British Journal of Educational Psychology (2001), **71**, 511–530 *Printed in Great Britain*

The shortened Study Process Questionnaire: An investigation of its structure and longitudinal stability using confirmatory factor analysis

Robin A. Fox*, **I.C. McManus** and **Belinda C. Winder**

Royal Free & University College Medical School, London, UK

Background. The Study Process Questionnaire (SPQ) is a widely used measure of learning approach and was proposed to have three orientations: surface, deep, and achieving, each with an underlying motive and strategy.

Aims. This study aimed to examine the factor structure and longitudinal stability over five to seven years of a modified shortened 18-item version of the SPQ.

Samples. A total of 1349 medical students completed the shortened SPQ at application and in their final year of medical school. Three additional cohorts of students completed the shortened SPQ during their third and fourth year of medical school (sample size: 194, 203, 174).

Method. Confirmatory factor analysis was used to examine the dimensionality and longitudinal stability of the shortened SPQ.

Results. Like the full 42-item version, the shortened SPQ has six subscales and the data are best fit by three second order shared indicator factors (surface, deep and achieving) and a single higher order composite deep-achieving factor. The longitudinal analysis found 26.8%, 26.3%, and 18.7% of the non-attenuated variance of the surface, deep and achieving factor scores in the final year is predicted from the shortened SPQ completed at application to medical school.

Conclusions. The shortened 18-item SPQ has the same six subscales as the full SPQ as well as three second order shared indicator factors (surface, deep, achieving) and one higher order deep-achieving factor similar to that suggested by Biggs (1987). The longitudinal analysis supports this hypothesis and suggests that these learning approaches are partly stable during medical school undergraduate training and partly modifiable under the influence of the educational environment.

*Requests for reprints should be addressed to Dr Robin Fox, Priory Farmhouse, Townsend, Marsh Gibbon, Bicester, Oxon, OX27 0EY.

Students approach their study in different ways. There appear to be two major theoretical standpoints for the source of current learning process questionnaires (Entwistle & Waterston, 1988).

i) The information processing (IP) position originating from cognitive psychology (e.g., Schmeck, Ribich, & Ramanaiah, 1977). The emphasis in the IP models is on a set of theoretical constructs about learning which apply irrespective of the learning environment.

ii) The student approach to learning (SAL) position (e.g., Study Process Questionnaire (SPQ; Biggs, 1987), Approaches to Studying Inventory (ASI; Ramsden & Entwistle, 1981)) arose partly out of dissatisfaction with the IP models, as the learning environment was felt to have profound effects on studying (Entwistle & Waterston, 1988). The SAL emphasises the context within which learning occurs.

The impetus for the SAL position was the phenomenographic approach of Marton and Säljö (1976). They asked students to read academic articles and to describe what they had learnt. Analysis of the interview data identified differences in the students' intentions with which they approached the task and the processes they used as they studied the article. One group had the intention of actively seeking the author's meaning – they appraised the evidence in relation to the conclusions and related the new ideas in the article to their previous knowledge/experience. This pattern of intention and process was called the *deep approach*. The other group's intention was to identify and then memorise the facts and ideas they deemed important in the text. As a result this second group failed to appreciate a lot of the article's structure and principles and they tended to rote learn facts they thought they would be required to produce at the end of the exercise – this pattern of intention and process was called the *surface approach*. This dichotomy of surface and deep approaches has been confirmed by a number of workers in higher education using semi-structured interviews (Hounsell, 1984; Morgan, Taylor, & Gibbs, 1982; Ramsden 1979). As the underlying theory of student approaches to learning has been described in detail elsewhere (e.g., Biggs, 1987, 1993; Kember & Leung, 1998; Watkins, 1996); it has not been extensively repeated in this article which is primarily concerned with the psychometric properties of the shortened SPQ. The interested reader is referred to the above references.

The Study Process Questionnaire (Biggs, 1987) is a widely used measure of students' approaches to learning. It is a 42-item questionnaire which is suggested to have three dimensions (surface (SA), deep (DA) and achieving (AA)) each with two subscales of motive and strategy (Biggs, 1985, 1987), surface motive (SM), surface strategy (SS), deep motive (DM), deep strategy (DS), achieving motive (AM), and achieving strategy (AS). As such there are seven questionnaire items for each of the six subscales. The achieving dimension is also described as 'strategic' in the literature (e.g., Newble & Entwistle, 1986). In this paper, we will use the term 'achieving' for the dimension and the term 'strategy' will only be used in reference to the motive/strategy subscales. A summary of the motives and strategy for each of the three learning approach dimensions is shown in Table 1, and has been reviewed by Newble & Entwistle (1986).

Biggs's original work for this instrument dates back to the 1960s. He hypothesised that 'variations on factors such as cognitive style, personality and values, would generate different emphases on coding and rehearsal strategies . . . . these would manifest in the academic context by different ways of studying' (Biggs, 1994). The 10-

Table 1. Summary of the differences in motivation and study process of surface, deep, and achieving approaches to study

Approach	Motivation	Process (strategy)
Surface	Fear of failure Desire to complete their course of study	Rote learning of facts and ideas Focusing on task components in isolation Little real interest in content
Deep	Interest in the subject Vocational relevance Personal understanding	Relate ideas to evidence Integration of material across courses Identifying general principles
Achieving	Achieving high grades Competing with others To be successful	Use any technique that achieves highest grades Level of understanding patchy and variable

scale Study Behaviour Questionnaire was the eventual result of these thought processes (Biggs, 1976). The large number of scales made the instrument difficult to use and this number was reduced using second order factor analysis, producing three higher order factors (Reproducing (surface), Internalising (deep) and Organising (achieving)). Correlating the items with each of these higher order factor scores revealed that items correlating highly with each factor score fell into two groups, affective and cognitive, which formed congruent motive-strategy combinations – this congruence theory forms the basis of the SPQ (Biggs, 1978).

The presence of an achieving factor is supported by the work of Miller and Parlett (1974). They described the cue-seeking behaviour of certain students who sought out hints for forthcoming examinations and tried to impress their teachers. Additional support for the three dimensions of the SPQ comes from the work of Ramsden and Entwistle (1981) who developed the Approaches to Studying Inventory independently of the work of Biggs. The ASI includes three orientations similar to that of the SPQ. The surface, deep and achieving orientations of these two instruments have been found to correlate when both were administered to the same group of students with attenuated correlation coefficients ($p < .01$) of .44–.61 suggesting both are measuring similar constructs (Wilson, Smart, & Watson, 1996). The presence of the achieving approach was not identified in Marton and Säljö's work (1976). This may be because they were asking students about isolated experimental reading of text whilst Biggs and Entwistle were both asking students to relate their approach to learning in relation to their actual courses (Newble & Entwistle, 1986), which appears more likely to tap into the cue-seeking behaviour suggested by Miller and Parlett (1974). The controversy regarding the existence of this third orientation will be discussed later.

Student approach to learning inventories are associated with quantitative and qualitative educational outcomes. Surface approaches generally correlate with poor (mean −0.11) and deep/strategic approaches (mean 0.20 and 0.19 respectively) with better academic grades (16 studies reviewed by Watkins, 1996). Similarly deep/strategic (mean 0.30 and 0.28 respectively) but not surface approaches correlate with higher student self-esteem measures (nine studies reviewed by Watkins, 1996).

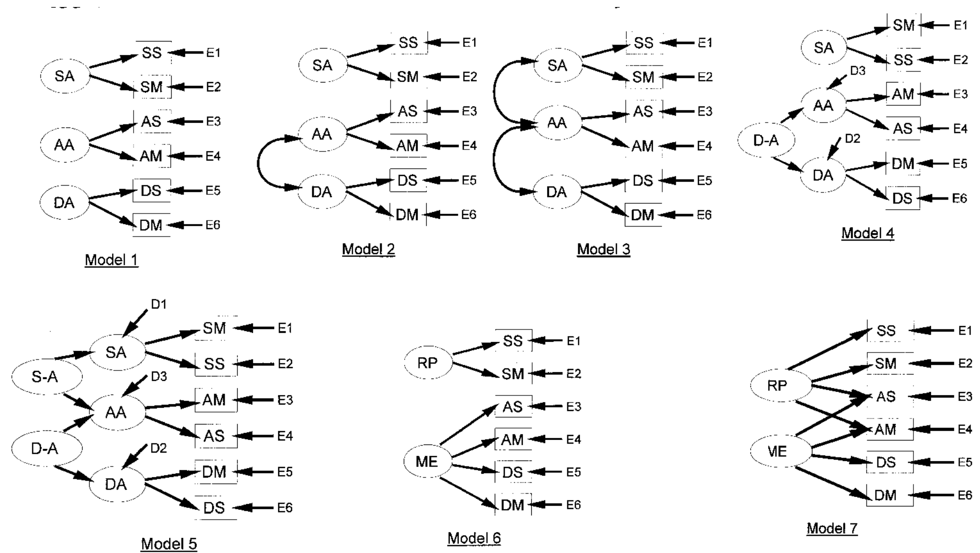
Previous studies have looked at the psychometric properties of the SPQ. Analysis of the validity of the subscales has typically been through reliability assessments or exploratory first order factor analysis. Hattie and Watkins (1981) found the internal consistency of each of the six subscales to be satisfactory and factor analysis revealed a six-factor solution consistent with these six subscales in Australian students. O'Neil and Child (1984) found similar results although four items did not load onto any of the six factors corresponding to the subscales in British students. Kember and Gow (1990) were unable to replicate the six-factor solution in Hong Kong students. Christiansen, Massey, and Isaacs (1991) found 11 of the 14 surface motive and strategy subscales to load significantly onto one or more of the other subscales. Biggs (1993) criticises this type of analysis for inappropriately extracting six *orthogonal* factors, which is not consistent with his motive strategy congruence theory (Biggs, 1978). He believes the subscales to be further validated on the basis of Lisrel goodness of fit indices and scale factor analysis in 4130 Hong Kong university students (Biggs, 1993). A number of researchers have assumed the six SPQ subscales to be useful and have explored the second order factor structure. Earlier studies suggest three second order factors consistent with surface, deep, and achieving (Biggs, 1985, 1987; Hattie & Watkins, 1981). A number of studies using exploratory factor analysis have suggested the presence of only two second order factors, one showing loading on deep and achieving motives and strategies and the other on surface and achieving motives, and surface strategy (Biggs & Rihn, 1984; Watkins & Akande, 1992).

Confirmatory factor analysis (CFA) is a superior methodology to exploratory factor analysis for examining the dimensionality and structure of learning approach questionnaires (Kember & Leung, 1998). With CFA the validity of a priori factor structures can be studied and compared using goodness of fit indices and the factors need not be orthogonal. Nested models can be directly compared using formal tests of statistical significance. Andrews, Violato, Rabb, and Hollingsworth (1994) used CFA on the LPQ (a version of SPQ designed for use with school children (Biggs, 1987)) and were able to identify the expected three second order factors suggested by Biggs, although they did not look at two-factor models. Wong, Lin, & Watkins (1996) tested six CFA models on 10 sets of LPQ data from various countries and found better fits with two second order factors or correlated three second order factors than the original orthogonal three second order factor model. The most comprehensive CFA study on the SPQ looked at the second order structure from data on 4843 Hong Kong university students (Kember & Leung, 1998). They tested seven a priori models based on the literature to date, as shown in Figure 1. They found the best fit to be with two second order factors, one loading onto surface and achieving strategies and motives, and the other onto the deep and achieving strategies and motives (model 7).

This study has examined the first and second order factor structure and the long-term stability of a shortened and slightly revised form of the SPQ using structural equation modelling. The rationale behind the modification of the SPQ was to:

- Shorten it to make it easier to administer to students, as part of a large questionnaire containing multiple other scales, and to make it more suitable for repeated administration (e.g., Tooth, Tonge, & McManus, 1989)
- Anglicise wording to make it appropriate for British students

Figure 1. Structural models of the shortened SPQ (models 1–7 as in Kember & Leung, 1998)



Note: SS = surface strategy, SM = surface motive, AS = achieving strategy, AM = achieving motive, DS = deep strategy, DM = deep motive, SA = surface approach, AA = achieving approach, DA = deep aproach, RP = reproducing, ME = meaning

- Make it suitable to assess students longitudinally so that the item wording is as relevant to an 18-year-old as to a 24-year-old student six years later, which is important for long courses such as medicine
- Choose items which have a purer factor structure.

Previous Varimax analyses of the full 42-item questionnaire have shown a lack of consistency for many of the items. In the 1984 study by O'Neil and Child, which shows the Varimax analysis for two separate data sets of Biggs and the authors' own data, the following problems can be seen.

i) There are items such as numbers 7, 10, 16 and 19 (original numbering for 42-item version) in which for none of the three factor analyses do any of the items show meaningful loadings (defined as > = 0.25).

ii) There are items such as numbers 13, 22, 26, and 30 where the items load on at least two factors.

iii) Of the 42 items, only 22 load above 0.25 on a single factor, in 14 cases for all three analyses, for 6 cases in two analyses and in 2 cases for one analysis (the remaining analyses being below the 0.25 criterion).

Method

Shortened SPQ

The 42-item SPQ was shortened to 18 items by one of the authors (ICM) as shown in column 1, Table 2. The rationale for this has been discussed above. Selection of the items was principally based on questions which loaded only on a single factor in as many of the Varimax analyses (O'Neil & Child, 1984) as possible, but with the constraint that three questions (out of seven) were selected for each of the six subscales. Each item is answered on a 5-point scale: (1) 'rarely true', (2) 'sometimes true', (3) 'true half the time', (4) 'frequently true', (5) 'usually true'.

Samples

Longitudinal sample

The shortened 18-item SPQ was sent to applicants to five British medical schools in 1990 (Time 1) who were resident in the European Community and again to the same students in their final year of medical school, between five and seven years (Time 2) later (McManus, Richards, Winder, Sproston, & Styles, 1995; McManus, Richards, & Winder, 1999). Non-respondents were sent reminders and additional copies of the questionnaire after five weeks and nine weeks. Of 5845 such applicants, 5361 (92%) returned the questionnaire at Time 1. Of these 5845 applicants, 1963 individuals were due to qualify at Time 2. Of these 1963 finalists in the study, 1386 (70.6%) returned the questionnaire at Time 2. Overall therefore, of 1963 students sent the questionnaire on both occasions, 1349 returned it at both Time 1 and Time 2, 539 only at Time 1,37 only at Time 2, and 38 on neither occasion. Only the 1349 students who completed the shortened SPQ on both occasions are included in the main analysis. A small percentage of data points were omitted on one occasion or the other. These missing values were imputed by a method in which a substitute value was obtained at random from another subject with similar scores on the same learning approach dimension. Only 0.6% of the questionnaire items were missing, the majority of which consisted of one missing item per questionnaire.

In order to assess possible response bias, we compared the mean scores on the subscales of the abbreviated Study Process Questionnaire at Time 1, of those replying and those not replying at Time 2. We similarly compared the very small number of people who completed the questionnaire only at Time 2 with those completing it on both occasions.

Additional samples

The shortened 18-item SPQ was given to third year medical students at University College London Medical School in 1998 (UCL98 cohort) after they had completed their end of year clinical examination and similarly to the next third year medical students the following year (UCL99 cohort). It was also given to fourth year medical students at St. Bartholomew's and the Royal London School of Medicine in 1998 at the end of one of their seminars (Barts98 cohort). In total 194/242 (80.2%) of the UCL98, 203/214 (94.8%) of the UCL99 and 174/226 (77.0%) of the Barts98 students completed the questionnaire. Ethics committee approval allowed for only one administration of this questionnaire. The small numbers of missing values were handled as for the longitudinal sample.

Table 2. a) Completely standardised first order factor loadings at Time 1 using Lisrel (Time 2 in parentheses). The shortened SPQ is administered to the students in the order of 1)-18) but items are grouped here by their subscales for clarity. b) Correlation matrix for subscales: at Time 1 below diagonal, Time 2 above diagonal

Item	Completely standardised factor loadings					
	SM	SS	DM	DS	AM	AS
2) I chose my present courses largely with a view to the job situation when I graduate rather than their intrinsic interest to me.	0.289 (0.297)	0[1]	0[1]	0[1]	0[1]	0[1]
13) I almost resent having to do further years studying after leaving school, but feel that the end results make it all worthwhile.	0.325 (0.385)	0[1]	0[1]	0[1]	0[1]	0[1]
16) Whether I like it or not, I can see that further education is for me a good way to get a well-paid or secure job.	0.448 (0.437)	0[1]	0[1]	0[1]	0[1]	0[1]
5) I think browsing around is a waste of time, so I only study seriously what's given out in class or in course outlines.	0[1]	0.630 (0.667)	0[1]	0[1]	0[1]	0[1]
12) I generally restrict my study to what is specifically set as I think it is unnecessary to do anything extra.	0[1]	0.671 (0.660)	0[1]	0[1]	0[1]	0[1]
15) I find it best to accept the statements and ideas of my lecturers and question them only under special circumstances.	0[1]	0.330 (0.486)	0[1]	0[1]	0[1]	0[1]
3) I find that at times studying gives me a feeling of deep personal satisfaction.	0[1]	0[1]	0.639 (0.676)	0[1]	0[1]	0[1]
10) I find that studying academic topics can at times be as exciting as a good novel or film.	0[1]	0[1]	0.751 (0.786)	0[1]	0[1]	0[1]
11) I usually become increasingly absorbed in my work the more I do.	0[1]	0[1]	0.687 (0.661)	0[1]	0[1]	0[1]
1) While I am studying, I often think of real life situations to which the material that I am learning would be useful.	0[1]	0[1]	0[1]	0.485 (0.475)	0[1]	0[1]
8) I find that I have to do enough work on a topic so that I form my own point of view before I am satisfied	0[1]	0[1]	0[1]	0.627 (0.602)	0[1]	0[1]
17) I try to relate new material, as I am reading it, to what I already know on the topic.	0[1]	0[1]	0[1]	0.479 (0.538)	0[1]	0[1]
4) I want top grades in most or all of my courses so that I will be able to select from among the best positions available when I graduate.	0[1]	0[1]	0[1]	0[1]	0.643 (0.769)	0[1]
7) I would see myself basically as an ambitious person and want to get to the top, whatever I do.	0[1]	0[1]	0[1]	0[1]	0.710 (0.704)	0[1]
14) I see getting high marks as a kind of competitive game, and I play it to win.	0[1]	0[1]	0[1]	0[1]	0.547 (0.694)	0[1]
6) I try to work consistently throughout the term and review regularly when the exams are close.	0[1]	0[1]	0[1]	0[1]	0[1]	0.724 (0.757)
9) I try to do all of my assignments as soon as possible after they have been set.	0[1]	0[1]	0[1]	0[1]	0[1]	0.615 (0.712)
18) I keep neat, well organised notes for most subjects.	0[1]	0[1]	0[1]	0[1]	0[1]	0.506 (0.524)

Note: [1]value fixed at 0

(a)

	SM	SS	DM	DS	AM	AS
SM		0.643	−0.149	0.057	0.219	0.057
SS	0.743		−0.234	−0.313	−0.071	−0.065
DM	−0.246	−0.274		0.682	0.523	0.356
DS	0.027	−0.354	0.678		0.502	0.430
AM	0.577	0.137	0.238	0.328		0.409
AS	−0.108	−0.173	0.455	0.402	0.343	

(b)

Confirmatory factor analysis

The LISREL8.30 program (Jöreskög & Sörbom, 1999) was used to obtain parameter estimates and goodness of fit indices of the models tested. Several goodness of fit indices were used as recommended by Bollen and Long (1993 p. 8). They include chi-square value (χ^2), with its degree of freedom (d.f.), Bentler's comparative fit index (CFI), the adjusted goodness of fit index (AGFI) and the standardised root mean square residual (SRMR). Typically as a model accounts for more of the covariation amongst test variables the value of χ^2 decreases, CFI/AGFI increase (to above 0.9 in good fitting models), and the SRMR decreases (to below 0.05 in good fitting models). The *p*-value significance of the χ^2 is not used as a goodness of fit parameter in this study, as in large samples trivial differences between the sample and estimated population covariance matrices are often deemed statistically significant (Ullman, 1996).

Models tested

In analysis one, the first order factor structure of the shortened questionnaire was examined at Time 1 and Time 2 using the longitudinal sample.

Three a priori models were tested: Model A assumes the six first order factors are orthogonal. Model B allows three correlations (between SS & SM, AS & AM and DS & DM), consistent with Biggs' congruent motive-strategy hypothesis. Model C allows all six first order factors to correlate, consistent with the more extensive factor correlations found by O'Neil and Child (1984).

In analysis two, the six first order factors suggested by Biggs are assumed. The second order structure was evaluated with the longitudinal data using the seven a priori models of Kember and Leung (1998) (see Figure 1). Two additional a priori models allowing the higher order factors of models 5 and 7 to correlate were also assessed in view of a negative correlation between surface and deep SPQ scores found in British medical students (McManus, unpublished observation) (Figure 2). Two a posteriori models were analysed to further evaluate this apparent negative correlation between surface and deep SPQ scores (models 8 and 9) at Time 1 and were analysed a priori at Time 2 (see Figure 2). The three questionnaire items for each of the six factors were summed to obtain the six subscale scores.

In analysis three, confirmatory factor analysis was used a priori to evaluate the hypothesised model 8 from the longitudinal data using the three additional cohorts (UCL98, UCL99, and Barts98).

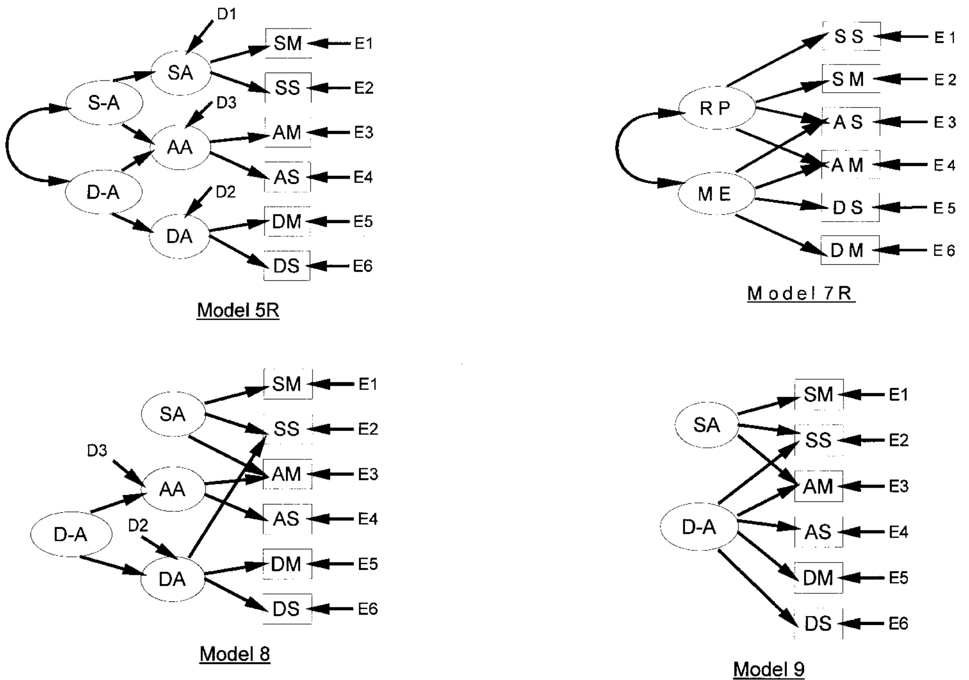
In analysis four, a longitudinal model is used to assess the stability of the shortened SPQ over five to seven years using the best fitting model from analyses two/three.

Results

Longitudinal sample sampling bias analysis

The mean scores for the subscales of the shortened SPQ are shown in Table 3. Overall there was no significant difference in the mean scores on the subscales of the abbreviated SPQ at Time 1 in those replying and those not replying at Time 2 (MANOVA, Wilks' Lambda 0.995, $p = .20$). Also there was no significant difference when comparing those few students who completed the questionnaire only at Time 2, with those completing it on both occasions (MANOVA, Wilks' Lambda 0.995,

Figure 2. New structural models of the shortened SPQ



Note: SS = surface strategy, SM = surface motive, AS = achieving strategy, AM = achieving motive, DS = deep strategy, DM = deep motive, SA = surface approach, AA = achieving approach, DA = deep approach, RP = reproducing, ME = meaning

$p =$ .36). There is therefore no systematic pattern in the process of attrition, which occurred in particular with the smaller sample at Time 2, compared to Time 1, lending confidence to the sampling process in the longitudinal study.

Analysis One

Table 4 shows the goodness of fit indices for the three models examining the first order structure of the shortened SPQ at Time 1 (application to medical school) and Time 2 (final year of medical school). Model C provides the best fit (and a good fit) to the data for both Time 1 and Time 2 and the completely standardised factor loadings are shown in Table 2. These results suggest that the first order factor structure of the shortened SPQ is consistent with the six factors originally suggested by Biggs for the 42-item SPQ, although the majority of these factors appear to be significantly correlated (in part as Biggs expected).

Analysis Two

The Cronbach alpha reliabilities for the six subscales are shown in Table 5. Table 6 shows the results of the fit of the 11 higher order factor models tested using the six subscales. All the paths in all the models were statistically significant.

Table 3. Mean scores (SD) for the subscales of the shortened SPQ according to time of completion (longitudinal sample)

Time of completion of shortened SPQ	Mean score (SD)											
	SM1	SS1	DM1	DS1	AM1	AS1	SM2	SS2	DM2	DS2	AM2	AS2
Time 1 only	7.36	5.93	10.51	10.79	11.39	11.66						
(N = 539)	(2.44)	(2.38)	(2.69)	(2.22)	(2.64)	(2.68)						
Time 2 only							7.14	6.30	8.08	9.65	6.38	7.24
(N = 37)							(1.86)	(2.21)	(2.38)	(1.96)	(2.95)	(3.09)
Time 1 & Time 2	7.22	5.83	10.49	10.58	11.12	11.74	7.45	6.54	8.75	9.75	6.94	8.31
(N = 1349)	(2.31)	(2.12)	(2.68)	(2.29)	(2.81)	(2.58)	(2.24)	(2.41)	(2.75)	(2.44)	(3.04)	(3.17)
t statistic	1.08	0.87	0.16	1.75	1.94	−0.64	−0.84	−0.61	−1.46	−0.25	−1.11	−2.01
(d.f.$_{1886}$ at Time 1) (d.f.$_{1384}$ at Time 2)	(p= .28)	(p= .38)	(p= .87)	(p= .08)	(p= .05)	(p= .52)	(p= .40)	(p= .54)	(p= .15)	(p= .80)	(p= .27)	(p= .04)

Note: SS = surface strategy, SM = surface motive, AS = achieving strategy, AM = achieving motive, DS = deep strategy, DM = deep motive. In row 2, 1 refers to Time 1, 2 refers to Time 2.

Table 4. Goodness of fit indices for the first order structure of the shortened SPQ using the longitudinal sample

	Time 1 (application to medical school)				Time 2 (final year of medical school)			
Analyses	d.f.	χ^2	AGFI	CFI	d.f.	χ^2	AGFI	CFI
Model A[1]	135	1458.2	0.85	0.67	135	1612.7	0.83	0.71
Model B[2]	132	933.3	0.91	0.80	132	1087.6	0.90	0.81
Model C[3]	120	519.5	0.94	0.90	120	644.4	0.93	0.90

Note:
1. Model A assumes all six first order factors are orthogonal
2. Model B allows congruent motive-strategy correlations only (i.e., between SS & SM, AS & AM and DS & DM)
3. Model C allows all six first order factors to correlate (all such correlations were significant ($p < 0.05$) except for SM/DS and SM/AS at Time 1 and SM/DS, SM/AS, SS/AM and SS/AS at Time 2).

At Time 1, as found by Kember and Leung (1998), the worst fit is obtained with the original orthogonal three-factor model and much better fits with the shared indicator two-factor models (models 7 and 7R). Model 7R appears to be the best fitting of the a priori models. Whilst this suggests that there is a negative correlation between the surface and deep latent factors it may have been due to some, rather than all, of the subscales being negatively correlated. A posteriori analysis of the data at Time 1 suggested that the surface strategy subscale is a negative indicator of the meaning factor whilst the achieving motive subscale is a shared indicator of both the meaning and the reproducing factors (see model 9, Figure 2). Model 8 was also evaluated a posteriori as it retains the shared indicator model of model 9 as well as building on the presence of a composite higher order deep-achieving factor suggested by Biggs (1987, p. 20). In model 8 the surface strategy subscale is a negative indicator of the deep factor. Model 9 is

Table 5. Cronbach alpha reliability values for the shortened SPQ subscales at Time 1 and Time 2 using the longitudinal sample

	Time 1	Time 2
Surface motivation	0.288	0.318
Surface strategy	0.506	0.618
Deep motivation	0.733	0.749
Deep strategy	0.533	0.549
Achieving motivation	0.654	0.762
Achieving strategy	0.642	0.699

Table 6. Goodness of fit indices for the second order structure of the shortened SPQ using the longitudinal sample

	Time 1 (application to medical school)					Time 2 (final year of medical school)				
Analyses	d.f.	χ^2	AGFI	CFI	SRMR	d.f.	χ^2	AGFI	CFI	SRMR
Model 1[1]	12	417.3	0.85	0.56	0.14	12	426.9	0.85	0.58	0.15
Model 2[2]	9	235.5	0.88	0.75	0.09	9	107.1	0.94	0.90	0.06
Model 3	7	189.7	0.86	0.80	0.09	7	81.4	0.94	0.93	0.06
Model 4[2,3]	9	235.6	0.88	0.75	0.09	9	107.1	0.94	0.90	0.06
Model 5[4]	7	189.7	0.86	0.80	0.09	7	81.4	0.94	0.93	0.06
Model 6[2]	10	250.8	0.88	0.74	0.09	10	119.8	0.94	0.89	0.06
Model 7	7	107.2	0.92	0.89	0.07	7	89.9	0.94	0.92	0.06
Model 7R	6	72.6	0.94	0.93	0.04	6	59.3	0.95	0.95	0.04
Model 5R[4]	6	149.7	0.88	0.84	0.07	6	50.9	0.96	0.96	0.04
Model 8[5]	6	51.6	0.96	0.95	0.04	6	39.0	0.97	0.97	0.03
Model 9	7	65.7	0.95	0.94	0.04	7	50.9	0.96	0.96	0.03

Note:
1. Equal weights were assigned to the two indicators for each of the 3 latent factors for model identification
2. Equal weights were assigned to SS and SM for model identification
3. The variance of disturbance term D2 was fixed at 0
4. The variance of disturbance terms D1, D2 were fixed at 0
5. The variance of disturbance term D2 was fixed at 0

nested within model 8. Model 8 provides a better fit to the data than model 9 (χ^2_{diff} 14.1, d.f. = 1, $p <$ .001).

At Time 2 the original orthogonal three-factor model is again clearly the worst fitting model. The hypothesised model 8 appears to fit the data best and significantly better than model 9 (χ^2_{diff} 11.9, d.f. = 1, $p <$ .001).

Analysis Three

Model 8 from analysis two appears to be the best fitting model for the shortened SPQ. In order to confirm this, model 8 was evaluated using the UCL98, UCL99 and Barts98 cohorts. As shown in Table 6 model 8 fits the data well for all three cohorts with all paths in all models significant ($p < 0.05$). The surface strategy subscale remains a significant negative indicator of the deep factor throughout.

Analysis Four

Model 8 from analysis two provides the best overall fit to the data at both Time 1 and Time 2. This model was used to assess the long-term stability of the shortened SPQ.

A priori four longitudinal models were examined. Longitudinal model I was tested in which each of the three second order factors (SA, DA, AA) and the third order factor (D-A) were stable across time. Longitudinal model II is as model I, but following Corballis and Traub (1970) each of the six first order measures (SM, SS, DM, DS, AM, AS) were correlated across time. Nested within model II were two other models (all with the six correlated first order measures) in which only the three second order factors (longitudinal model III) or only the third order deep-achieving factor and the second order surface approach factor (SA) (longitudinal model IV) were stable across time.

To summarise:

Longitudinal Model I- The three second order factors (SA, DA, AA) and the third order factor (D-A) of Time 2 are predicted using their corresponding factors at Time 1.

Longitudinal Model II- The three second order factors (SA, DA, AA) and the third order factor (D-A) of Time 2 are predicted using their corresponding factors at Time 1 with the six subscales correlated across time.

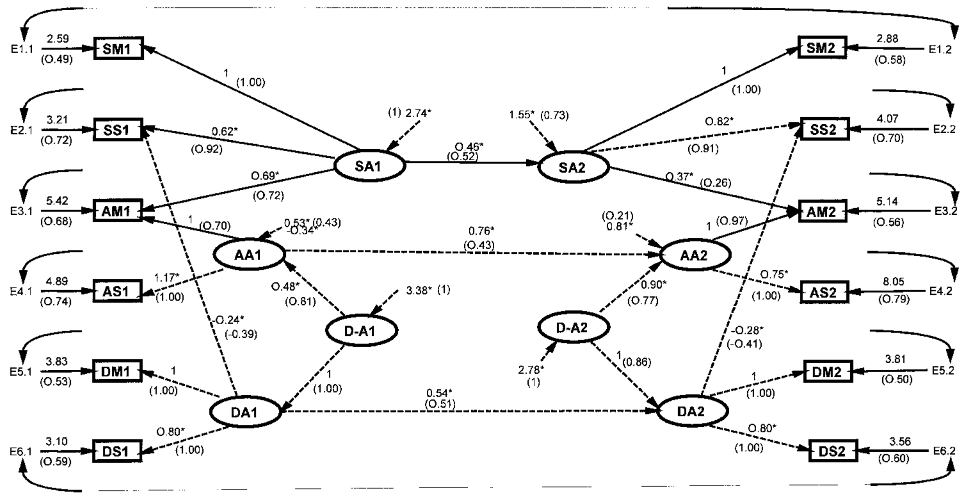
Longitudinal Model III- The three second order factors (SA, DA, AA) Time 2 are predicted using their corresponding factors at Time 1 with the six subscales correlated across time (nested within Longitudinal Model II).

Longitudinal Model IV- The third order factor (D-A) and the second order factor SA of Time 2 are predicted using their corresponding factors at Time 1 with the six subscales correlated across time (nested within Longitudinal Model II).

The fit of longitudinal model I is not good (χ^2 499.5, d.f. = 45, AGFI = 0.89, CFI = 0.84, SRMR = 0.05, variance of AA at Time 2 fixed at 0 for identification). Longitudinal model II significantly improves the fit of the model (χ^2 124.1, d.f. = 38, AGFI = 0.97, CFI = 0.97, SRMR = 0.03)—five of the six subscale correlations were highly significant ($p < .01$) with the surface motive subscale autocorrelation just not significant ($.1 > p > .05$). The path coefficients for the three second order factors at Time 1 predicting those at Time 2 were significant ($p < .01$) whilst that of the third order factor (D-A) was not ($p > .05$). The fit of the two nested models of longitudinal model II was then examined as shown in Table 8. Longitudinal model III represents no worse a fit than longitudinal model II, whilst longitudinal model IV fits the data significantly worse than longitudinal model II ((χ^2_{diff} 8.0, d.f. = 2, $p < .05$).

The path model representing the longitudinal stability of the shortened SPQ using longitudinal model III is shown in Figure 3. This suggests that the proposed factor structure of the SPQ is stable over the five to seven years of medical school and that the surface, deep and achieving latent variables at Time 2 are predicted from those at Time 1 (26.8%, 26.3% and 18.7% of the disattenuated variance respectively).

Figure 3. Unstandardised path model (completely standardised solution in parentheses) of the longitudinal second order factor structure of shortened SPQ



Note: SS = surface strategy, SM = surface motive, AS = achieving strategy, AM = achieving motive, DS = deep strategy, DM = deep motive. In the boxes/ellipses 1 refers to Time 1, 2 refers to Time 2. The six pairs of E (error) terms are all correlated across time but only E1.1/E1.2 and E6.1/E6.2 are shown in the diagram, for clarity. All path coefficients are significant ($p < .01$) except E1.1/E1.2 which is just not significant ($.1 > p > .05$).

Table 7. Goodness of fit indices for the higher order factor structure of the shortened SPQ

	Model 8					Model 9				
Analyses	d.f.	χ^2	AGFI	CFI	SRMR	d.f.	χ^2	AGFI	CFI	SRMR
UCL 98	6	5.4	0.97	1.00	0.04	7	6.7	0.97	1.00	0.04
UCL 99	6	9.3	0.95	0.96	0.04	7	9.3	0.95	0.97	0.04
Barts 98	6	6.4	0.96	1.00	0.04	7	7.1	0.96	1.00	0.04

Table 8. Goodness of fit indices for the longitudinal structure of the shortened SPQ

Analyses	d.f.	χ^2	AGFI	CFI	SRMR
Longitudinal Model I[1]	45	499.5	0.89	0.84	0.05
Longitudinal Model II	38	124.1	0.97	0.97	0.03
Longitudinal Model III	39	124.4	0.97	0.97	0.03
Longitudinal Model IV	40	132.1	0.97	0.97	0.03

Note: Variance of AA at Time 2 fixed at 0 for identification

Discussion

The shortened 18-item SPQ has the same underlying first order factor structure as the full 42-item SPQ suggesting good within-construct validity at this level (Watkins & Akande, 1992). The second order factor structure appears to preserve the presence of the three learning approaches (surface, deep, achieving) suggested by Biggs (1987) with the presence of a composite higher order deep-achieving factor as has previously been suggested by Biggs himself (1987, p. 20). The longitudinal stability of this structure is confirmed in the structural equation model of Figure 3 over five to seven years. In a prospective study using the shortened SPQ and two other cohorts of medical students, high deep and achieving, and low surface scores were predictive of passing medical finals (Time 2) and attaining more clinical experience at medical school (Time 1 and Time 2), suggesting good between-construct predictive validity (McManus, Richards, Winder, & Sproston, 1998). The shortened 18-item questionnaire appears useful for large scale, particularly longitudinal, questionnaire research. It is easy to administer and is easily and fully completed by almost all students. For individual student assessment the full 42-item questionnaire is to be preferred in view of its higher reliability and hence lower error variance for individual diagnosis.

The six first order factors suggested by Biggs appear present at both Time 1 and Time 2 in the shortened SPQ. We have found a second order factor structure consisting of the three learning approaches (surface, deep, achieving) with the presence of a composite higher order deep- achieving factor to fit the data well, for both the longitudinal study of over five years duration as well as with the three other cohorts of British medical students. This study, like those of O'Neil and Child (1984), Wong *et al.* (1996) and Kember and Leung (1998) using the full SPQ, found significant correlations between the subscales, which are more extensive than the congruent motive-strategy model suggested by Biggs (1978). Like Kember and Leung's (1998) study with the full SPQ we also found the achieving motive subscale as an indicator of the surface factor but we also found the surface strategy subscale to be a negative indicator of the deep factor. The tendency for students using surface-like methods of study to be inversely associated with deep learning approaches does not appear surprising in Western cultures although it was not found by Kember and Leung (1998) in Hong Kong students. This may reflect cross-cultural differences as it has been suggested that in Asian cultures understanding may come through memorisation (Biggs, 1993; Marton, Watkins, & Tang 1997) and hence not expected to be negatively associated. Alternatively this may reflect the selection of items based on a purer factor structure used in constructing the shortened SPQ or a peculiarity to medical students. Further studies using confirmatory factor analysis of SPQ should clarify this issue.

Kember and Leung (1998) using the full SPQ suggested that its factor structure was better explained with two second order factors of *meaning* (deep-achieving) and *reproducing* (surface-achieving). This view was shared by Richardson (1994, p. 463) in his review of the learning approach literature, who felt that there was 'little unambiguous supportive evidence for any separate 'strategic' approach to academic assessment of the sort that was originally proposed by Ramsden (1979), based upon an 'achieving' orientation towards studying', a view more recently supported by Kember, Wong, and Leung (1999). After selecting items based upon a purer factor structure from the study of O'Neil and Child (1984) we are able to reproduce three second order factors along the lines suggested by Biggs. The achieving orientation rather than the deep one has also been found to be predictive of academic performance in both British (Tooth *et al.*, 1989 using the shortened SPQ) and Canadian medical students (Arnold & Feighny, 1995) as well as in American psychology students (Hall, Bolen, & Gupton, 1995).

The longitudinal study is novel, in that we have assessed the long-term stability of the shortened SPQ over between five and seven years. This study suggests that 26.8%, 26.3%, and 18.7% of the non-attenuated variance of the surface, deep, and achieving factor scores at Time 2 is predicted from the same scores at Time 1. This suggests that while a student's approach to learning is partly fixed over a medical school undergraduate training it is also modified according to the influence of the learning environment (Biggs, 1993). As put by Ramsden (1988), 'Approaches to learning have to be understood as being both variable and consistent … as long as experiences and contexts differ as in 'natural' learning settings and as long as students strive to adapt to the learning environment, an observer will note elements of both consistency and of variability'.

We did not find that the composite higher order deep-achieving factor at Time 1 predicted that at Time 2 although this may represent a lack of power to determine this effect, particularly because strong longitudinal effects for the first and second order factors had already been taken into account. The first order measures (subscales) appear to be correlated over time in the longitudinal analysis (with the possible exception of the surface motive subscale). This supports the existence of the subscales which appear to be measuring important stable constructs in addition to the higher order factors identified, although it is possible, though unlikely, that these correlations are due to the students becoming familiar with the questionnaire items over time.

The alpha reliability of the present subscales, which have only three items, is somewhat less than that of the original subscales that have seven items. Thus for the deep strategy subscales, the two alpha coefficients are .53 and .54 in the present study. However, it is not surprising that the reliabilities are less since the subscales have fewer items. The Spearman-Brown formula allows one to calculate precisely by how much the reliability will decrease in a scale of reduced number of items. In published studies (Balla, Stokes, & Stafford, 1991; Biggs, 1987; Hattie & Watkins, 1981; Kember & Gow, 1990; Kember & Leung, 1998; O'Neil & Child, 1984) the deep strategy subscale with seven items has alpha coefficients in the range .60 to .75. Taking a middle value, of .67, then the Spearman-Brown formula predicts that the reliability with three items should be 0.47. That our alpha reliabilities are actually .53 and .54 suggests that we have indeed chosen our items well, and that they are more reliable individually than those in the original scale. For the surface motive subscale, the range of reliabilities in the published literature is from .51 to .61, with a middle value of .56. Biggs (1993) explained the low alpha value for the surface motive subscale as being due to it containing two components (fear of failure and the extrinsic motivation of obtaining a qualification). It can be predicted that reduction from seven to three items will reduce the reliability to .34, a value broadly compatible with our values of .29 and .32.

It might be argued reducing the reliability of a scale could only reduce its utility. This is not necessarily the case. It is certainly true for diagnostic studies involving a single individual, where reliability should be as high as possible. However for large-scale research purposes, the important measure is the power of the study, and if reducing a questionnaire in length, and hence reliability, also allows very much larger sample sizes to be attained, then the power of the study will in fact increase. In the present case, in our cross-sectional survey of medical students at Time 1 we had a sample size approaching 6000, and in our longitudinal study we had one of 1349. Neither value would have been feasible if we had used the longer, full questionnaire as part of much larger, more broadly focused questionnaires covering a wide range of issues, of which the Study Process Questionnaire was only a small part. As an example of the effects upon power, detecting a significant correlation of .30 with some other measure requires four times the sample size for detecting a correlation of .55 (Faul & Erdfelder, 1992). A similar proportional effect applies if the other measure is itself measured unreliably. As long therefore as a shortened test allows a sample size of at least four times that which could be obtained with the full questionnaire, then the shortened questionnaire is more powerful and hence more efficient.

In summary we believe the shortened SPQ to have a factor structure broadly similar to that proposed by Biggs for the full SPQ and we have found in previous studies that it is predictive of medical student performance. It is ideally suited to large questionnaires where the SPQ is but one of many scales and hence brevity is of crucial importance in order to achieve a high response rate.

References

Arnold, L., & Feighny, K. M. (1995). Students' general learning approaches and performances in medical school: A longitudinal study. *Academic Medicine, 70*, 715–722.

Andrews, J., Violato, C., Rabb, K., & Hollingsworth, M. (1994). A validity study of Biggs' three-factor model of learning approaches: A confirmatory factor analysis employing a Canadian sample. *British Journal of Educational Psychology, 64*, 179–185.

Balla, J., Stokes, M., & Stafford, K (1991). *Using the Study Process Questionnaire to its full potential: A Hong Kong view*. Technical Report No. 1. Educational Technology Centre, City Polytechnic of Hong Kong.

Biggs, J. B. (1976). Dimensions of study behaviour: Another look at ATI. *British Journal of Educational Psychology, 46*, 68–80.

Biggs, J. B. (1978). Individual and group differences in study process. *British Journal of Educational Psychology, 48*, 266–279.

Biggs, J. B. (1985). The role of metalearning in study processes. *British Journal of Educational Psychology, 55*, 185–212.

Biggs, J. B. (1987). *Student approaches to learning*. Hawthorn, Vic.:, Australian Council for Educational Research.

Biggs, J. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology, 63*, 3–19.

Biggs, J. B. (1994). Approaches to learning: Nature and measurement of. *Encyclopedia of Education* (pp. 318–322). Oxford: Pergamon Press.

Biggs, J. B., & Rihn, B. A. (1984). The effects of intervention on deep and surface approaches to learning. In J. R. Kirby (Ed.), *Cognitive strategies and educational performance* (pp. 279– 293). Orlando, FL: Academic Press.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA:, Sage.

Christensen, C. A., Massey, D. R., & Isaacs, P. J. (1991). Cognitive strategies and study habits: An analysis of the measurement of tertiary students' learning. *British Journal of Educational Psychology, 61*, 290–299.

Corballis, M. C., & Traub, R. E. (1970). Longitudinal factor analysis. *Psychometrika, 35*, 79–98.
Entwistle, N., & Waterston, S. (1988). Approaches to studying and levels of processing in university students. *British Journal of Educational Psychology, 58*, 258–265.
Faul, F., & Erdfelder, E. (1992). *Gpower: A priori, post-hoc, and compromise power analyses for MS-DOS* {computer program}. Bonn, FRG, Bonn University, Department of Psychology.
Hall, C. W., Bolen, L. M., & Gupton, R. H. (1995). Predictive validity of the Study Process Questionnaire for undergraduate students. *College Student Journal, 29*, 234–239.
Hattie, J., & Watkins, D. (1981). Australian and Filipino investigations of the internal structure of Biggs' new Study Process Questionnaire. *British Journal of Educational Psychology, 51*, 241–244.
Hounsell, D. (1984). Learning and essay writing. In F. Marton, D. Hounsell, & N. Entwistle (Eds), *The experience of learning* (pp. 103–123)., Edinburgh: Scottish Academic Press.
Jöreskog, K., & Sörbom, D. (1999). *Lisrel 8.30*. Mooresville, IN: Scientific Software International.
Kember, D., & Gow, L. (1990). Cultural specificity of approaches to study. *British Journal of Educational Psychology, 60*, 356–363.
Kember, D., & Leung, D. Y. P. (1998). The dimensionality of approaches to learning: An investigation with confirmatory factor analysis on the structure of the SPQ and LPQ. *British Journal of Educational Psychology, 68*, 395–407.
Kember, D., Wong, A., & Leung, D. Y. P. (1999). Reconsidering the dimensions of approaches to learning. *British Journal of Educational Psychology, 69*, 323–343.
Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology, 46*, 4–11.
Marton, F., Watkins, D., & Tang, C. (1997). Discontinuities and continuities in the experience of learning: An interview study of high school students in Hong Kong. *Learning and Instruction, 7*, 21–48.
McManus, I. C., Richards, P., & Winder, B. C. (1999). Intercalated degrees, learning styles, and career preferences: Prospective longitudinal study of UK medical students. *British Medical Journal, 319*, 542–546.
McManus, I. C., Richards, P, Winder, B. C., & Sproston, K. A. (1998). Clinical experience, performance in final examinations, and learning style in medical students: Prospective study. *British Medical Journal, 316*, 345–350.
McManus, I. C., Richards, P., Winder, B. C., Sproston, K. A., & Styles, V. (1995). Medical school applicants from ethnic minorities: Identifying if and when they are disadvantaged. *British Medical Journal, 310*, 496–500.
Miller, C. M. L., & Parlett, M. (1974). *Up to the mark: A study of the examination game*. London: Society for Research in Higher Education.
Morgan, A., Taylor, E., & Gibbs, G. (1982). Variations in students' approaches to studying. *British Journal of Educational Technology, 13*, 107–113.
Newble, D. I., & Entwistle, N. J. (1986). Learning styles and approaches: Implications for medical education. *Medical Education, 20*, 162–178.
O'Neil, M. J., & Child, D. (1984). Biggs' SPQ: A British study of its internal structure. *British Journal of Educational Psychology, 54*, 228–234.
Ramsden, P. (1979). Student learning and perceptions of the academic environment. *Higher Education, 8*, 411–427.
Ramsden, P. (1988). Context and strategy-situational influences on learning. In R. R. Schmeck (Ed.), *Learning strategies and learning styles*. New York: Plenum Press.
Ramsden, P., & Entwistle, N. J. (1981). Effects of academic departments on students' approaches to studying. *British Journal of Educational Psychology, 51*, 368–383.
Richardson, J. T. E. (1994). Cultural specificity of approaches to studying in higher education: A literature survey. *Higher Education, 27*, 449–468.
Schmeck, R. R., Ribich, F., & Ramanaiah, N. (1977). Development of a self report inventory for using individual differences in learning processes. *Applied Psychological Measurement, 1*, 413–431.
Tooth, D., Tonge, K., & McManus, I. C. (1989). Anxiety and study methods in preclinical students: Causal relation to examination performance. *Medical Education, 23*, 416–421.

Ullman, J. B. (1996). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (3rd ed., pp. 709–812). New York: Harper Collins.
Watkins, D. A. (1996). Learning theories and approaches to research: A cross-cultural perspective. In D. A. Watkins & J. B. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 3–24). Hong Kong: CERC and ACER.
Watkins, D., & Akande, A. (1992). Assessing the approaches to learning of Nigerian students. *Assessment and Evaluation in Higher Education, 17*, 11–20.
Wilson, K. L., Smart, R. M., & Watson, R. J. (1996). Gender differences in approaches to learning in first year psychology students. *British Journal of Educational Psychology, 66*, 59–71.
Wong, N. Y., Lin, W. Y., & Watkins, D. (1996). Cross-cultural validation of models of approaches to learning: An application of confirmatory factor analysis. *Educational Psychology, 16*, 317–327.

Received 28 February 2000; revised version received 25 April 2001

Appendix

1) Covariance matrix for the 18 items of the shortened SPQ at Time 1 below the diagonal and Time 2 above diagonal (longitudinal sample $N = 1349$). Variances given at bottom of table.

Item No.	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
01		0.143	0.385	0.199	−0.083	0.253	0.179	0.336	0.080	0.288	0.325	−0.116	−0 .080	0.107	−0.113	−0.064	0.414	0.138
02	0.004		0.017	0.045	0.178	0.007	0.010	−0.048	0.020	0.003	−0.045	0.159	0.111	0.071	0.166	0.174	0.023	0.024
03	0.316	−0.025		0.439	−0.053	0.328	0.370	0.301	0.176	0.665	0.55	−0.147	− 0.139	0.265	0.001	−0.007	0.218	0.238
04	0.089	0.172	0.194		0.026	0.490	0.881	0.411	0.254	0.472	0.356	−0.179	−0.008	0.738	0.015	0.128	0.197	0.279
05	−0.139	0.190	−0.065	0.112		−0.016	−0.022	−0.183	0.033	−0.148	−0.127	0.497	0.184	0.023	0.360	0.169	−0.065	0.047
06	0.119	−0.057	0.285	0.146	−0.071		0.461	0.367	0.947	0.371	0.214	−0.236	−0.020	0.329	0.099	−0.035	0.221	0.690
07	0.091	0.090	0.232	0.578	0.078	0.316		0.562	0.351	0.480	0.299	−0.168	−0.032	0.731	0.000	0.246	0.221	0.201
08	0.385	−0.091	0.292	0.152	0.091	0.263	0.305		0.377	0.500	0.426	−0.213	−0. 029	0.343	−0.147	0.003	0.343	0.224
09	0.116	−0.023	0.304	0.130	−0.037	0.596	0.242	0.256		0.355	0.141	−0.106	0. 061	0.254	0.106	0.084	0.144	0.708
10	0.363	−0.185	0.576	0.096	−0.143	0.305	0.160	0.470	0.392		0.676	−0.205	−0.116	0.320	−0.051	−0.036	0.257	0.192
11	0.286	−0.073	0.501	0.113	−0.116	0.226	0.117	0.371	0.264	0.680		−0.160	− 0.119	0.268	−0.080	0.000	0.317	0.167
12	−0.186	0.170	−0.122	0.018	0.377	−0.149	−0.021	−0.130	−0.110	−0.181	− 0.130		0.179	−0.061	0.336	0.142	−0.094	−0.103
13	−0.044	0.095	−0.099	0.054	0.184	−0.084	0.001	0.000	−0.072	−0.105	−0.061	0.146		0.061	0.228	0.245	−0.033	−0.012
14	0.027	0.163	0.186	0.505	0.148	0.137	0.639	0.181	0.165	0.071	0.095	0.066	0.122		0.117	0.224	0.146	0.202
15	−0.053	0.167	0.003	0.102	0.195	0.105	0.068	−0.188	0.078	−0.096	−0.056	0.186	0.146	0.260		0.357	−0.083	0.081
16	−0.039	0.245	−0.023	0.416	0.167	0.000	0.301	0.012	0.033	−0.096	−0.058	0.200	0.171	0.470	0.384		0.211	0.131
17	0.253	−0.024	0.150	0.111	−0.090	0.164	0.098	0.292	0.103	0.199	0.200	−0.099	−0.031	0.048	0.000	0.218		0.238
18	0.065	0.004	0.177	0.077	−0.021	0.402	0.184	0.102	0.425	0.117	0.122	−0.105	−0.085	0.086	0.104	0.034	0.128	
Variance																		
Time 1	1.363	1.625	1.079	1.171	0.903	1.155	1.337	1.285	1.631	1.413	1.187	0.795	0.743	1.950	1.274	1.942	0.732	1.017
Time 2	1.367	1.236	1.185	1.519	1.079	1.844	1.798	1.503	1.680	1.308	1.299	1.056	1.143	1.230	1.270	1.563	0.920	1.849

2) Covariance matrix for the longitudinal study of the shortened SPQ ($N = 1349$)

	SM1	SS1	DM1	DS1	AM1	AS1	SM2	SS2	DM2	DS2	AM2	AS2
SM1	5.332											
SS1	1.754	4.488										
DM1	−0.725	−0.903	7.192									
DS1	0.005	−0.976	2.645	5.241								
AM1	1.790	0.830	1.264	1.102	7.901							
AS1	−0.251	−0.205	2.191	1.316	1.481	6.648						
SM2	1.568	0.695	−0.268	−0.093	0.921	0.046	5.003					
SS2	1.034	1.574	−0.793	−0.927	0.442	−0.075	1.762	5.791				
DM2	−0.428	−0.527	2.745	1.490	0.631	0.855	−0.441	−0.970	7.575			
DS2	−0.014	−0.483	1.360	1.827	0.883	0.690	0.125	−1.098	3.017	5.975		
AM2	0.565	0.007	0.974	0.949	2.783	0.921	0.744	−0.248	3.269	2.365	9.245	
AS2	−0.237	0.092	1.408	0.619	0.866	3.653	0.261	−0.096	2.182	2.043	2.821	10.064

3) The covariance matrices for the UCL98, UCL99 and Barts98 cohorts are available from the authors