-SPQ20;
   CATEGORICAL ARE SPQ01-SPQ20;
   MISSING ARE ALL (-1);

ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION = THETA;
ITERATION = 1000;
MODEL:
Deep by SPQ01 SPQ02 SPQ05 SPQ06 SPQ09 SPQ10 SPQ13
SPQ14 SPQ17 SPQ18;
Surface by SPQ03 SPQ04 SPQ07 SPQ11 SPQ12 SPQ15 SPQ16
SPQ19 SPQ20;

OUTPUT: MODINDICES (ALL) STDYX TECH4;
```

```
TITLE: CFA unidimentionality deep approach
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   USEVARIABLES ARE SPQ01 SPQ02 SPQ05 SPQ06 SPQ09
SPQ10 SPQ13 SPQ14 SPQ17 SPQ18;
   CATEGORICAL ARE SPQ01 SPQ02 SPQ05 SPQ06 SPQ09
SPQ10 SPQ13 SPQ14 SPQ17 SPQ18;
   MISSING ARE ALL (-1);

ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION = THETA;
ITERATION = 10000;
MODEL:
Deep by SPQ01 SPQ02 SPQ05 SPQ06 SPQ09 SPQ10 SPQ13
SPQ14 SPQ17 SPQ18;

OUTPUT: STANDARDIZED STDY TECH4;
```

```
TITLE: CFA unidimentionality surface
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   USEVARIABLES ARE SPQ03 SPQ04 SPQ07 SPQ11 SPQ12
SPQ15 SPQ16 SPQ19 SPQ20;
   CATEGORICAL ARE SPQ03 SPQ04 SPQ07 SPQ11 SPQ12
SPQ15 SPQ16 SPQ19 SPQ20;
   MISSING ARE ALL (-1);
 ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION = THETA;
ITERATION = 10000;
MODEL:
Surface by SPQ03 SPQ04 SPQ07 SPQ11 SPQ12 SPQ15 SPQ16
SPQ19 SPQ20;
SPQ03 WITH SPQ07;

OUTPUT: STANDARDIZED STDY TECH4;
```

References

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*(3), 243–260. https://doi.org/10.1016/j.edurev.2010.06.001.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.

Biggs, J. B. (2012). What the student does: Teaching for enhanced learning. *Higher Education Research & Development, 31*(1), 39–55. https://doi.org/10.1080/07294360.2012.642839.

Biggs, J. B., Kember, D., & Leung, D. Y. P. (2001). The revised two factor study processquestionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*, 133–149.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research.* New York, London: The Guilford Press.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230–258. https://doi.org/10.1177/0049124192021002005.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* New York: Routledge, Taylor & Francis Group.

Chan, Y. M., & Sheung Chan, C. M. (2010). Approach to learning of sub-degree students in Hong Kong. *Research in Education, 84*(1), 65–78.

Diseth, Å. (2001). Validation of a Norwegian version of the approaches and study skills inventory for students (ASSIST): Application of structural equation modelling. *Scandinavian Journal of Educational Research, 45*(4), 381–394.

Diseth, Å., Pallesen, S., Brunborg, G. S., & Larsen, S. (2010). Academic achievement among first semester undergraduate psychology students: The role of course experience, effort, motives and learning strategies. *Higher Education, 59*, 335–352.

Fryer, L. K., Ginns, P., Walker, R. A., & Nakao, K. (2012). The adaptation and validation of the CEQ and the R-SPQ-2F to the Japanese tertiary environment. *The British Journal of Educational Psychology, 82*(Pt 4), 549–563. https://doi.org/10.1111/j.2044-8279.2011.02045.x.

Fryer, L. K., & Vermunt, J. D. (2018). Regulating approaches to learning: Testing learning strategy convergences across a year at university. *The British Journal of Educational Psychology, 88*(1), 21–41. https://doi.org/10.1111/bjep.12169.

Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research, 41*(1), 65–83. https://doi.org/10.1207/s15327906mbr4101_5.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Immekus, J. C., & Imbrie, P. K. (2010). A test and cross-validation of the revised two-factor study process questionnaire factor structure among western university students. *Educational and Psychological Measurement, 70*(3), 495–510. https://doi.org/10.1177/0013164409355685.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: Guide to the program and applications* (2nd ed.). Chicago, IL: SPSS.

Justicia, F., Pichardo, M. C., Cano, F., Berbén, A. B. G., & De la Fuente, J. (2008). The revised two-factor study process questionnaire (R-SPQ-2F): Exploratory and confirmatory factor analyses at item level. *European Journal of Psychology of Education, 23*(3), 355–372.

Kember, D. (1990). Cultural specificity of approaches to study. *The British Journal of Educational Psychology, 60*, 356–363.

Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances. *Sociological Methods & Research, 41*(1), 124–167. https://doi.org/10.1177/0049124112442138.

Lake, W. W., Boyd, W. E., & Boyd, W. (2017). Understanding how students study: The genealogy and conceptual basis of a widely used pedagogical research tool, Biggs' study process questionnaire. *International Education Studies, 10*(5), 100–108. https://doi.org/10.5539/ies.v10n5p100.

Leung, D. Y. P., Ginns, P., & Kember, D. (2008). Examining the cultural specificity of approaches to learning in universities in Hong Kong and Sydney. *Journal of Cross-cultural Psychology, 39*(3), 251–266. https://doi.org/10.1177/0022022107313905.

López-Aguado, M., & Gutiérrez-Provecho, L. (2018). Checking the underlying structure of R-SPQ-2F using covariance structure analysis / Comprobación de la estructura subyacente del R-SPQ-2F mediante análisis de estructura de covarianza. *Cultura y Educación/ Culture and Education, 30*(1), 105–141. https://doi.org/10.1080/11356405.2017.1416787.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.

Maciejewski, W., & Merchant, S. (2016). Mathematical tasks, study approaches, and course grades in undergraduate mathematics: A year-by-year analysis. *International Journal of Mathematical Education in Science and Technology, 47*(3), 373–387. https://doi.org/10.1080/0020739x.2015.1072881.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling A Multidisciplinary Journal, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2.

Marton, F., & Säljö, R. (1976a). On qualitative differences in learning - II: Outcome as a function of the learner's conception of the task. *The British Journal of Educational Psychology, 46*, 115–127.

Marton, F., & Säljö, R. (1976b). On qualitative differences in learning I: Outcome and process. *The British Journal of Educational Psychology, 46*, 4–11.

Merino, S. C., & Kumar, P. R. (2013). Validación estructural del R-SPQ-2F: Un análisis factorial confirmatorio. *Revista Digital de Investigación en Docencia Universitaria, 7*, 113–127.

Muthén, L. K., & Muthén, B. O. (1998-2017). Mplus user's guide. (Eighth ed.). Los Angeles, CA: Muthén & Muthén.

Önder, I., & Besoluk, S. (2010). Adaptation of revised two factor study process questionnaire (R-SPQ-2F) to Turkish. *Education and Science, 35*(157), 55–67.

Raykov, T., & Marcoulides, G. A. (2016). Scale reliability evaluation under multiple assumption violations. *Structural Equation Modeling A Multidisciplinary Journal, 23*(2), 302–313. https://doi.org/10.1080/10705511.2014.938597.

Shaik, S. A., Almarzuqi, A., Almogheer, R., Alharbi, O., Jalal, A., & Alorainy, M. (2017). Assessing Saudi medical students learning approach using the revised two-factor study process questionnaire. *International Journal of Medical Education, 8*, 292–296. https://doi.org/10.5116/ijme.5974.7a06.

Socha, A., & Sigler, E. A. (2014). Exploring and "reconciling" the factor structure for the revised two-factor study process questionnaire. *Learning and Individual Differences, 31*, 43–50. https://doi.org/10.1016/j.lindif.2013.12.010.

Steiger, J. H. (2016). Notes on the Steiger–Lind (1980) handout. *Structural Equation Modeling A Multidisciplinary Journal, 23*(6), 777–781. https://doi.org/10.1080/10705511.2016.1217487.

Stes, A., De Maeyer, S., & Van Petegem, P. (2013). Examining the cross-cultural sensitivity of the revised two-factor study process questionnaire (R-SPQ-2F) and validation of a Dutch version. *PloS One, 8*(1), e54099. https://doi.org/10.1371/journal.pone.0054099.

Suh, Y. (2015). The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Structural Equation Modeling A Multidisciplinary Journal, 22*(4), 568–580. https://doi.org/10.1080/10705511.2014.937669.

Tait, H., Entwistle, N. J., & McCune, V. (1998). ASSIST: A re-conceptualisation of the approaches to studying inventory. In C. Rust (Ed.). *Improving students as learners* (pp. 262–271). Oxford: Oxford Brookes University.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10.

Valadas, S., Gonçalves, F., & Faísca, L. (2010). Approaches to studying in higher education Portuguese students: A Portuguese version of the approaches and study skills inventory for students. *Higher Education, 59*, 259–275.

Weller, J. M., Henning, M., Civil, N., Lavery, L., Boyd, M. J., & Jolly, B. (2013). Approaches to learning for the ANZCA final examination and validation of the revised study process questionnaire in specialist medical training. *Anaesthesia and Intensive Care, 41*, 631–640.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 76*(6), 913–934. https://doi.org/10.1177/0013164413495237.

Xie, Q. (2014). Validating the revised two-factor study process questionnaire among Chinese university students. *The International Journal of Educational and Psychological Assessment, 16*(1).