A Re-Evaluation of the Approaches to Learning Questionnaire

Alexander Whitelock-Wainwright
Portfolio of the DVC & VP (Education)
Monash University
alex.wainwright@monash.edu

Dragan Gašević
Faculty of Information Technology
Monash University
dragan.gasevic@monash.edu

Trevor Woods
Monash Education Innovation
Monash University
trev.wood@monash.edu

Kris Ryan
Portfolio of the DVC & VP (Education)
Monash University
kris.ryan@monash.edu

20/03/2020

Abstract

Scores obtained from the Revised Two Factor Study Process Questionnaire are regularly used to profile students as being *Deep* or *Surface* learners. A critical examination of how this particular instrument was both developed and validated does, however, raise questions over what is being measured. Possible factor models have been tested – guided by the original theorising – yet consensus has not been achieved over what is a permissible representation. Aside from uncertainties around model candidature agreement, researchers continue to adopt item parcelling when analysing data collected using this instrument. This is often undertaken without any exploration of whether the assumptions of item-parcelling (e.g., no cross-loadings) are violated or not. Two research questions are explored in this work, motivated by the associated problems with model structure and analysis strategies. Using data collected from a final sample of 1,158 students, three model representations were tested using both Exploratory Structural Equation Modelling and Confirmatory Factor Analysis. Findings were not in-line with the originally presented results in terms of model acceptability. More importantly, the results raise questions around the construct validity of the Revised Two Factor Study Process Questionnaire. Discussion of these results centres on why item parcelling is not recommended when using this instrument and general concerns around what is being measured. Recommendations on alternative approaches to measuring learning strategies are offered, alongside a critical examination of the underlying model.

Key Words: *Approaches to Learning, R-SPQ-2F, Factor Analysis*

1 Introduction

Measurement of student learning has predominately been carried out through the use of questionnaires, spanning such constructs as achievement goal orientation (Elliot, Murayama, & Pekrun, 2011) and meta-cognition (Tock & Moxley, 2017). As an approach, questionnaires are advantageous as they can be used to measure traits (personality, motivations, or self-efficacy) that do not arise from analysis of alternative data streams (event traces; Winne & Perry, 2000). Even so, there are problems with the use of questionnaires in educational settings such as the expectation that students are able to provide accurate and valid retrospective accounts of their learning processes (Richardson, 2004).

The Study Process Questionnaire (SPQ) of Biggs (1987) or its revised version (the Revised Two Factor Study Process Questionnaire, R-SPQ-2F; Biggs, Kember, & Leung, 2001) are examples of instruments wherein students provide self-reports that are not without response bias (Richardson, 2004). Both questionnaires aim to measure student attitudes towards study strategies and motivations, with students being directed to think about a subject of importance when responding to items (Biggs et al., 2001). Thus, there is a reliance upon students giving retrospective accounts of their motivations and strategy enactment within a specific setting. It is possible that such reports are distorted accounts due to issues associated with memory or social desirability (Richardson, 2004). Additional questions pertaining to the validity of the R-SPQ-2F have also been raised by several authors (Justicia, Pichardo, Cano, Berbén, & De la Fuente, 2008; Socha & Sigler, 2014), which have prompted questions about the scale development approach undertaken by Biggs et al.. This paper puts aside such questions about possible response bias; instead, focusing on exploring the issues in the development of the R-SPQ-2F and follow-up psychometric evaluations.

The need to re-evaluate the R-SPQ-2F is also required since the application of the underlying model has been widely applied in both face-to-face and online learning settings. Examples of which include Ginns and Ellis (2007), wherein clustering was applied to student responses to the R-SPQ-2F as a means to create learning profiles. Bliuc, Ellis, Goodyear, and Piggott (2010) explored the relationship that face-to-face and online learning variants of the R-SPQ-2F scales (e.g., surface online approaches) have with final grades. In one instance, a Higher-Education institution has incorporated the approaches to learning framework as a means to develop learning and teaching (e.g., staff development; Parpala & Lindblom-Ylänne, 2012). Across each example, the authors have assumed that both measure and underlying framework are robust, which may in fact not be the case. It is for this reason that the questionnaire requires re-evaluation, as if problems are detected then previous empiric (e.g., Bliuc et al., 2010) work may be called into question.

1.1 Approaches to Learning

Biggs (1987) proposed the 3P (Presage, Process, Product) model of student learning that sought to capture the interplay of personal and situational factors in learning processes. Parallels can be drawn with the influence of internal and external conditions on study behaviour that are detailed in the self-regulated learning model proposed by (Butler & Winne, 1995; Winne & Hadwin, 1998). In either case, student strategies and motives are believed to be affected by an array of factors including: prior-knowledge, self-efficacy, personality, learning design, resources, and time-on-task. This drew upon prior work exploring differences in strategy and motive across academic disciplines (Arts and Science; Biggs, 1970). Findings were indicative of Arts student learning approaches being focused on what were termed *simplifying strategies* due to having a subject structure that was not well-defined. Science students were reported to be more *intrinsically motivated* and *dogmatic* in their approach to learning. *Product* covers general outcomes such as grades and satisfaction, but also extends into a student's future goal setting and self-efficacy beliefs.

Central to the 3P model is the learning process complex, which Biggs viewed as embodying the students' *motives* and *strategies* (Biggs, 1987). A *motive* can generally be thought of as the reasons as to why the students approaches the task (e.g., to get a good grade). *Strategy* represents the way in which the task is approached, covering such things as making notes, reading widely, or rote learning. Biggs believed that together, *motives* and *strategies* define three approaches to learning: *Achieving*, *Deep*, and *Surface*. *Achieving Approaches* are centred around students viewing learning as a competition, wherein the goal is to attain the highest grade. *Deep Approaches* are associated with students wanting to be competent in a subject due to having an intrinsic interest in the topic, inciting them to undertake strategies such as reading widely. *Surface Approaches*, on the other hand, are typified by a reluctance to deeply understand the topic; rather, there is a preference to do the minimum expected without failing. Together, these three approaches (*Achieving, Deep, and Surface*) and *Motive-Strategy* decomposition provides a 3x2 model of learning approaches.

To measure the constructs of the 3x2 model (i.e., *Deep-Motive*, *Surface-Strategy*, *Achieving-Strategy*), Biggs developed and validated the SPQ (Biggs, 1987). This was later revised to become the R-SPQ-2F (Biggs et al., 2001). What differentiates the R-SPQ-2F from the SPQ, apart from the reduction in item numbers, is the removal of the *Achieving Approach* constructs. Reasoning behind this decision was due to *Achieving Approach* referring the location and time at which a task is undertaken. Biggs et al. also viewed the *Achieving Approach* factor as ambiguous for the purpose of monitoring learning environments compared to either the *Deep* and *Surface* factors (Biggs et al., 2001). The development of the R-SPQ-2F was therefore about measuring a 2x2 model: *Deep-Motive*, *Deep-Strategy*, *Surface-Motive*, and *Surface-Strategy*.

1.2 Proposed Models of Learning Approaches

1.2.1 Biggs' Four Factor Model

Our aim is to explore the process followed by Biggs et al. (2001) to both refine and validate the purported four factor (*Deep-Motive*, *Deep-Strategy*, *Surface-Motive*, and *Surface-Strategy*) model of learning approaches. This will involve an exploration into how select items were dropped and how specified models were evaluated.

Reducing an initial item pool of 43 down to 20 − those contained in the R-SPQ-2F − was informed by: Cronbach's α values, inter-item correlations, and modification indices (Biggs et al., 2001). Refinement of scale items in the way adopted by Biggs et al. can be scrutinised on two grounds: tau-equivalence (Tavakol & Dennick, 2011) and capitalisation of chance (MacCallum, Roznowski, & Necowitz, 1992).

Evaluating scale reliability has often relied upon Cronbach's α, which assumes items measure a single latent variable (tau-equivalence; Tavakol & Dennick, 2011). Utilising Cronbach's α as an approach to item reduction, as used by Biggs et al. (2001), becomes questionable given the hypothesised four factor structure. Even if overlooked, reported Cronbach's α values for the retained 20-item measure fall in a range (.57-.72; Biggs et al., 2001) that does not overwhelmingly satisfy the heuristic of .70, the cut-off for acceptability (Tavakol & Dennick, 2011).

Capitalising on chance is exemplified by Biggs et al. (2001) omitting items based on modification indices obtained from the application of CFA. These indices identify points of localised strain within the factor model; addition or removal of parameters are steps to remedy such ill-fit (MacCallum et al., 1992). This represents a problematic approach for the sole reason that modifications are made with a view of improving model fit over a critical evaluation of why the model may not align with reality. These modifications are often carried out in an iterative process of evaluating and modifying a model, introducing many forking paths in how an eventual 'good' fitting model is obtained. Simulation studies of MacCallum et al. (1992) also show that such modifications are reflective of sampling fluctuations; misspecifcations in Sample A may not be the same for Sample B.

Difficulty in evaluating how Biggs et al. (2001) dropped 23 of the 43 item pool is due to such details being omitted. What is provided are the details on how unidimensional models had acceptable fits to the data, despite a multidimensional (four factor) model originally being hypothesised. This presents problems associated with Biggs et al. later adopting an item parcel approach as it ignores potential cross-loading items; discussion of these details are presented below. When the multidimensional models are eventually tested, two approaches are followed:

1. A first-order factor model, wherein correlations are specified within approaches (*Deep* and *Surface*) and between opposing motives (not strategies; Model A; Figure 1); *and*

2. A two-factor higher-order model (*Deep* and *Surface*), each of which is measured by two first-order factors (*Strategy* and *Motive*) represented as item parcels (Model B; Figure 2).

The Comparative Fit Index (CFI) and Standardized Root Mean Square Residual (SRMR) were used to evaluate each model: .904 (CFI) and .058 (SRMR) for Model A; .992 (CFI) and .015 (SRMR) for Model B. These fit indices clearly show a fundamental problem in the work of Biggs et al. as Model A does not fully meet those guidelines of acceptable fit (CFI = .95, SRMR = .08; Hu & Bentler, 1999); Model B, through the use of item parcelling, would however be indicative of good fit.

To understand the discrepancies between Model A and B, there is a need to re-consider the adopted approach of Biggs et al. (2001) to separately test four factors, assuming they were unidimensional. To be correctly considered as unidimensional, it is necessary to show that the assumptions of a CFA model (e.g., no cross-loadings or correlated errors) are not violated (Marsh, Lüdtke, Nagengast, Morin, & von Davier, 2013). Only when such conservative assumptions are supported is the item parcelling strategy appropriate. Model A is a test of the item level representation and the reported fit does suggest that a misspecified model due to values falling short of the acceptability criteria; sources of local misfit were not reported by Biggs et al. (2001). The likelihood is that by testing each factor in isolation has ignored items that cross-load, resulting in an inadequate fit. Take items:

1. *I find that at times studying gives me a feeling of deep personal satisfaction* (Item 1; *Deep-Motive*); *and*

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* (Item 2; *Deep-Strategy*).

From the content, it is clear that Item 1 is associated with intrinsic interests and wanting to study as it is personally rewarding, whilst Item 2 captures the strategy to achieve this end state. Nevertheless, Item 2 still asks about the satisfaction that arises from studying so it is expected that method effects would be present, biasing its association with other factors. Item parcelling effectively hides these misspecifications as each of the four factors is represented by the sum total of five items, possibly resulting in the improved fit reported for Model B.

Despite warnings against item parcelling (Marsh et al., 2013), authors such as Fryer, Ginns, Walker, and Nakao (2012) have similarly adopted this approach

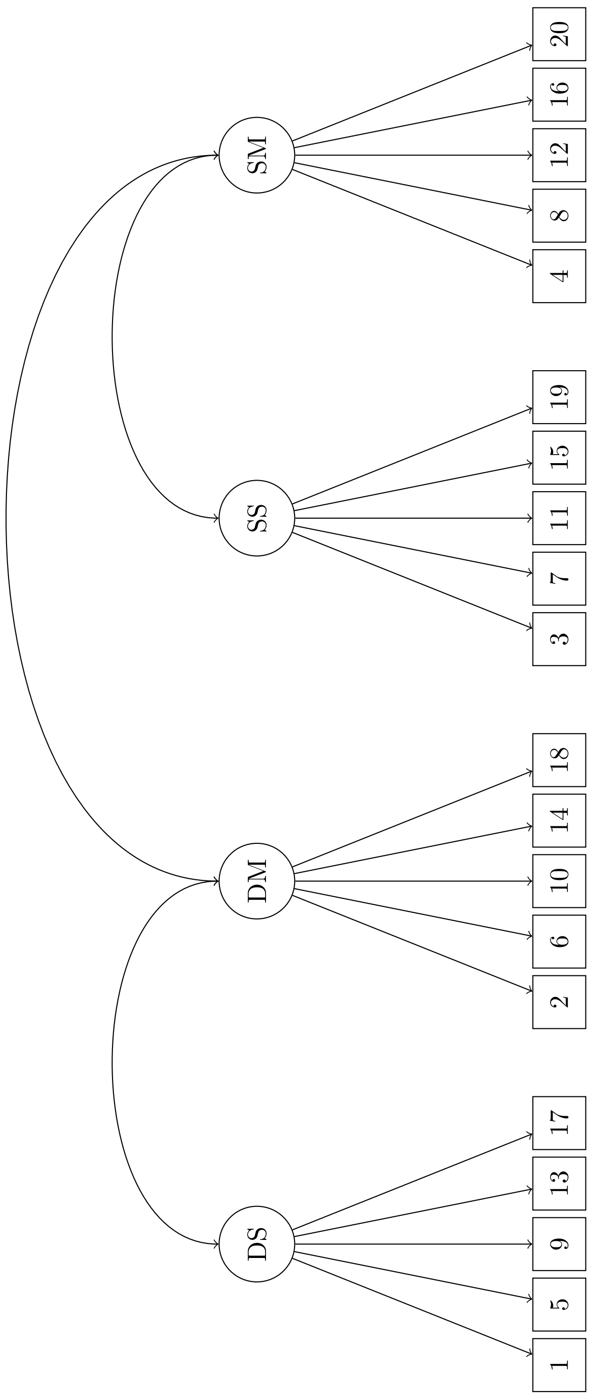


Figure 1: First-Order Factor Model Tested by Biggs et al. (2001) (DS = Deep-Strategy, DM = Deep-Motive, SS = Surface-Strategy, SM = Surface-Motive; Model A). Freely estimated correlations between: DS and DM, DM and SM, and SS and SM. Model Fit: CFI = .904, SRMR = .058.

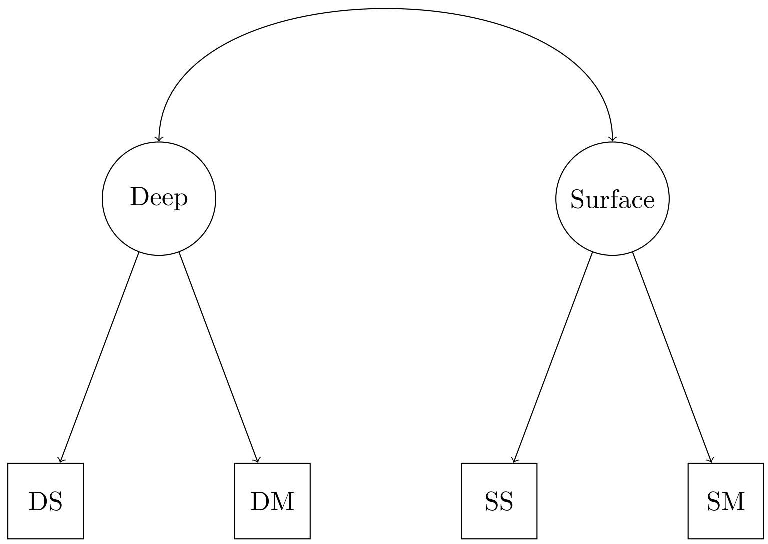


Figure 2: Higher-Order Factor Structure Tested by Biggs et al. (2001) using an Item Parcelling Approach (DS = Deep-Strategy, DM = Deep-Motive, SS = Surface-Strategy, SM = Surface-Motive; Model B). Model Fit: CFI = .992, SRMR = .015.

when it comes to evaluating the R-SPQ-2F. Findings closely paralleling those of Biggs et al. (2001) are found, wherein a four-factor solution falls short of what may be considered an acceptable fit (CFI = .78, TLI (Tuker-Lewis Index) = .73, RMSEA (Root Mean Square Error of Approxmiation) = .063; Fryer et al., 2012). When sub-scales (*Motive* and *Strategy*) for each learning approach (*Deep* and *Surface*) are instead summated, problems again become masked. This is exemplified by the model fit indices being indicative of an excellent fitting model (CFI = 1, TLI = 1, RMSEA = 0), despite the item-based approach suggesting the model was misspecified (Biggs et al., 2001).

There is insufficient evidence presented by Biggs et al. (2001) to fully support conjectures of model misspecification, but they are not isolated instances (Justicia et al., 2008; López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014). Even so, the limited critique of using item parcelling with R-SPQ-2F items necessitates the following question to be asked:

- Are item parcels an appropriate approach when analysing data from the R-SPQ-2F? (Research Question 1; RQ1)

1.2.2 Alternative Model Representations

Investigations into the factor structure of the R-SPQ-2F have resulted in findings that are contradictory to those reported by Biggs et al. (2001) (see Table 1).

The work of Justicia et al. (2008) is an example of such contradictions. Through the evaluation of four model representations, Justicia et al. found the best fitting model to include two factors representing *Deep* and *Surface Approaches* (Table 1). No evidence to support the decomposition of *Deep* and *Surface* factors into respective *Motive* and *Strategy* factors was found. Similar conclusions were drawn by Xie (2014), who found a better model fit for a two-factor (*Deep* and *Surface Approaches*) model over a four-factor model (*Deep-Strategy*, *Deep-Motive*, *Surface-Strategy*, and *Surface-Motive*; Table 1).

Immekus and Imbrie (2010) evaluated the suitability of a four, first-order factor model; the model deviated from what was tested by Biggs et al. due to all factors being free to correlate. In the Biggs et al. (2001) model, correlations were only specified between: *Deep-Strategy* and *Deep-Motive*, *Surface-Strategy* and *Surface-Motive*, and *Deep-Motive* and *Surface-Motive*. Results obtained from the model of Immekus and Imbrie did not provide overwhelming support – based on fit indices of CFI, RMSEA, and SRMR – for a four-factor model (Table 1). Issues of discriminant validity were also raised due to the strength of inter-factor correlations: .96 between *Deep-Strategy* and *Deep-Motive*, and .90 between *Surface-Strategy* and *Surface-Motive*. Based on such correlation values, a four factor model appears as an over-extraction of factors. This was the line of thinking adopted by Immekus and Imbrie, who evaluated a two-factor model (*Deep* and *Surface Approaches*). Contrary to what Justicia et al. (2008) reported, Immekus and Imbrie did not find support for a two-factor model (Table 1). The model eventually accepted by Immekus and Imbrie – reassessed using exploratory factor analysis – was a four-factor model with five items (1, 3, 7, 13, and 15) being dropped (Table 1).

Socha and Sigler (2014) tested a series of different model representations (first-order, hierarchical, and bi-factor) based on the aforementioned work of Biggs et al. (2001); Justicia et al. (2008); Immekus and Imbrie (2010). Models proposed by Biggs et al. were found to be non-admissible. The 15-item, four-factor model offered by Immekus and Imbrie met the commonly used Hu and Bentler (1999) criteria (CFI = .95, TLI = .95, RMSEA = .06, SRMR = .08) for acceptability (Table 1); however, factor definitions became distorted due to item loadings (Socha and Sigler). For Socha and Sigler, a two-factor model (*Deep* and *Surface Approaches*) was accepted with two items (7 and 8) being omitted as they were sources of local misfit; López-Aguado and Gutiérrez-Provecho (2018) also found support for this model (Table 1). Shahrazad, Sulaiman, and Dzulkifli (2013) find support for a two-factor (*Deep* and *Surface*) model, with six items being dropped based on modification indices and factor loadings (items 4, 5, 11, 12, 15, and 17; Table 1) . A limitation here is the use of modification indices that may reflect sampling fluctuations; thus, the model of Shahrazad et al. should be viewed with caution. Zakariya, Bjørkestøl, Nilsen, Goodchild, and Lorås (2020) offer an alternate two-factor (*Deep* and *Surface Approaches*) model with one item removed (item 8) based on having a nominal target factor loading ($\lambda_{Surface-Approach} = .06$).

Researchers have also explored the proposed hierarchical model proposed by Biggs et al. (2001), whereby two factors of *Deep* and *Surface* are each measured by two first-order factors (*Motive* and *Strategy*). Justicia et al. (2008), for instance, found the fit to meet what could be considered acceptable (Table 1). Socha and Sigler (2014), on the other hand, found this model to not converge. Zakariya et al. (2020) found the hierarchical model to converge, but the fit was inadequate (Table 1). In light of such modelling exercises, two points can hereby be made on the topics of model parsimony and identification:

1. if a two first-order factor model is supported, there is no cause to pursue a hierarchical model; *and*
2. for a hierarchical model to be identified, it must have a minimum of three first-order factors loading (Kline, 2015).

Pursuit of a hierarchical representation of learning approaches is therefore needless and precedence should be placed on clarifying the first-order factor model. This leads to the second research question:

- What factor model is the best representation for items of the R-SPQ-2F? (Research Question 2; RQ2)

1.3 Research Aim

The foremost aim of this work is to undertake a critical appraisal of the R-SPQ-2F factor structure (Biggs et al., 2001) and analytical steps undertaken by recent researchers (Justicia et al., 2008; Immekus & Imbrie, 2010; Shahrazad et al., 2013; Socha & Sigler, 2014; Xie, 2014; López-Aguado & Gutiérrez-Provecho, 2018; Zakariya et al., 2020). Although these more recent works have sought to explore a variety of model representations, there is little consensus as to what factor structure to agree upon. Results of such work have offered models containing different numbers of factors, different numbers of items, and different ways to represent the factors. Answers to the two aforementioned research questions are intended to evaluate the utility of the R-SPQ-2F in measuring what it is intended to measure, with a view of offering recommendations on how best to proceed.

2 Method

2.1 Sample

A total of 1,636 student (Undergraduate and Postgraduate) responses were initially collected using opportunity sampling from an Australian University. Of the 1636 responses, 456 (27.873%) contained missing values for all 20 items and 4 (.244%) responses contained missing student identification details. The 456 missing responses were omitted, in addition to those with 4 instances of missing personal details. Having personal details was a necessity for the analysis to determine how representative the sample was of the student population; thus,

Table 1: Alternative Model Representations of the R-SPQ-2F Factor Structure Evaluated by Other Researchers

Authors	Model Tested	χ^2	*df*	P-Value	CFI	TLI	RMSEA	SRMR
Justicia et al. (2008)	Four-factor model	667.83	168	$< .001$	.92	–	.07	.12
	Two-factor model	645.77	169	$< .001$	.92	–	.07	.09
	Hierarchical model	623.19	165	$< .001$	.93	–	.07	.09
Immekus and Imbrie (2010)	Four-factor model	598.60	79	$< .01$	.87	–	.06	.06
	Two-factor model	607.94	79	$< .01$	.86	–	.10	.06
	Four-factor model[a]	160.78	49	$< .01$	.96	–	.05	.04
Shahrazad et al. (2013)	Two-factor model[b]	111.48	73	$< .05$	.94	.93	.04	–
Socha and Sigler (2014)	Four-factor model[a]	282.336	84	$< .001$	.95	–	.05	.045
	Two-factor model[c]	504.826	134	$< .001$	.95	–	.06	.051
Xie (2014)	Four-factor model	463.64	164	$< .001$	.92	–	.065	–
	Two-factor model	489.40	169	$< .001$	.92	–	.067	–
López-Aguado and Gutiérrez-Provecho (2018)	Two-factor model[c]	226.53	134	$< .001$	.914	.900	.050	.052
Zakariya et al. (2020)	Two-factor model	522.179	169	$< .001$	.769	.740	.091	.081
	Two-factor model[d]	377.676	151	$< .001$	.844	.824	.077	.072
	Hierarchical model	521.114	168	$< .001$	.769	.739	.091	.081

[a]Removal of five items: 1, 3, 7, 13, and 15
[b]Removal of six items: 4, 5, 11, 12, 15, and 17
[c]Removal of two items: 7 and 8
[d]Removal of one item: 8

Table 2: Sample and 2019 Population Academic Discipline Enrolment Counts and Percentages

Academic Discipline	Sample		Population	
	n	%	*n*	%
Art, Design, and Architecture	21	1.81	2773	3.30
Arts	90	7.77	7954	9.47
Business and Economics	341	29.40	21007	25.00
Education	87	7.51	7079	8.42
Engineering	179	15.50	9011	10.70
Information Technology	137	11.80	7169	8.53
Law	32	2.76	4095	4.87
Medicine, Nursing, and Health Sciences	127	11.00	15893	18.90
Pharmacy and Pharmaceutical Sciences	–	–	2546	3.03
Science	144	12.40	6500	7.74

they were omitted. Eighteen (1.100%) responses were identified as being duplicates on the basis of the student identification information; all responses given by these students (original and duplicate) were dropped.

One-thousand and fifty eight (1,158) complete responses were retained for the analyses that followed. The gender breakdown for the sample was as follows: 737 (63.60%) Female, 419 (36.20%) Male, and 2 (.173%) Indeterminate/Intersex/Unspecified. To evaluate the sample characteristics, they can be compared to the student population from which they were sampled. Across the University, the gender counts and percentages are as follows: 47,629 (56.60%) Female, 36,410 (43.30%) Male, and 44 (.052%) Indeterminate/Intersex/Unspecified students. Ages of the respondents in the sample ranged from 17 to 53 (M = 21.60, SD = 4.74, *Skewness* = 3.25). For the student population, age ranged from 15 to 87 (M = 23.30, SD = 6.56, *Skewness* = 2.77; sample and population age distributions are presented in Appendix A.1). The number of Undergraduate students totalled 905 (78.20%), whilst 253 (21.80%) were Postgraduate students. The University population values for course levels were as follows: 4812 (5.72%, Higher Degree Research), 24,495 (29.10%, Postgraduate), 54,351 (64.60%, Undergraduate), and 425 (.505%, Non-Award). Of the 10 academic disciplines at the university, all but one (Pharmacy and Pharmaceutical Sciences) was represented (Table 2).

Comparing the sample to the 2019 Academic Discipline enrolment data (Table 2), it was clear that the student population was not well represented. In certain cases, Academic Disciplines were under-represented: *Art, Design, and Architecture*, *Arts*, *Education*, *Law*, *Medicine, Nursing, and Health Sciences*, and *Pharmacy and Pharmaceutical Sciences*. Whereas, the Academic Disciplines of *Business and Economics*, *Engineering*, *Information Technology*, and *Science* were over-represented in the sample.

2.2 Instrument

Data was collected using the 20-item R-SPQ-2F (Appendix A.2; Biggs et al., 2001). Slight amendments were made to the item wordings to align with the institutional context. In the original item wordings Biggs et al. referred to ***course*** (e.g., *My aim is to pass the **course** while doing as little work as possible*); for the context of this study, ***unit*** was a more appropriate wording choice (e.g., *My aim is to pass the **unit** while doing as little work as possible*). Each of the 20-items were answered on a 5-point Likert scale ranging from *This is never or only rarely true of me* (1) to *This is always or almost always true of me* (5).

2.3 Procedure

Survey links for the R-SPQ-2F (Biggs et al., 2001) were distributed to students (Undergraduates and Postgraduates) through direct emails. To receive the email, the student had to be enrolled within a course using live-streaming services. The R-SPQ-2F was used as a set of introductory items before a series of items designed to measure live streaming attitudes; results of the latter are not applicable to the current work. Completion of the survey was voluntary; respondents to the survey were eligible to participate in a voluntary prize-draw to win one of five $20 gift cards. Respondents were only able to access the survey by using their university credentials, which were logged and used to connect survey responses to demographic and educational information. An explanatory statement form was given to students before completing the survey detailing how their university credentials would be collected and used to connect responses to university records; consent was obtained from all respondents.

2.4 Analysis

To answer RQ1 and RQ2, three models (Figure 3) were tested using Confirmatory Factor Analysis (CFA) and Exploratory Structural Equation Modelling (ESEM; Figure 3). Response distributions for the 20 items (Appendix A.3) show that normality cannot be assumed, a commonality with ordinal level data. In light of the normality violation, the weighted least square mean and variance adjusted (WLSMV) estimator was used in place of the maximum likelihood estimator; delta scaling was also used. All analyses were run using Mplus 8.3 (Muthén & Muthén, 2017).

Each model tested was informed by the original work of Biggs et al. (2001) and the work that followed (Fryer et al., 2012; Immekus & Imbrie, 2010; Justicia et al., 2008; López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014; Xie, 2014). Figure 3a (Model 1) is the item-level model proposed by Biggs et al. where four factors (*Deep-Motive*, *Deep-Strategy*, *Surface-Motive*, and *Surface-Strategy*) are measured by five items. Correlations between factors are specified between Motive and Strategy within the same approach (*Surface-Motive* and *Surface-Strategy*); correlations across approaches are only specified for the *Deep-Motive* and *Surface-Motive* factors. Figure 3b (Model 2) deviates from

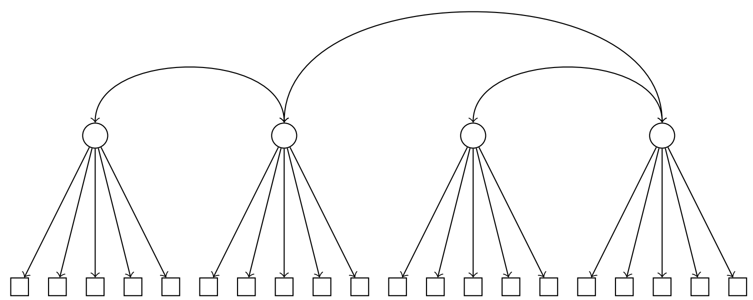
Table 3: Indicator to Factor Specification for 3 Tested Models (Item wordings presented in Appendix A.2)

Model	Factor	Items
1 & 2	Deep-Motive	1 5 9 13 17
	Deep-Strategy	2 6 10 14 18
	Surface-Motive	3 7 11 15 19
	Surface-Strategy	4 8 12 16 20
3	Deep Approach	1 2 5 6 9 10 13 14 17 18
	Surface Approach	3 4 7 8 11 12 15 16 19 20

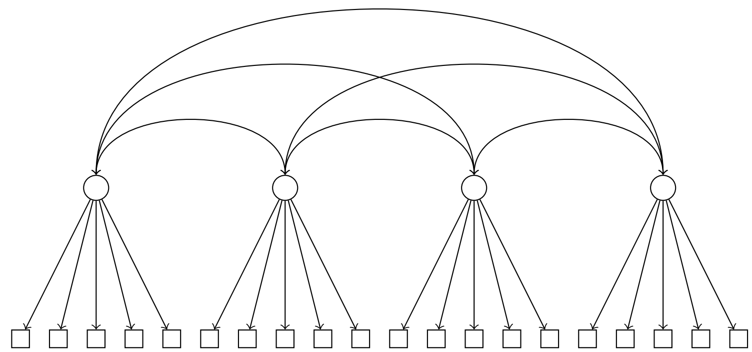
the prior model as correlations between the *Deep-Strategy* and *Surface-Strategy* are specified. This model is a reasonable adjustment to Figure 3a (Model 1), as undertaking a *Surface-Strategy* would be associated with a lower likelihood of strategies focused on obtaining a deep understanding of the subject (*Deep-Strategy*. Figure 3c (Model 3) is based on the findings of Justicia et al. (2008) that a two-factor (Deep and Surface) representation offers a better fit to the data over a four-factor model. It also accounts for the high correlation between *Motive* and *Strategy* factors that are reported (Immekus & Imbrie, 2010). The indicator to factor specification is presented in Table 3.

The models in Figure 3 are presented as confirmatory factor models with the assumption that there are no cross-loading factors and errors are not correlated. As discussed by Marsh, Morin, Parker, and Kaur (2014), the restriction of zero cross-loadings is too restrictive due to indicator fallibility. Marsh et al. instead advocate the use of ESEM, wherein the analyst can specify items to load onto multiple factors. This may be a suitable reflection of reality as items may have nominal loadings onto non-target factors; without specification, the latter may result in sources of ill-fit. The recommendation outlined by Marsh et al. (2014) was therefore followed: a confirmatory factor model was first tested, followed by an ESEM (RQ 1 and 2). ESEM was applied to all but Model 3a as a target oblique rotation procedure was used, wherein factor correlations are freely estimated. For the target rotation, the only factor loadings specified (= 0) were those, based on the work of Biggs et al. (2001) and Justicia et al. (2008), not loading onto the target factors. Those indicators that have target factors (Table 3) were not specified. Results obtained from the ESEM approach help answer RQ1 as they show whether items only measure a single factor or not. Thus, providing evidence on whether item-parcelling can be applied when analysing responses from the R-SPQ-2F.

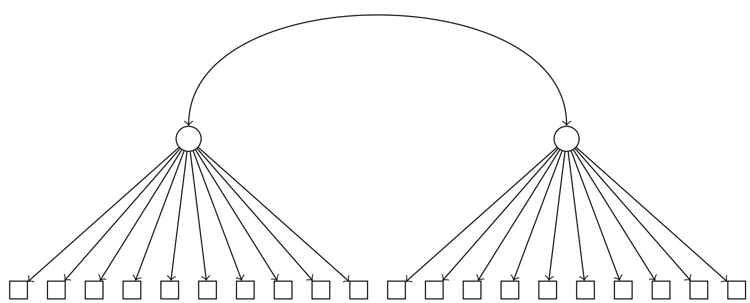
How the ESEM models differed from those presented in Figure 3 was the permittance of cross-loadings; no correlated errors were specified. Comparisons of both models (CFA and ESEM) were made on general fit measures (χ^2, CFI, TLI, RMSEA, and SRMR), local fit (residuals), and measurement quality (factor loadings). Although cut-offs are regularly cited to judge the *goodness* of fit (CFI = .95, TLI = .95, RMSEA = .06, SRMR = .08; Hu & Bentler, 1999), they are based on a narrow range of models that are not reflective of those



(a) Model proposed by Biggs et al. (2001) (Model 1). Four factors (DS = Deep-Strategy, DM = Deep-Motive, SS = Surface-Strategy, SM = Surface-Motive). Five indicators per factor. Freely estimated correlations between: DS and DM, DM and SM, and SS and SM.



(b) Adaptation of the Biggs et al. (2001) model with four freely correlated factors (Deep-Strategy, Deep-Motive, Surface-Strategy, Surface-Motive), each with five indicators (Model 2).



(c) Two-Factor (Deep and Surface) Model (Model 3). Freely estimated correlation between factors, each of which has 10 indicators.

Figure 3: Three Confirmatory Factor Models Tested in the Current Work

tested (Marsh, Hau, & Wen, 2004). Whilst remaining cognisant of such criteria, their interpretation requires an assessment of how well indicators measure the construct of interest and an exploration of possible sources of misfit. Identification of localised strain in the model was carried out by inspecting the residual correlations, using $\geq .10$ as a guideline (Kline, 2015); modification index (MI) and standardised/expected parameter change (S/EPC) values were also provided (Saris, Satorra, & Veld, 2009). This model evaluation provides an answer to RQ2 as the results can be used to determine whether a suitable candidate factor model (Model 1-3) can be agreed upon.

With 20 indicators, the number of observations totalled 270 (190 polychoric correlations and 80 thresholds). For Model 1, the confirmatory factor model required 103 parameters to be estimated (16 factor loadings, 80 thresholds, 3 factor covariances, and 4 factor variances), equating to 167 degrees of freedom (*df*). For Model 2, the confirmatory factor model estimated 106 parameters (16 factor loadings, 80 thresholds, 6 factor covariances, and 4 factor variances; *df* = 164). The ESEM model required 154 parameters to be estimated (74 factor loadings and 80 thresholds; *df* = 116). Model 3 required 101 paramaters to be freely estimated for the CFA (18 factor loadings, 80 thresholds, 1 factor covariance, and 2 factor variances; *df* = 169) and 119 for the ESEM (39 factor loading and 80 thresholds; *df* = 151). In all instances, the models were overidentified.

3 Results

Results are presented as follows: each proposed model (Model 1 to 3) are evaluated through both CFA and ESEM. RQ1 seeks to understand whether item parcelling is acceptable when analysing R-SPQ-2F data, the answer to which would be obtained through the use of ESEM. RQ2 focuses on the identification of a representative factor model, explored through both CFA and ESEM. Thus, the results are presented with a view of simultaneously addressing RQ1 and RQ2.

3.1 Model 1

The first confirmatory factor model (Model 1; Figure 3a) resulted in a latent variable covariance matrix that was not positive definite. A detailed inspection of the output showed estimated correlations between latent variables to exceed the value of one (*Deep-Motive* and *Deep-Strategy* = 1.008; *Surface-Motive* and *Surface-Strategy* = 1.032).

3.2 Model 2

3.2.1 CFA Model

When all inter-factor correlations are specified (Model 2; Figure 3b) in the confirmatory factor model, the results were again not positive definite. This was attributed to the correlations between the constructs of a particular approach exceeding 1 (*Deep-Motive* and *Deep-Strategy* = 1.008; *Surface-Motive* and *Surface-Strategy* = 1.030).

3.2.2 ESEM Model

The ESEM version of Model 2 converged; $\chi^2(116) = 695.729$, $p < .001$, CFI = .948, TLI = .915, RMSEA = .066 (90% CI = .061—.070), SRMR = .026. Inter-factor correlations were small (Table 4). A low to moderate amount of the underlying continuous latent response variables was accounted for by the factors (R^2 = .309—636). Nine instances of residual variances being > 50% were identified (Table 5).

Factor loadings are presented in Table 5, showing the following absolute value ranges:

- .021—.589 (M = .276; *Deep-Motive*);
- .082—.670 (M = .282; *Deep-Strategy*);
- .005—.670 (M = .274; *Surface-Motive*); *and*
- .013—.706 (M = .216; *Surface-Strategy*).

Focusing on target factor loadings only, $|\lambda|$ values ranged from:

- .172—.589 (M = .453; *Deep-Motive*);
- .082—.670 (M = .404; *Deep-Strategy*);
- .202—.670 (M = .524; *Surface-Motive*); *and*
- .240—.706 (M = .385; *Surface-Strategy*).

An inspection of the residual correlation matrix (Appendix) showed three locations with absolute values $\geq$ .10. The areas of localised strain are presented in Table 6, along with MI and S/EPC values.

3.3 Model 3

3.3.1 CFA Model

Model 3 (Figure 3c), as a confirmatory factor model, successfully converged and was positive definite. The obtained fit was as follows: $\chi^2(169) = 1844.252$, $p < .001$, CFI = .851, TLI = .832, RMSEA = .093 (90% CI = .089—.096), and SRMR = .066. Both unstandardised and standardised factor loadings for the model are presented in Table 7. Standardised loadings ranged from:

- .513—.769 (M = .662; *Deep Approach*); *and*
- .513—.733 (M = .618; *Surface Approach*).

Table 4: Factor Correlation Matrix for the ESEM Version of Model 2 (Figure 3b)

	Deep-Motive	Deep-Strategy	Surface-Motive	Surface-Strategy
Deep-Motive	1			
Deep-Strategy	.332	1		
Surface-Motive	-.125	-.200	1	
Surface-Strategy	.103	.127	.386	1

Table 5: ESEM Factor Loadings for Model 2 (Figure 3b)

Items	Deep-Motive		Deep-Strategy		Surface-Motive		Surface-Strategy		δ
	λ	S.E.	λ	S.E.	λ	S.E.	λ	S.E.	
1	.589	.029	.111	.030	-.217	.029	.147	.031	.489
2	.574	.032	.082	.039	-.215	.035	.139	.040	.533
3	-.066	.033	.137	.035	.626	.033	.078	.033	.570
4	.147	.030	-.262	.031	.334	.044	.240	.058	.690
5	.517	.032	.250	.033	.022	.029	-.036	.030	.595
6	.445	.033	.558	.035	.114	.025	-.097	.028	.372
7	-.206	.030	.097	.028	.670	.029	.067	.027	.466
8	-.096	.037	.139	.062	.005	.030	.706	.038	.468
9	.451	.029	.422	.026	-.041	.026	-.013	.029	.481
10	.403	.035	.253	.047	-.248	.028	.315	.030	.509
11	-.198	.037	.221	.060	.202	.042	.648	.036	.378
12	.045	.024	-.103	.027	.410	.035	.371	.040	.561
13	.535	.025	.342	.024	-.166	.022	.033	.024	.400
14	.326	.041	.670	.039	.182	.026	-.141	.030	.364
15	-.094	.029	.186	.028	.600	.030	.208	.029	.493
16	.297	.030	-.360	.035	.465	.048	.270	.072	.443
17	.172	.038	.559	.029	.066	.035	.062	.045	.589
18	.154	.029	.455	.031	-.039	.032	.102	.038	.691
19	.176	.029	-.307	.028	.524	.046	.299	.059	.397
20	-.021	.027	.121	.031	.324	.032	.339	.034	.687

Table 6: Areas of Localised Strain for ESEM Model 2

Items		Residual Correlation	MI	EPC	SEPC
3	4	.111	50.903	.162	.258
4	12	.100	64.774	.148	.238
17	18	.103	74.832	.160	.251

The two factors of *Deep Approach* and *Surface Approach* were not strongly correlated (r = -.107). A low to moderate amount of the variance in the latent continuous response variables is accounted for by the factors (R^2 = .263—.592). Residual variances for 15 of the 20 indicators showed the factors to account for less than 50% of the latent response variable (Table 7).

Residual correlations for the CFA model are presented in Appendix A.5, wherein 41 sources of ill-fit can be identified using $\geq$ .10 as a cut-off. Table 8 presents the residual correlations that were $\geq$ .10, along with MI, EPC, and SEPC values.

3.3.2 ESEM Model

The ESEM version of Model 3 converged and achieved a slightly improved global fit over the confirmatory factor model, $\chi^2(151) = 1607.452$, $p < .001$, CFI = .870, TLI = .837, RMSEA = .091 (90% CI = .087—.095), and SRMR = .043. The inter-factor correlation was found to be low (-.090). Loadings are presented in Table 9, with $|\lambda|$ values ranging from:

- .020—.771 (M = .383; *Deep Approach*); *and*
- .000—.708 (M = .337; *Surface Approach*).

For target factor loadings, the $|\lambda|$ values ranged from:

- .527—.771 (M = .662; *Deep Approach*); *and*
- .486—.708 (M = .618; *Surface Approach*).

A low to moderate amount of the underlying continuous latent response variable was accounted for by the factors (R^2 range = .252—.591). There were 15 instances of residual variance values being > 50% (Table 9).

The residual correlation matrix (Appendix A.6) showed nine locations where absolute values $\geq$ .10. Table 10 presents the locations at which this localised strain within the model occurs, along with MI, EPC, and SEPC values.

4 Discussion

As shown in the presented results, the adoption of item parcelling is not suitable when utilising the R-SPQ-2F (RQ 1). Added to this is the failure to identify a suitable model that aligns with the original work of Biggs et al. (2001), which

Table 7: Unstandardised and Standardised CFA Factor Loadings for Model 3

	Unstandardised Solution				Standardised Solution				
Items	Deep Approach[a]		Surface Approach[b]		Deep Approach		Surface Approach		δ
	λ	S.E.	λ	S.E.	λ	S.E.	λ	S.E.	
DA1	1.000	.000	-	-	.661	.019	-	-	.563
DA2	.947	.035	-	-	.626	.020	-	-	.608
DA3	.931	.036	-	-	.615	.020	-	-	.622
DA4	1.160	.037	-	-	.767	.014	-	-	.412
DA5	1.094	.039	-	-	.723	.017	-	-	.477
DA6	.969	.037	-	-	.640	.019	-	-	.590
DA7	1.164	.037	-	-	.769	.015	-	-	.408
DA8	1.101	.039	-	-	.728	.016	-	-	.471
DA9	.871	.039	-	-	.575	.021	-	-	.669
DA10	.776	.040	-	-	.513	.022	-	-	.737
SA1	-	-	1.000	.000	-	-	.606	.023	.633
SA2	-	-	.847	.045	-	-	.513	.024	.737
SA3	-	-	1.086	.049	-	-	.658	.021	.567
SA4	-	-	.851	.047	-	-	.516	.023	.734
SA5	-	-	1.063	.050	-	-	.644	.020	.585
SA6	-	-	1.092	.048	-	-	.662	.018	.562
SA7	-	-	1.101	.049	-	-	.667	.022	.554
SA8	-	-	1.081	.049	-	-	.655	.019	.571
SA9	-	-	1.209	.052	-	-	.733	.016	.463
SA10	-	-	.860	.048	-	-	.521	.023	.728

[a]DA (Deep Approach) = Items (in order) 1, 2, 5, 6, 9, 10, 13, 14, 17, 18
[b]SA (Surface Approach) = Items (in order) 3, 4, 7, 8, 11, 12, 15, 16, 19, 20

Table 8: Areas of Localised Strain for CFA Model 3

Items		Residual Correlation	MI	EPC	SEPC
1	2	.141	62.226	.191	.326
1	7	-.177	39.251	-.179	-.317
1	8	.135	20.937	.136	.212
2	7	-.184	40.513	-.187	-.318
2	8	.114	15.096	.115	.172
3	7	.108	29.283	.152	.253
3	13	-.129	19.828	-.131	-.257
3	16	-.101	13.462	-.123	-.204
3	19	-.105	17.648	-.136	-.251
4	12	.107	28.080	.141	.220
4	14	-.109	14.859	-.110	-.187
4	15	-.103	14.219	-.123	-.192
5	7	-.130	20.280	-.132	-.222
6	8	.160	32.555	.161	.294
6	11	.118	16.081	.119	.243
6	20	.124	18.070	.125	.229
7	8	-.112	18.053	-.134	-.208
7	9	-.105	13.441	-.107	-.205
7	13	-.270	120.434	-.277	-.575
8	9	.115	16.362	.116	.197
8	10	.281	108.881	.283	.431
8	11	.226	186.216	.322	.492
8	14	.111	15.195	.112	.190
8	17	.184	43.491	.185	.264
8	18	.164	30.085	.164	.224
9	19	-.135	23.851	-.137	-.292
10	11	.172	34.945	.174	.297
11	14	.153	28.953	.155	.296
11	16	-.141	30.964	-.176	-.305
11	17	.177	40.210	.179	.286
11	18	.129	20.284	.130	.198
11	19	-.108	22.388	-.145	-.278
11	20	.101	22.224	.130	.199
13	19	-.133	24.152	-.136	-.313
14	15	.129	20.217	.131	.256
14	20	.135	23.203	.136	.233
15	17	.152	25.987	.154	.253
16	19	.152	137.765	.284	.552
17	18	.160	59.159	.193	.275
17	20	.114	15.203	.114	.164
18	20	.143	22.925	.144	.196

Table 9: ESEM Factor Loadings for Model 3

Items	Deep Approach		Surface Approach		δ
	λ	S.E.	λ	S.E.	
1	.648	.020	-.061	.022	.569
2	.612	.020	-.064	.021	.615
3	-.039	.024	.599	.024	.635
4	-.087	.023	.486	.024	.748
5	.610	.020	.000	.021	.628
6	.771	.013	.027	.017	.409
7	-.190	.021	.626	.022	.551
8	.226	.022	.567	.021	.651
9	.714	.017	-.039	.017	.483
10	.655	.018	.043	.019	.574
11	.165	.021	.683	.018	.526
12	-.021	.020	.654	.018	.570
13	.751	.016	-.109	.016	.409
14	.733	.016	.048	.017	.466
15	.020	.021	.677	.021	.543
16	-.062	.021	.636	.019	.585
17	.596	.020	.107	.021	.645
18	.527	.022	.049	.022	.725
19	-.120	.019	.708	.017	.469
20	.116	.023	.549	.022	.696

Table 10: Areas of Localised Strain for ESEM Model 3

Items		Residual Correlation	MI	EPC	SEPC
1	2	.147	123.885	.196	.332
3	7	.112	57.996	.156	.263
4	12	.121	65.949	.158	242
8	10	.107	36.658	.134	.220
8	11	.156	184.341	.249	.425
11	16	-.137	55.602	-.175	-.315
13	16	.101	28.853	.128	.261
16	19	.164	288.489	.296	.566
17	18	.144	90.215	.176	.258

raises questions as to what is being measured by the R-SPQ-2F (RQ 2). What follows is a discussion pertaining to each research question with a view of fleshing out the study results and offering directions for future work.

4.1 Item Parcelling and the R-SPQ-2F

Under particular conditions, the application of item parcelling may be justified such as when items exclusively load onto a target factor (i.e., no cross-loadings; Marsh et al., 2013). The present work sought to investigate whether this was a viable approach by comparing CFA and ESEM representations of the Biggs et al. (2001) model (Figure 3b). When modelled using CFA, with cross-loadings fixed at zero, the tested model failed to converge. An admissible solution was instead found when ESEM was applied, which also permitted cross-loadings to be freely estimated. The latter was advantageous as it would uncover areas of misspecifications (non-target factor loadings) that would go otherwise undetected. For the investigation into whether item parcelling was an acceptable undertaking when using the R-SPQ-2F, the specification of cross-loadings was a necessity to assess unidimensionality and measurement quality.

Output of the ESEM model factor loadings indicated a number of items that do not have a clear target factor, despite this being previously hypothesised (Biggs et al., 2001). For Item 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*), the hypothesised target factor was *Deep-Strategy*. The results fail to support this as Item 2 was found to moderately load onto *Deep-Motive* ($\lambda = .574$), but only a target factor (*Deep-Strategy*) loading of .082. An examination of the item wording may help in understanding why this item does not work as intended. For one, the item asks about satisfaction, similar to Item 1 (I find that at times studying gives me a feeling of deep personal satisfaction). It is possible that there is a wording effect due to the items being framed around the feeling of satisfaction that arises from studying. Alternatively, Item 2 could be interpreted as measuring the *Deep-Motive* factor given that it essentially asks why the student approaches the task (i.e., to form conclusions). Item 2 cannot then be viewed as an acceptable indicator of *Deep-Strategy*.

Item 4 (*I only study seriously what's given out in class or in the unit guide*) was found to be a poor indicator of its target factor (*Surface-Strategy*; $\lambda = .240$). A slightly higher loading for this item was found for the factor of *Surface-Motive* ($\lambda = .334$). In this instance, Item 4 cannot be viewed as a good measure of either the *Surface-Strategy* or *Surface-Motive* factors, even more so with a residual variance of 69.40%. Wording of this item does not clearly align with either: *Strategy*, how a task is approached; or *Motive*, why is a task approached. The item asks if the student studies seriously, which does not convey any explicit details on the strategy itself, nor the motive. As a whole the item intention is to explore whether students go beyond the material covered in class or not, but this may not appear to be the case.

Proposed as an indicator of *Deep-Strategy*, Item 6 (*I find most new topics interesting and often spend extra time trying to obtain more information about* *them*) showed similar issues of moderate loadings on the target factor ($\lambda =$ .558) and a non-target factor (*Deep-Motive*; $\lambda = .445$). As Item 6 asks students about their interest in a topic, this clearly taps into the *Motive* facet as it is asking why the student is approaching the task. The latter half of the item does explore the *Strategy* factor by asking how they approach the task, specifically by allocating more time. Thus, the item problems can be attributed to it being 'double-barrelled', resulting in students answering either the *Motive* or *Strategy* aspect of the item (Bandalos, 2018). It is for this reason that Item 6 cannot be regarded as a well functioning with regards to measuring a single factor.

Moderate cross-loading was found with Item 9 (*I find that studying academic topics can at times be as exciting as a good novel or a movie*), with a value of .422 for *Deep-Strategy*; a loading value of .451 was reported for the target factor (*Deep-Motive*). While this item appears to touch upon motivations for the approach, specifically the excitement students experience whilst studying, it is not explicit in asking whether this is the underlying reason. There may be good reason to view this item as measuring *Deep-Strategy* as it seeks to evaluate the strategy used, rather than ask why it is used.

Item 10 (*I test myself on important topics until I understand them completely*) was identified as problematic for two reasons: one, it had a low target factor loading ($\lambda_{Deep-Strategy} = .253$); and two, low to moderate non-target factor loadings ($\lambda_{Deep-Motive} = .403$; $\lambda_{Surface-Strategy} = .315$). From the interpretation of the item alone, it could be argued that Item 10 is a *Surface-Strategy* as it appears to ask if students repeat a behaviour (testing) until understanding is achieved. Having a larger loading on *Deep-Motive* could be attributed to the item asking about wanting a complete understanding of the topic. In this sense, the item is framed as a *Motive* that looks to explore the reasoning behind the adopted approach (i.e., to improve topic understanding) more than how it was approached. Put differently, the terminology of testing is rather vague given the various manners in which someone can test themselves about a topic (e.g., multiple choice questions or essay questions).

Factor loading values for Item 12 (*I generally restrict my study to what is specifically set as I think it is unnecessary to do anything extra*) raise questions as to what is being measured. The highest factor loading was for a non-target factor ($\lambda_{Surface-Motive} = .410$), although the target factor loading value was moderate ($\lambda_{Surface-Strategy} = .371$). Wording of this item can be argued as problematic as it encompasses two themes: *Motive* and *Strategy.* The first part of the statement asks about how the student approaches their studying, i.e. restricting the amount studied. The second half, however, asks about the underlying motivation for the approach adopted (the student sees any additional work as unnecessary). There are two things essentially being asked of students with Item 12, what is there approach (*Strategy*) and why adopt this approach (*Motive*).

With Item 16 (*I believe that lecturers shouldn't expect students to spend significant amounts of time studying material*), loadings are higher for non-target factors ($\lambda_{Deep-Motive} = .297$; $\lambda_{Surface-Motive} = .465$) than the target factor ($\lambda_{Surface-Strategy} = .270$). A moderate loading on the *Surface-Motive* factor is understandable as the item can be interpreted as asking students their motivations for adopting a particular approach (the belief that lecturers should not expect such workload levels from students). For the *Deep-Motive* factor, item 16's loading is not straightforward. There is a possibility that this presents a difficult item for students to respond to as it is negatively phrased. For example, a student who agrees with the statement has to respond with '*This is never or only rarely true of me*'. It may be necessary for work to explore how this particular item is interpreted by respondents as a way to explore if there are any associated difficulties. Even so, as a measure of *Surface-Strategy* Item 16 does not appear to be particularly good.

Finally, Item 20 (*I find the best way to pass examinations is to try to remember answers to likely questions*) had low loadings on its target factor (*Surface-Strategy*; $\lambda = .339$), comparable to the non-target factor loading on *Surface-Motive* ($\lambda = .324$). This is another instance where an item can be viewed as tapping into two constructs: there is an element of *Motive* by framing the statement around passing an exam, which is followed by the *Strategy* of focusing on likely questions.

Based on the above-mentioned discussion, the practice of item-parcelling as used by Biggs et al. (2001) is not supported. There are multiple instances where the assumption of unidimensionality is violated, with items having non-target factor loadings that exceed the values of target factor loadings. In light of the results presented here, the use of item-parcels in Biggs et al. (2001) only masked localised strains in the model, all for the purpose of obtaining a good fit.

The recommendations for future work are to analyse R-SPQ-2F at an item level and do not assume that the indicators have negligible cross-loadings. As the original work of Biggs et al. (2001) adopted the questionable practice of analysing each factor in isolation of one another, it thereby ignored cross-loadings (Marsh et al., 2013). This is concerning as it has invariably led to researchers continuing this practice to create scores (e.g., Bliuc et al., 2010) without acknowledgement of potential misspecifications. As a result, the findings of this previous literature may reflect biased estimates (Marsh et al., 2013). It becomes important that item-parcelling becomes discouraged when analysing responses to the R-SPQ-2F, unless they can demonstrate that the assumption of unidimensionality is supported.

4.2 Model Representation and the R-SPQ-2F

Three models were tested – informed by prior work using the R-SPQ-2F – with a view of establishing whether a suitable candidate model could be accepted. Unlike previous instances of model evaluations, ESEM was used in conjunction with CFA. This approach was followed as ESEM can identify localised sources of ill-fit (i.e., cross-loadings) that may be overlooked in the case of confirmatory factor models.

Based on the results presented, the original model (Figure 1) of Biggs et al. (2001) can be out right rejected as the solution was nonadmissible. This was attributed to the inter-factor correlations exceeding the value of 1, which was also identified by Zakariya et al. (2020). For these authors, they report the following values of 1.018 between *Deep-Motive* and *Deep Surface* and 1.048 between *Surface-Motive* and *Surface-Strategy*, similar to those reported here. Socha and Sigler (2014) have also found there inter-factor correlations to exceed the value of 1 when the proposed model of Biggs et al. (2001) is evaluated.

When correlations are not fixed at zero – as in Figure 1 – an admissible solution is found, but only in the ESEM representation. A possible reason behind the CFA model not converging could be a result of all cross-loadings being fixed at zero, resulting in an inflated correlation between factors (Marsh et al., 2014). Thus, by freely estimating item cross-loadings, the ESEM model converged. As discussed above, the ESEM model provides information about the indicator fallibility that has not been previously discussed by other authors (e.g., Justicia et al., 2008; Socha & Sigler, 2014; Zakariya et al., 2020). Whilst the global fit indices may suggest a permissible model, the construct validity of the four factors can be questioned based on the pattern coefficients. In this case, it is wrong to accept a particular model off the reported fit indices, particularly when the indicators are not measuring the intended constructs. The unidimensionality originally proposed by Biggs et al. (2001) can also be viewed as questionable given the reported cross-loadings. We therefore reject Model 2 (Figure 3b) as a suitable representation of the data and advise that the undertaking of item-parcelling should not be carried out.

Model 3 (Figure 3c) was found to be admissible for both the ESEM and CFA approaches, allowing for a comparison of model fits. Starting with the global fit, the ESEM does show nominal differences to the CFA model, suggesting that freely estimating cross-loadings is an admissible approach. Differences between these two models become more apparent when inspecting the local fit, with the CFA model having a substantially large amount of strains compared to the ESEM model. There may be an inclination to assume that the ESEM model would be accepted over the CFA model, yet the issue of measurement quality cannot be ignored. For either model, the majority of residual variance values for indicators exceeded 50%, indicating that most of the variance in the underlying continuous latent variable was not well captured by the factors. From this perspective, the utility of the R-SPQ-2F in measuring *Deep* and *Surface Approaches* needs to scrutinised, as we cannot be assured that indicators are measuring what is theorised.

Other authors have settled upon a two-factor solution (López-Aguado & Gutiérrez-Provecho, 2018; Shahrazad et al., 2013; Socha & Sigler, 2014; Zakariya et al., 2020) of *Deep* and *Surface Approaches*, following an evaluation of various model representations. Whilst these different works converge upon the same model, they are not comparable. López-Aguado and Gutiérrez-Provecho (2018) and Socha and Sigler (2014) dropped two items (7 and 8) from the model, Shahrazad et al. (2013) removed six items (4, 5, 11, 12, 15, and 17), and Zakariya et al. (2020) omitted only item 8. Across four studies, there appears to be an inconsistent approach of researchers dropping items for the pursuit of a good fitting model, without any attempt to build a reasonable argument to support the approach. In certain cases (e.g., Shahrazad et al., 2013) based decisions on modifications, an approach that capitalises on chance (MacCallum et al., 1992). Our approach was solely confirmatory, with a view of evaluating the suitability of the proposed factor models for the R-SPQ-2F without undertaking any post-hoc model modifications. Having an array of inconsistent model representations, with varying numbers of items, does little to facilitate progression nor does it help users of the questionnaire. At most, the current work shows that the a two-factor solution offers a better solution compared to four factors. As for the items, there are clear problems that may be attributable to how they were written, attested by the number of non-negligible cross-loadings. Therefore, a full revision of the R-SPQ-2F is warranted with a view to address such issues.

4.3 Biggs' Model of Learning Approaches

Where does this leave the model 3P model of student learning proposed by Biggs (1987)? Based on the measurement tool developed on the basis of this framework, there is little support for a *Motive-Strategy* decomposition; this is not exclusive to the current work (e.g., Socha & Sigler, 2014). As for measuring the general constructs of *Deep Approach* and *Surface Approach*, this remains unclear. Although a permissible model, the two-factor representation cannot readily be accepted as there are issues associated with the indicator variables. At this point, the central facet of the 3P model (Process) cannot be measured well, thus raising questions about the model in general.

A critique of the 3P model is offered by Howie and Bagnall (2013), who describe it as being within an '*underdeveloped state*' (p. 4). Reasoning for the model development lag is attributable to a general readiness to accept such superficial categorisations of students (*Deep Approach* or *Surface Approach* learners) at face value, without a rigorous assessment of its claims. For example, the approach a learner takes is typically only discussed with reference to academic outcomes, resulting in a good-bad dichotomy represented by *Deep* and *Surface*, respectively. Contradiction of *Surface Approaches* having negative connotations for grades comes from the work of Kember (2000), who challenged this naïve line of thinking. Research into approaches to learning inadvertently created a stereotype of Asian students, characterising them as *rote learners* (Howie & Bagnall, 2013). Academic performance of these students were not consistent with Biggs' model as they were performing well, although it was claimed that the memorisation technique was not strictly a *Surface Approach* (Kember, 2000). This again brings issues of conceptual clarity and the appropriateness of the terminology beyond the context in which it was originally developed.

An additional complication of Biggs' conceptualisation of *Deep* and *Surface Learning* has been the view that learning approaches are fixed. Consider the work by Biggs (1989) that discusses student profiles, based on the results of their developed instruments. This would suggest that the R-SPQ-2F can be effectively used to cluster students into distinct groups based on a single data collection point. Such lines of thinking have now flipped, with ? (?) now stating that the constructs of *Deep* and *Surface* cannot be regarded as personality traits. Instead, approaches are viewed as dynamic reactions of the students to certain conditions within the teaching environment.

Not only are there measurement issues associated with the R-SPQ-2F, which is further compounded by a model that contains numerous inconsistencies. There is a need to re-address the 3P Model of Biggs (1989) and underdeveloped areas could be rectified to offset such practices as the static labelling of students or the creation of good-bad learning strategy dichotomies. As for measuring the *Process* element (*Strategy* and *Motive*) of the 3P model (Biggs, 1989), the use of questionnaires may not be entirely appropriate given that learning unfolds over time (Winne & Perry, 2000). Take for instance the strategies enacted by a student such as making notes, updating goals, and watching a lecture video, yet they are not easily captured by conventional self-report measures. An alternative approach would be to leverage event data as this captures the operations performed as students engage with a task (Winne & Perry, 2000). Instances of trace data being used this way can be found in learning analytic methodologies, wherein data mining techniques are applied to describe patterns of strategy enactments (Fincham, Gašević, Jovanović, & Pardo, 2019; Matcha, Gašević, Uzir, Jovanović, & Pardo, 2019; Saint, Gašević, & Pardo, 2018). While offering a perspective of strategy usage not easily attainable through questionnaire approaches such as the R-SPQ-2F – a sense of dynamism – they still become reductive by creating static profiles of students, as exemplified in Tempelaar, Rienties, Mittelmeier, and Nguyen (2018). It was previously discussed that profiling students could result in an interpretation that such 'profiles' are traits, as opposed to dynamic reactions to various conditions (internal and external). There is then a risk that learning analytic approaches may be applied in the same reductive manner as the R-SPQ-2F; that is, to categorise students, as opposed to understanding the learning processes of students.

These aforementioned data types and associated analyses are potential avenues for work associated with understanding the *Process* facet of the 3P Model (Biggs, 1989). However, the underdevelopment of the model remains an important shortcoming as the constructs are ill-defined, resulting in uncertainty over what being measured. Progress can only be made if the model is revised in order to address its fundamental weaknesses, in addition to a re-visitation of how such theorised constructs should be measured.

4.4 Limitations

Data was obtained through the use of opportunity sampling, as opposed to the use of stratified or random sampling approaches. Based on the sample and population demographic breakdowns, it was clear that the sample was not representative. Follow-up work should seek to obtain a more representative sample of the student population, which is also large enough to conduct measurement invariance tests across various sub-groups.

References

Bandalos, D. L. (2018). *Measurement Theory and Applications for the Social Sciences.* New York, UNITED STATES: Guilford Publications.

Biggs, J. B. (1970, August). Faculty patterns in study behaviour. *Australian Journal of Psychology*, *22*(2), 161–174. doi: 10.1080/00049537008254570

Biggs, J. B. (1987). *Student Approaches to Learning and Studying. Research Monograph.* Australian Council for Educational Research Ltd.

Biggs, J. B. (1989, January). Approaches to the Enhancement of Tertiary Teaching. *Higher Education Research & Development*, *8*(1), 7–25. doi: 10.1080/0729436890080102

Biggs, J. B., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, *71*(1), 133–149. doi: 10.1348/000709901158433

Bliuc, A.-M., Ellis, R., Goodyear, P., & Piggott, L. (2010). Learning through face-to-face and online discussions: Associations between students' conceptions, approaches and academic performance in political science. *British Journal of Educational Technology*, *41*(3), 512–524. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8535.2009.00966.x) doi: 10.1111/j.1467-8535.2009.00966.x

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245. doi: 10.2307/1170684

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3 2 achievement goal model. *Journal of Educational Psychology*, *103*(3), 632–648. doi: 10.1037/a0023952

Fincham, E., Gašević, D., Jovanović, J., & Pardo, A. (2019, January). From Study Tactics to Learning Strategies: An Analytical Method for Extracting Interpretable Representations. *IEEE Transactions on Learning Technologies*, *12*(1), 59–72. (Conference Name: IEEE Transactions on Learning Technologies) doi: 10.1109/TLT.2018.2823317

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. *British Journal of Educational Psychology*, *82*(4), 549–563. doi: 10.1111/j.2044-8279.2011.02045.x

Ginns, P., & Ellis, R. (2007, January). Quality in blended learning: Exploring the relationships between on-line and face-to-face teaching and learning. *The Internet and Higher Education*, *10*(1), 53–64. doi: 10.1016/j.iheduc.2006.10.003

Howie, P., & Bagnall, R. (2013, May). A critique of the deep and surface approaches to learning model. *Teaching in Higher Education*, *18*(4), 389–400. (Publisher: Routledge _eprint: https://doi.org/10.1080/13562517.2012.733689) doi: 10.1080/13562517.2012.733689

Hu, L., & Bentler, P. M. (1999, January). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives.

Structural Equation Modeling: A Multidisciplinary Journal, *6*(1), 1–55. doi: 10.1080/10705519909540118

Immekus, J. C., & Imbrie, P. (2010, June). 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. doi: 10.1177/0013164409355685

Justicia, F., Pichardo, M. C., Cano, F., Berbén, A. B. G., & De la Fuente, J. (2008, September). 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. doi: 10.1007/BF03173004

Kember, D. (2000). Misconceptions about the Learning Approaches, Motivation and Study Practices of Asian Students. *Higher Education*, *40*(1), 99–121. (Publisher: Springer)

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling, Fourth Edition.* Guilford Publications. (Google-Books-ID: HyavCgAAQBAJ)

López-Aguado, M., & Gutiérrez-Provecho, L. (2018, January). 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*, *30*(1), 105–141. doi: 10.1080/11356405.2017.1416787

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992, May). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*(3), 490–504. doi: 10.1037/0033-2909.111.3.490

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004, July). 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. doi: 10.1207/s15328007sem1103$_2$

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & von Davier, M. (2013, September). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—Camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, *18*(3), 257–284. doi: 10.1037/a0032773

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annual review of clinical psychology*, *10*, 85–110. doi: 10.1146/annurev-clinpsy-032813-153700

Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., & Pardo, A. (2019, March). Analytics of Learning Strategies: Associations with Academic Performance and Feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 461–470). Tempe, AZ, USA: Association for Computing Machinery. doi: 10.1145/3303772.3303787

Muthén, L., & Muthén, B. (2017). *Mplus User's Guide* (Eighth Edition ed.). Los Angeles, CA: Muthén & Muthén.

Parpala, A., & Lindblom-Ylänne, S. (2012, November). Using a research instrument for developing quality at the university. *Quality in Higher Education*, *18*(3), 313–328. (Publisher: Routledge _eprint: https://doi.org/10.1080/13538322.2012.733493) doi: 10.1080/13538322.2012.733493

Richardson, J. T. E. (2004, December). Methodological Issues in Questionnaire-Based Research on Student Learning in Higher Education. *Educational Psychology Review*, *16*(4), 347–358. doi: 10.1007/s10648-004-0004-z

Saint, J., Gašević, D., & Pardo, A. (2018). Detecting Learning Strategies Through Process Mining. In V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, & M. Scheffel (Eds.), *Lifelong Technology-Enhanced Learning* (pp. 385–398). Cham: Springer International Publishing. doi: 10.1007/978-3-319-98572-5$_2$9

Saris, W. E., Satorra, A., & Veld, W. M. v. d. (2009, October). Testing Structural Equation Models or Detection of Misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582. (Publisher: Routledge _eprint: https://doi.org/10.1080/10705510903203433) doi: 10.1080/10705510903203433

Shahrazad, W. S. W., Sulaiman, W. S. W., & Dzulkifli, M. A. (2013). Reliability of Second-order Factor of a Revised Two-factor Study Process Questionnaire (R-SPQ-2F) among University Students in Malaysia. *ASEAN Journal of Teaching & Learning in Higher Education*, *5*(2), 1–13.

Socha, A., & Sigler, E. A. (2014, April). Exploring and "reconciling" the factor structure for the Revised Two-factor Study Process Questionnaire. *Learning and Individual Differences*, *31*, 43–50. doi: 10.1016/j.lindif.2013.12.010

Tavakol, M., & Dennick, R. (2011, June). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. doi: 10.5116/ijme.4dfb.8dfd

Tempelaar, D., Rienties, B., Mittelmeier, J., & Nguyen, Q. (2018, January). Student profiling in a dispositional learning analytics application using formative assessment. *Computers in Human Behavior*, *78*, 408–420. doi: 10.1016/j.chb.2017.08.010

Tock, J. L., & Moxley, J. H. (2017, April). A comprehensive reanalysis of the metacognitive self-regulation scale from the MSLQ. *Metacognition and Learning*, *12*(1), 79–111. doi: 10.1007/s11409-016-9161-y

Winne, P. H., & Hadwin, A. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: Lawrence Erlbaum.

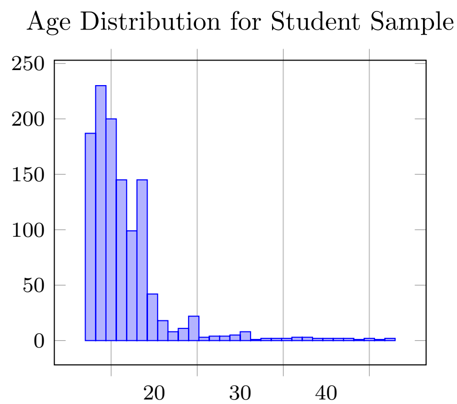
Winne, P. H., & Perry, N. E. (2000, January). Measuring Self-Regulated Learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 531–566). San Diego: Academic Press. doi: 10.1016/B978-012109890-2/50045-7

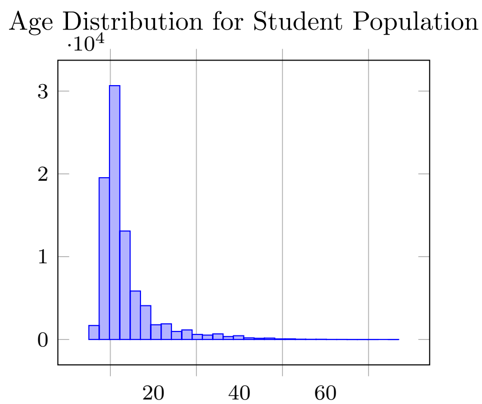
Xie, Q. (2014). Validating the Revised Two-Factor Study Process Questionnaire among Chinese University Students. *The International Journal of Educational and Psychological Assessment*, *16*(1), 18.

Zakariya, Y. F., Bjørkestøl, K., Nilsen, H., Goodchild, S., & Lorås, M. (2020, March). University students' learning approaches: An adaptation of the revised two-factor study process questionnaire to Norwegian. *Studies in Educational Evaluation*, *64*, 100816. doi: 10.1016/j.stueduc.2019.100816

A Appendices

A.1 Age Distributions for Student Sample and Population





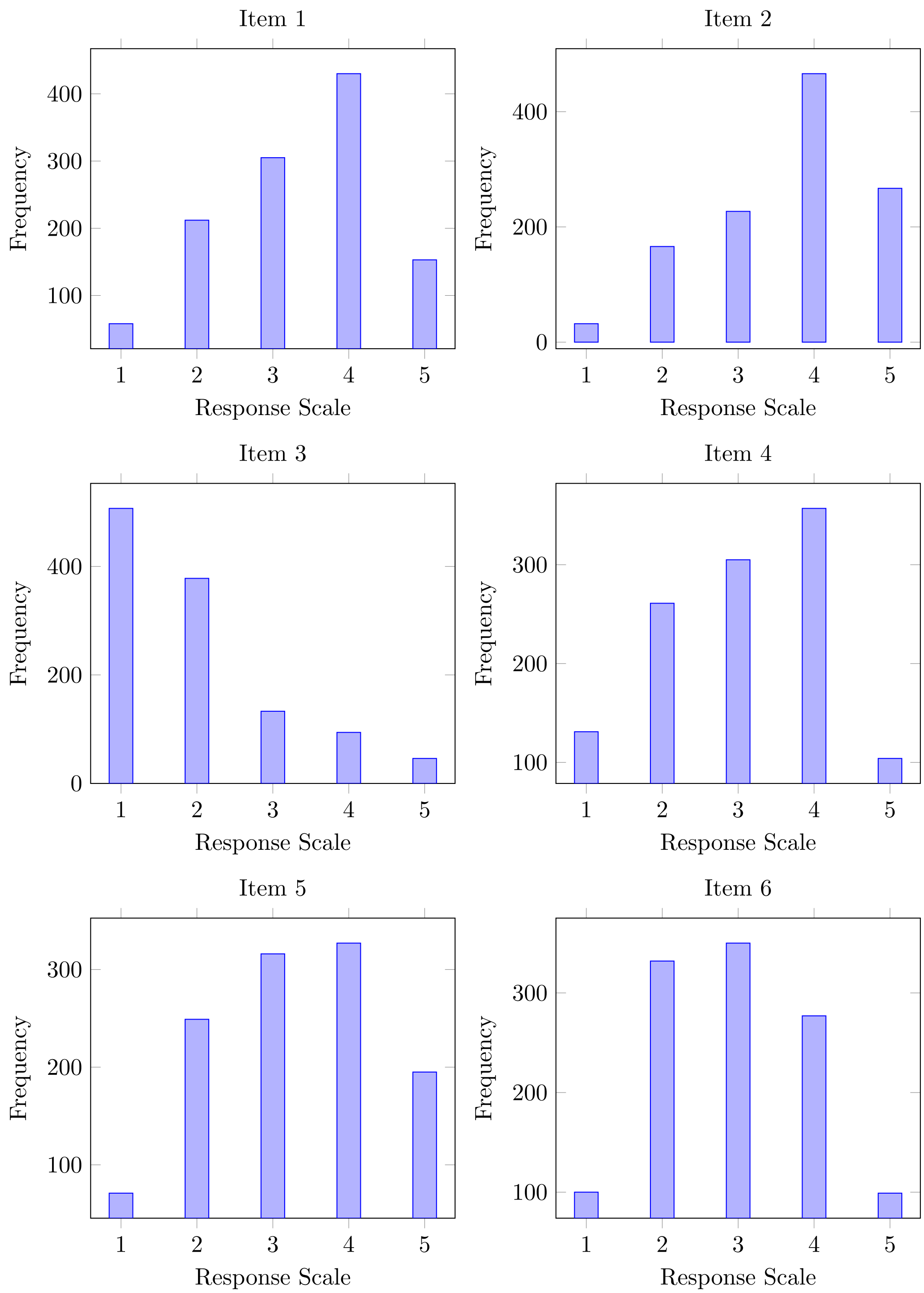
Draft version 1, 20/03/2020. This paper has not been peer reviewed. Please do not copy or cite without author's permission.

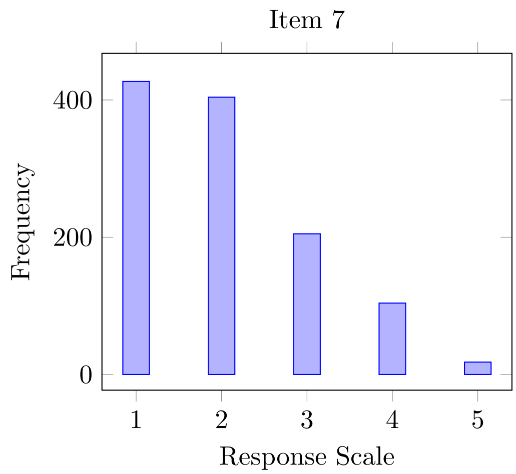
A.2 Biggs et al. (2001) R-SPQ-2F

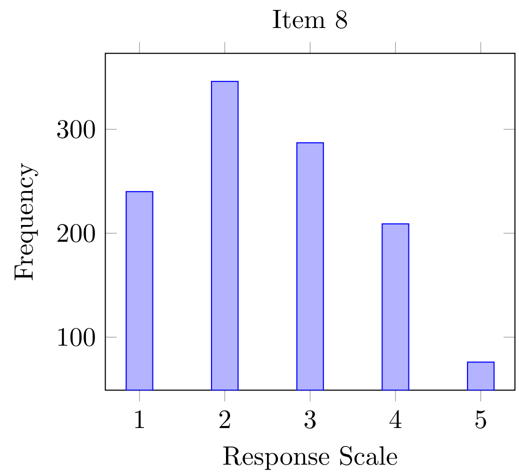
Item	Wording
1	I find that at times studying gives me a feeling of deep personal satisfaction
2	I find that I have to do enough work on a topic so that I can form my own conclusions before I am satisfied
3	My aim is to pass the unit while doing as little work as possible
4	I only study seriously what's given out in class or in the unit guide
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 unit very interesting so I keep my work to the minimum
8	I learn some things by rote, going over and over them until I know them by heart even I do not understand them
9	I find that studying academic topics can at times be as exciting as a good novel or a movie
10	I test myself on important topics until I understand them completely
11	I find I can get by in most assessments by memorising key sections rather than trying to understand them
12	I generally restrict my study to what is specifically set as I think it is unnecessary to do anything extra
13	I work hard at my studies because I find the material interesting
14	I spend a lot of my free time finding out more about interesting topics which have been discussed in different classes
15	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
16	I believe that lecturers shouldn't expect students to spend significant amounts of time studying material everyone knows won't be examined
17	I come to most classes with questions in my mind I want answering
18	I make a point of looking at most of the suggested readings that go with the lectures
19	I see no point in learning material which is not likely to be in the examination
20	I find the best way to pass examinations is to try to remember answers to likely questions

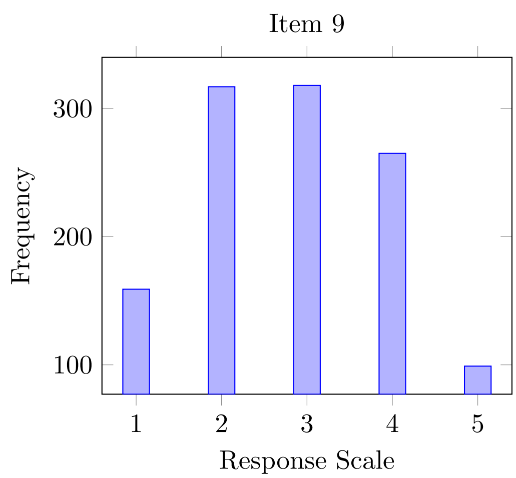
A.3 Response Distributions for the R-SPQ-2F Items

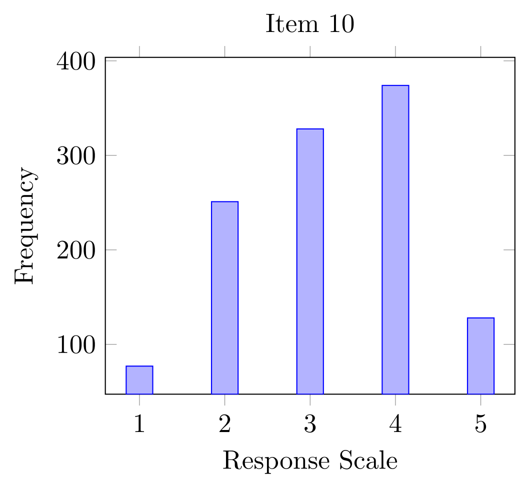
A.3 Response Distributions for the R-SPQ-2F Items

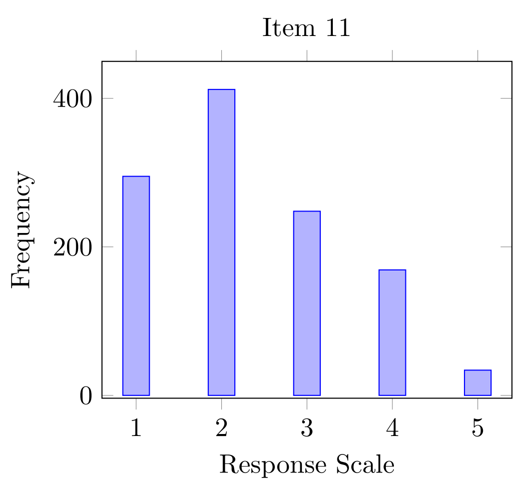

Item 1
Frequency
100
200
300
400
1
2
3
4
5
Response Scale
Item 2
Frequency
0
200
400
1
2
3
4
5
Response Scale
Item 3
Frequency
0
200
400
1
2
3
4
5
Response Scale
Item 4
Frequency
100
200
300
1
2
3
4
5
Response Scale
Item 5
Frequency
100
200
300
1
2
3
4
5
Response Scale
Item 6
Frequency
100
200
300
1
2
3
4
5
Response Scale

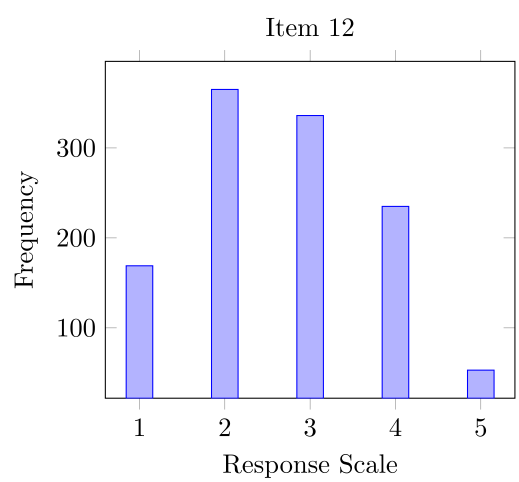

Item 7
400
200
0
Frequency
1 2 3 4 5
Response Scale

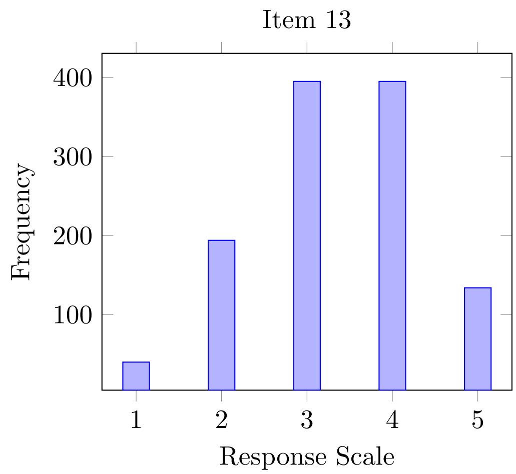

Item 8
300
200
100
Frequency
1 2 3 4 5
Response Scale

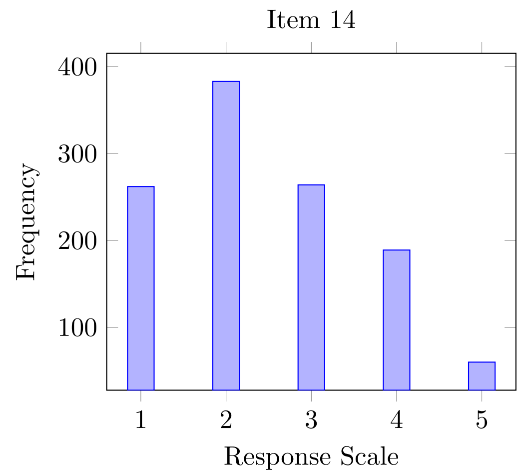

Item 9
300
200
100
Frequency
1 2 3 4 5
Response Scale

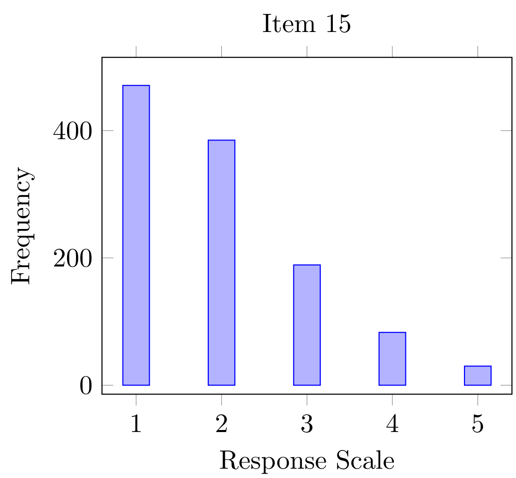

Item 10
400
300
200
100
Frequency
1 2 3 4 5
Response Scale

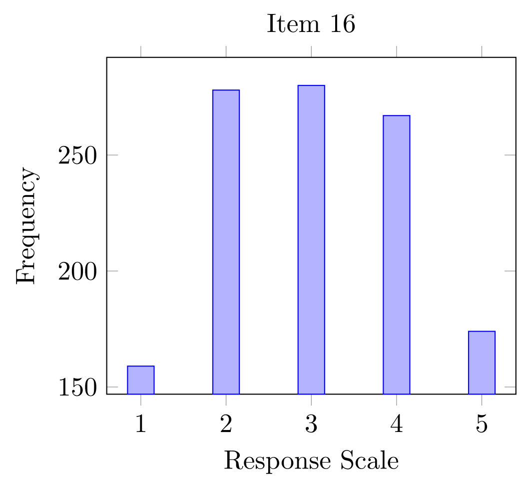

Item 11
400
200
0
Frequency
1 2 3 4 5
Response Scale

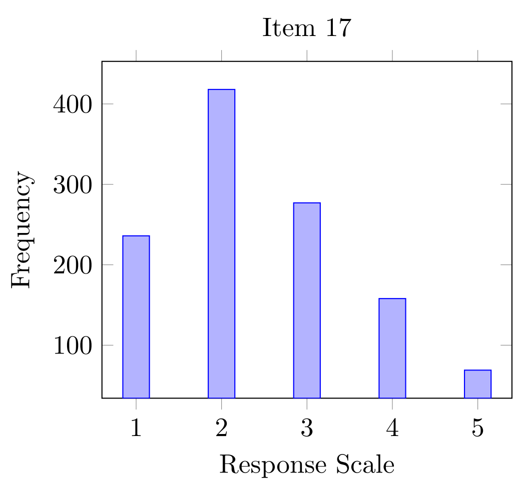

Item 12
300
200
100
Frequency
1 2 3 4 5
Response Scale

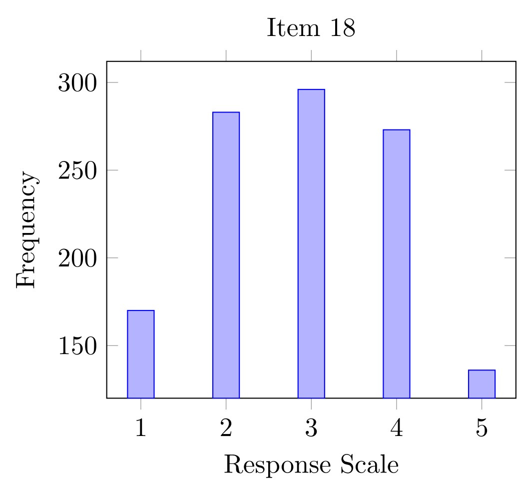

Item 13
Frequency
100
200
300
400
1
2
3
4
5
Response Scale

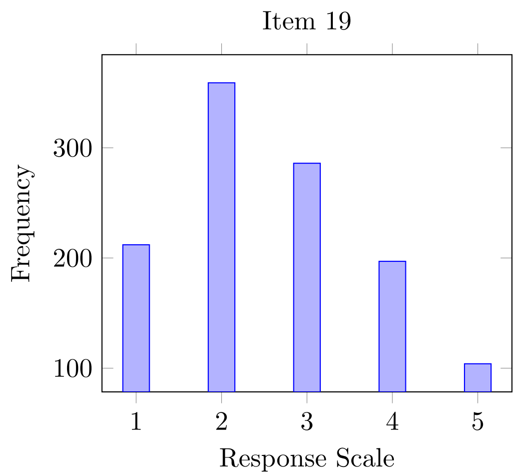

Item 14
Frequency
100
200
300
400
1
2
3
4
5
Response Scale

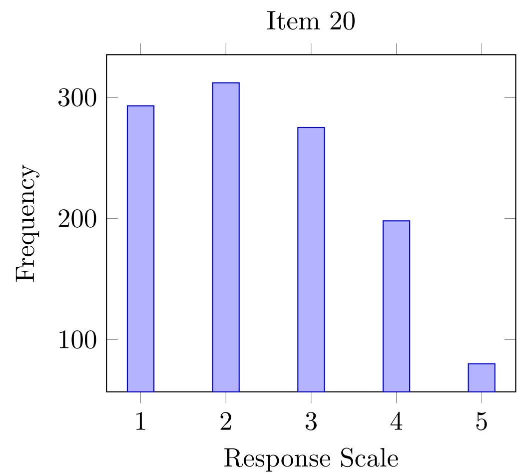

Item 15
Frequency
0
200
400
1
2
3
4
5
Response Scale


Item 16
Frequency
150
200
250
1
2
3
4
5
Response Scale


Item 17
Frequency
100
200
300
400
1
2
3
4
5
Response Scale


Item 18
Frequency
150
200
250
300
1
2
3
4
5
Response Scale


Item 19
Frequency
300
200
100
1
2
3
4
5
Response Scale


Item 20
Frequency
300
200
100
1
2
3
4
5
Response Scale

A.4 Residual Correlations for Model 2 (ESEM)

A.4 Residual Correlations for Model 2 (ESEM)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-																			
2	.066	-																		
3	.036	.053	-																	
4	.019	.031	.111	-																
5	.009	.007	.041	.022	-															
6	-.018	-.004	.005	-.025	.078	-														
7	.041	.030	.040	.023	-.024	-.019	-													
8	.015	.008	.022	-.006	.004	.038	-.010	-												
9	.009	-.033	.015	-.002	-.029	-.003	.031	.002	-											
10	-.010	.008	-.043	.009	-.001	-.004	.062	.026	.007	-										
11	-.004	-.014	.002	-.016	.025	.014	-.017	.003	.005	-.012										
12	-.007	-.021	.002	.100	.019	.012	-.007	-.017	.010	-.039	.036	-								
13	-.012	-.059	-.035	.028	-.029	-.036	-.051	-.051	.042	.001	.007	.057	-							
14	-.011	.005	-.015	-.030	-.047	.015	-.011	-.015	.002	-.001	.022	-.023	.024	-						
15	047	.012	-.009	-.059	-.007	-.015	.025	.024	-.017	-.014	-.014	.006	-.017	.009	-					
16	-.079	-.034	-.062	-.061	-.033	-.024	-.034	.004	.026	.016	-.023	-.030	.058	.014	.013	-				
17	-.005	-.013	-.002	-.007	-.025	-.045	.002	-.023	-.011	.019	-.032	-.007	.016	-.008	.021	.020	-			
18	-.052	.010	-.037	-.029	-.036	-.035	-.031	-.020	-.034	-.001	-.041	-.014	.049	-.020	.005	.036	.103	-		
19	-.028	-.042	-.077	-.058	-.024	-.008	-.028	-.016	-.042	.014	-.007	-.039	-.009	.032	.012	.061	.045	.064	-	
20	-.020	-.006	-.003	-.034	-.011	.021	-.015	-.003	-.003	-.025	.023	-.022	.011	.017	-.037	.012	-.030	.039	.046	-

A.5 Residual Correlations for Model 3 (CFA)

A.5 Residual Correlations for Model 3 (CFA)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-																			
2	.141	-																		
3	-.064	-.050	-																	
4	.004	.019	.079	-																
5	.030	.028	.036	.013	-															
6	-.063	-.052	.054	-.085	.084	-														
7	-.177	-.184	.108	-.015	-.130	-.081	-													
8	.135	.114	-.045	-.059	.079	.160	-.112	-												
9	-.003	-.047	-.009	-.072	-.029	.002	-.105	.115	-											
10	.036	.051	-.092	-.006	-.007	-.039	-.093	.281	-.001	-										
11	.028	.006	-.028	-.089	.056	.118	-.066	.226	.074	.172	-									
12	-.022	-.038	-.020	.107	.025	.011	-.038	-.005	-.022	-.003	.035	-								
13	.010	-.038	-.129	-.057	-.025	-.055	-.270	.051	.036	.013	.036	-.009	-							
14	-.093	-.081	.071	-.109	-.052	.079	-.024	.111	.014	-.051	.153	-.025	-.005	-						
15	-.025	-.066	.046	-.103	.006	.066	.078	.020	-.011	-.007	.013	-.009	-.081	.129	-					
16	-.048	-.002	-.101	.015	.005	-.057	-.092	-.082	-.022	.027	-.141	-.014	-.005	-.053	-.045	-				
17	-.072	-.084	.082	-.056	-.049	-.009	.011	.184	-.009	.015	.177	.034	-.014	.066	.152	-.026	-			
18	-.097	-.038	-.009	-.090	-.063	-.020	-.072	.164	-.036	.010	.129	-.003	.029	.021	.077	-.030	.160	-		
19	-.065	-.076	-.105	-.003	-.037	-.076	-.062	-.098	-.135	-.016	-.108	-.032	-.133	-.054	-.032	.152	-.005	-.010	-	
20	.010	.017	.000	-.064	.046	.124	-.029	.079	.058	.082	.101	-.021	.036	.135	-.010	-.029	.114	.143	.008	-

A.6 Residual Correlations for Model 3 (ESEM)

A.6 Residual Correlations for Model 3 (ESEM)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-																			
2	.147	-																		
3	-.010	.005	-																	
4	.082	.096	.089	-																
5	.038	.036	.053	.060	-															
6	-.057	-.045	.059	-.040	.087	-														
7	-.025	-.037	.112	-.011	-.023	.037	-													
8	.018	.007	-.053	-.045	-.062	-.033	-.081	-												
9	.003	-.040	.033	.001	-.022	.006	.045	-.029	-											
10	.037	.053	-.099	.024	-.010	-.049	-.004	.107	-.004	-										
11	-.044	-.058	-.034	-.075	-.049	-.034	-.041	.156	-.024	.032	-									
12	.023	.008	-.015	.121	.030	.000	-.028	-.017	.009	-.024	.027	-								
13	.014	-.033	-.043	.053	-.015	-.046	-.069	-.064	.042	.015	-.022	.068	-							
14	-.086	-.073	.063	-.076	-.050	.075	.075	-.084	.018	-.062	-.005	-.049	.005	-						
15	-.005	-.043	.044	-.092	-.013	.024	.087	-.010	-.008	-.055	-.012	-.009	-.031	.074	-					
16	.021	.068	-.093	.029	.034	-.037	-.085	-.081	.037	.032	-.137	-.002	.101	-.047	-.040	-				
17	-.069	-.080	.035	-.062	-.053	-.021	.048	-.011	-.010	-.003	.004	-.029	-.006	.051	.063	-.064	-			
18	-.096	-.036	-.023	-.073	-.066	-.029	-.010	.016	-.038	-.003	.006	-.030	.032	.011	.028	-.035	.144	-		
19	.046	.033	-.098	.008	.027	-.015	-.064	-.086	-.033	.023	-.097	-.021	.023	-.011	-.026	.164	-.018	.011	-	
20	-.037	-.025	-.004	-.053	-.029	.015	-.011	.027	-.009	-.020	.058	-.027	.002	.020	-.028	-.026	-.016	.053	.017	-