



Is PHEEM a multi-dimensional instrument? An international perspective

David Wall, Mike Clapham, Arnoldo Riquelme, Joaquim Vieira, Richard Cartmill, Knut Aspegren & Sue Roff

To cite this article: David Wall, Mike Clapham, Arnoldo Riquelme, Joaquim Vieira, Richard Cartmill, Knut Aspegren & Sue Roff (2009) Is PHEEM a multi-dimensional instrument? An international perspective, Medical Teacher, 31:11, e521-e527, DOI: [10.3109/01421590903095528](https://doi.org/10.3109/01421590903095528)

To link to this article: <https://doi.org/10.3109/01421590903095528>



Published online: 12 Nov 2009.



Submit your article to this journal [↗](#)



Article views: 2338



View related articles [↗](#)



Citing articles: 2 View citing articles [↗](#)

WEB PAPER

Is PHEEM a multi-dimensional instrument? An international perspective

DAVID WALL¹, MIKE CLAPHAM², ARNOLDO RIQUELME³, JOAQUIM VIEIRA⁴, RICHARD CARTMILL⁵,
KNUT ASPEGREN⁶ & SUE ROFF⁷

¹West Midlands Deanery, ²West Midlands Deanery, Education, ³Pontificia Universidad Catolica, Internal Medicine, ⁴University of Sao Paulo, ⁵Good Hope Hospital, ⁶University of Southern Denmark, Medical Education, ⁷University of Dundee, Centre for Medical Education

Abstract

Aim: To look at the characteristics of Postgraduate Hospital Educational Environment Measure (PHEEM) using data from the UK, Brazil, Chile and the Netherlands, and to examine the reliability and characteristics of PHEEM, especially how the three PHEEM subscales fitted with factors derived statistically from the data sets.

Methods: Statistical analysis of PHEEM scores from 1563 sets of data, using reliability analysis, exploratory factor analysis and correlations of factors derived with the three defined PHEEM subscales.

Results: PHEEM was very reliable with an overall Cronbach's alpha of 0.928. Three factors were derived by exploratory factor analysis. Factor One correlated most strongly with the teaching subscale ($R=0.802$), Factor Two correlated most strongly with the role autonomy subscale ($R=0.623$) and Factor Three correlated most strongly with the social support subscale ($R=0.538$).

Conclusions: PHEEM is a multi-dimensional instrument. Overall, it is very reliable. There is a good fit of the three defined subscales, derived by qualitative methods, with the three principal factors derived from the data by exploratory factor analysis.

Introduction

The Postgraduate Hospital Educational Environment Measure (PHEEM) is now an internationally used instrument for measuring the educational climate for doctors in training. This study, using an international collaboration, has brought together data from the use of PHEEM in the UK (in intensive care medicine and the Foundation Programme), from Brazil, Chile and The Netherlands. In this way we have been able to look at the characteristics of the instrument with large number of data sets, and examine the characteristics of PHEEM with an international perspective.

What is the educational environment? It is a set of factors that describe a learner's experiences within that organisation. Chambers & Wall (2000) considered the educational climate in three parts. These were 'the physical environment' (safety, food, shelter, comfort and other facilities), 'the emotional climate' (security, constructive feedback, being supported and the absence of bullying and harassment) and 'the intellectual climate' (learning with patients, relevance to practice, evidence-based, active participation by learners, motivating and planned education).

The good clinical teaching environment (Spencer 2003) should ensure that the teaching and learning is relevant to patients, has active participation by learners and shows professional thinking and behaviours. There needs to be good preparation and planning, both of the structure and content, reflection on learning and evaluation of what has happened in the teaching and learning. Problems include a

Practice points

- PHEEM is now being used internationally to measure the educational climate for doctors in training.
- PHEEM is a multi-dimensional instrument both in terms of the defined subscales and exploratory factor analysis.
- PHEEM is a very reliable instrument.

lack of clear objectives, a focus on knowledge rather than problem-solving skills, teaching at the wrong level, passive observation, little time for reflection and discussion and teaching by humiliation. Behaviours including teaching by humiliation, bullying and harassment are still a big problem. An editorial in *Medical Education* (Spencer & Lennard 2005) discussed teaching by humiliation. They called for an end to a culture of bullying, which had set in place a self perpetuating culture of abuse, in which the victims become the perpetrators in the next round. Bullying is commonly reported, with up to 84% of trainees reporting one or more bullying behaviours (Quine 2002). However, bullying was reported not just in macho specialties, but in psychiatry at 47% of trainees (Hoosen & Callaghan 2004) and even in palliative care at 40% of trainees (Keeley et al. 2004).

For these reasons it is essential to evaluate and monitor the educational climate in postgraduate medical education, using valid and reliable tools to do the job. One such tool is the PHEEM, developed in Scotland and the West Midlands (Roff et al. 2005). These researchers used a combination of

Correspondence: David Wall, Deputy Regional Postgraduate Dean, West Midlands Deanery, 150 Lichfield Road, Four Oaks, Sutton Coldfield, West Midlands B74 2TF, UK. Tel: 0121 323 5325; email: d.w.wall@btopenworld.com

grounded theory, focus groups, a nominal group method and a Delphi technique in a two-stage process to produce an inventory of 90 items. This was then reduced to a 40-item inventory by a focus group method using consultants and junior doctors in Birmingham. The PHEEM is a 40-item inventory using a series of statements, each marked on a five-item 0–4 scale. There are three subscales: perceptions of role autonomy, perceptions of teaching and perceptions of social support. Initial results showed a very high reliability using Cronbach's alpha of 0.91 in Birmingham.

In North America, attention is also being paid to the educational climate both for medical students and residents. For example, Grant et al. (2008) used the *Learners' Perceptions Survey* to examine students' and trainees' levels of satisfaction with the four separate domains of learning environment, clinical faculty, working environment and the physical environment.

Clapham et al. (2007) used the PHEEM for doctors in training in nine intensive care schemes throughout England and Wales, and again demonstrated a high reliability of 0.921. They were also able to demonstrate significant differences in the climate perceived by junior and senior trainees, with junior trainees scoring the educational climate better, and between different intensive care units. Factor analysis of their data showed 10 factors, responsible for 67% of the total variance. The top three factors encompassed 18 of the 40 questions and were described as issues relating to firstly 'the teacher', secondly 'learning doctoring skills in a safe environment' and thirdly 'a happiness index'. Looking at the questions loading onto these factors, there was good agreement with Roff's original three subscales. Clapham's *teacher* factor had six out of the eight of these questions from Roff's teaching perceptions subscale. The *learning in a safe environment* factor had three out of the five questions from Roff's role autonomy subscale. The *happiness* factor had three out of the four questions from Roff's social support subscale. So here there was some degree of statistical agreement using factor analysis with Roff's original three subscales produced using the qualitative methods of grounded theory, focus groups, a nominal group method and a Delphi technique.

Aspegren et al. (2007) validated the PHEEM in a wide selection of hospital departments in Denmark. It had excellent reliability measured by Cronbach's alpha of 0.93, and good agreement between the validations of PHEEM in the UK and in Denmark, with only four questions considered not relevant in Denmark. These related to hours of work, catering and accommodation. Again, Aspegren et al. used the three original subscales on the PHEEM (role autonomy, teaching and social support).

Lucas & Samarage (2008) used PHEEM in evaluating the clinical learning environment in paediatrics in Sri Lanka. They compared the educational environment in their three stages of postgraduate training, and found a significant difference. The more junior trainees scored significantly higher than their senior trainees. This fits with Clapham et al.'s (2007) findings (above) in intensive care medicine. However, these authors did not report the reliability of their data.

Boor et al. (2007) used factor analysis to study the psychometric properties of PHEEM in The Netherlands. Their study was of 595 junior doctors. Reliability was good. One of

their aims was to try to validate the construct validity of the three subscales of PHEEM. They claimed that no study had looked at this – although Clapham et al. (2007) had published their study in the same year, showing a total of 10 factors, of which the top three showed some agreement with Roff's three subscales (see above). They used exploratory factor analysis and found one factor responsible for 30.6% of the total variance. From their work they claimed that PHEEM was in fact a uni-dimensional instrument, but nevertheless was a reliable questionnaire for measuring the clinical learning environment.

One of the problems with exploratory factor analysis and the number of factors produced is the way the cut off point is chosen. Clapham et al. (2007) used a cut off accepting Eigenvalues of above 1.0. Boor et al. (2007) used the scree plot of Eigen values, and accepted one factor, and discarded factors below an Eigenvalue of 2.1. Field (2000) makes the point that there is debate as to how many factors should be accepted. Often, an Eigenvalue of above 1.0 is taken as the cut off – the Kaiser or Kaiser–Guttman criterion (Field 2000). A scree plot shows how the factors are distributed. However, Field points out that Cattell (1966) suggested that the inflexion point on the scree plot is the cut off, and all factors above that should be accepted. Sometimes the steep slope of the plot shows a sharp inflexion of the curve, but in other analyses there is a more gradual change in shape. It is then a matter of judgement and interpretation as to where to cut off the 'scree' and accept the factors. Also the amount of variance is important. Ideally, the factors should capture most of the variance rather than less than 50% of the total. Field (2000) suggested that research has shown that the Kaiser criterion should be used when there are fewer than 30 variables to put into the factor analysis. Above 30 variables, this criterion tends to retain too few factors and then the scree plot inflexion point should be used instead. As PHEEM has 40 items in its inventory, the scree plot inflexion point should probably be used in establishing how many factors are present.

In addition, there is the theoretical epistemological point here of whether an interpretivist or a positivist framework is the 'correct' one to follow. Roff et al. (2005) used a number of interpretivist methods to discover and construct the social reality of what is the educational environment. Can this be proven or not on the basis of another different epistemological process, a positivist mathematical and statistical approach? Some would say not (Grix 2001). However, some would feel vindicated if both viewpoints came to similar conclusions.

Because of these differences, we decided to try to resolve the issue of the characteristics of PHEEM and its three subscales. Perhaps, if we could recruit large numbers of sets of data from several countries, we could look again at the reliability and at the factor analysis and how the factor or factors correlated with Roff's three subscales. Our research question was as follows: Is PHEEM a uni-dimensional or multi-dimensional instrument when looked at using statistical methods?

Methods

Discussions with colleagues using PHEEM from Brazil, Chile, Denmark, Scotland and England enabled us to gather data

from doctors' evaluations of the educational environment from Brazil, Chile, Denmark, The Netherlands and the UK.

The UK data was from senior house officers and specialist registrars from 10 hospitals in Intensive Care Medicine in England and Scotland, and from the Foundation Programme (doctors in the first 2 years of postgraduate training) in a large teaching hospital on three sites in Birmingham in England. The data from Brazil was from junior doctors in all specialties at the Clinics Hospital of Sao Paulo, a tertiary hospital and the largest in Brazil. The data from Chile was from clerks in years six and seven (comparable to the UK Foundation years) from eight specialties in four different hospitals as part of the Pontificia Universidad Católica de Chile Medical School in Santiago. The data from The Netherlands was from clerks from 14 specialties in six different hospitals, and from registrars in paediatrics from 25 hospitals and from obstetrics in 40 hospitals. These sites and researchers were chosen because of their research interests in researching the educational climate using PHEEM.

After further discussion, the Denmark data was not utilised further because their data was a validation study of how appropriate the questions of PHEEM were, not a measure of the Danish hospital educational environment (Aspegren et al. 2007).

All sets of data were used with the original 0 to 4 scale (as described by Roff et al. 2005) rather than the 1–5 scale used by Boor et al. 2007) and by Lucas & Samarage (2008). The scores for questions 7, 8, 11 and 13 of the PHEEM, which contained negative statements, were reverse coded in line with Roff's original guidance (Roff et al. 2005) before any further statistical work was carried out.

Statistical analysis was carried out using SPSS version 15. Descriptive statistics were calculated in terms of the overall scores and the scores for the three subscales, both for the pooled data and for each of the five separate data sets (Brazil, Chile, The Netherlands, UK Intensive Care and UK Foundation Programme). Tests of statistical significance were carried out using the Kruskal Wallis test and ANOVA. Overall reliability was calculated using Cronbach's alpha.

Exploratory factor analysis was carried out using both the Kaiser criterion (accepting factors with an Eigenvalue above 1.0) and the Cattell criterion (accepting factors above the inflexion point in the scree plot curve), as discussed above. Exploratory factor analysis included the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of adequacy of sampling of the data. Varimax rotation was used, accepting factor loadings above 0.5.

Once factors were derived, the questions loading on each derived factor were matched against the questions in the original PHEEM three subscales (perceptions of role

autonomy, perceptions of teaching and perceptions of social support). In addition, correlations were calculated between the scores of the derived factors and the scores of the three original PHEEM subscales.

Results

A total of 1563 sets of data from doctors in training were used. The numbers from each of the five sources were as follows:

Brazil	306
Chile	125
The Netherlands	595
UK Intensive Care	278
UK Foundation	259

Results for the overall scores and for the scores of the three subscales are shown in Table 1. Also, the maximum score possible overall and in each subscale are indicated in this table. A box plot of the overall scores for the five sources appears in Figure 1.

Both a Kruskal Wallis test and ANOVA showed significant differences in the total scores of the five sources, with a *p*-value of less than 0.001 for the Kruskal Wallis test and a *p*-value of less than 0.001 for the ANOVA.

Reliability of the PHEEM using the whole 1563 sets of data was excellent at 0.928. Exploratory factor analysis revealed excellent adequacy of sampling with a KMO test value of 0.952. Bartlett's test was highly statistically significant.

Using the Kaiser criterion accepting Eigenvalues above 1.0, seven factors were extracted after nine iterations accounting for 56% of the total variance. Looking at the scree plot (Figure 2), there is an inflexion point in the plot after three factors.

Therefore the factor analysis was run again, fixed at extracting three factors (the Cattell criterion). Using this method, three factors were extracted after six iterations, responsible for 43% of the total variance. There were 30 questions (out of the total of 40 questions in the PHEEM inventory) which loaded onto these three factors. The way these 30 questions loading onto these three factors is shown in Table 2. Factor One appears to be largely related to questions about teaching, Factor Two appears to be largely related to questions about role autonomy and Factor Three appears to be largely related to social support.

Factor One contained 18 questions, of which 11 were from the teaching subscale, four from the social support subscale and four from the role autonomy subscale. Factor Two contained eight questions, of which four were from the role autonomy subscale, three from the teaching subscale and one

Table 1. Results for the overall PHEEM scores and for the three subscales.

PHEEM domain	Brazil	Chile	Netherlands	UK ITU	UK Foundation	Overall	Maximum scores possible
Global score	95	98	105	112	113	105	160
Role autonomy	34	34	37	40	38	37	56
Teaching	35	37	41	43	45	40	60
Social support	27	27	27	30	30	28	44

from the social support subscale. Factor Three contained four questions, of which the two strongest loaded were from the social support subscale and two others from the role autonomy subscale.

Correlations between the three factors and the three subscales are shown in Table 3. Factor One correlated most strongly with the perceptions of teaching subscale ($R=0.802$, $p < 0.01$). Factor Two correlated most strongly with the perceptions of role autonomy ($R=0.623$, $p < 0.01$). Factor Three correlated most strongly with the perceptions of social

support ($R=0.538$, $p < 0.01$). In addition, the three subscale values from the PHEEM did correlate with each other, as shown in the Table 3.

Discussion

We believe that we have shown that PHEEM is not a uni-dimensional instrument. Indeed, using exploratory factor analysis and applying the Cattell criterion (the inflexion point

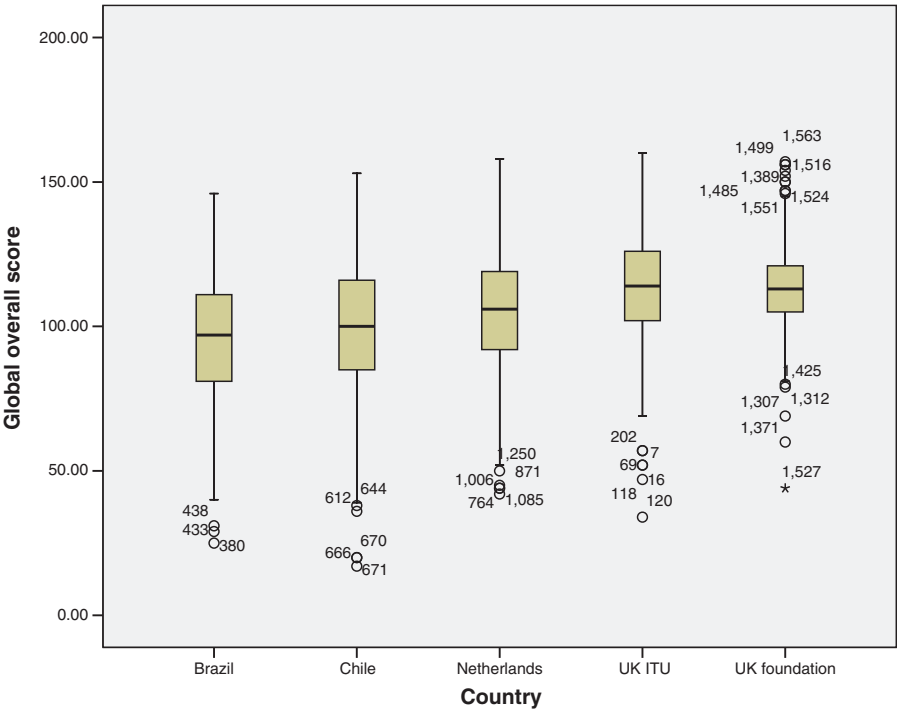


Figure 1. The box plot of the overall scores from the five data sources of PHEEM.

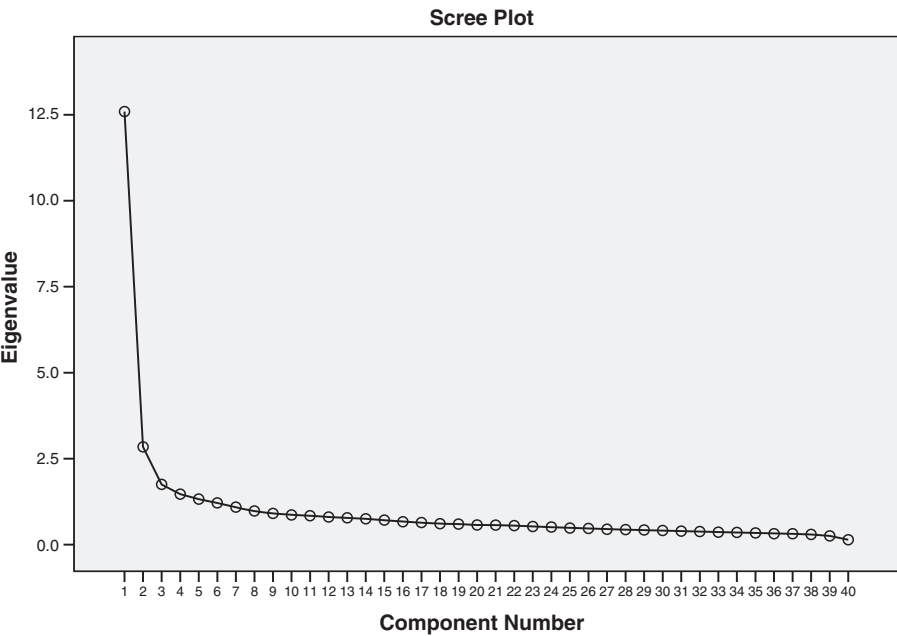


Figure 2. The scree plot from the exploratory factor analysis.

on the scree plot graph), as recommended by Field (2000) for a data set with more than 30 variables, we find that there are three distinct factors. The strongest factor, Factor One of 18 questions, contained mainly questions about teaching, and is

Table 2. Results of the factor analysis showing the loadings of PHEEM questions onto the three factors.

Question number	Factor One	Factor Two	Factor Three
q35	0.741 SS		
q22	0.730 T		
q39	0.704 T		
q33	0.695 T		
q31	0.664 T		
q23	0.655 T		
q29	0.648 RA		
q37	0.642 T		
q40	0.631 RA		
q28	0.619 T		
q15	0.611 T		
q30	0.580 RA		
q2	0.566 T		
q19	0.539 SS		
q10	0.539 T		
q36	0.538 SS		
q27	0.532 T		
q34	0.505 RA		
q38			
q21			
q18			
q5			
q4			
q9			
q17		0.679 RA	
q3		0.668 T	
q32		0.659 RA	
q1		0.588 RA	
q12		0.552 T	
q14		0.508 RA	
q6		0.505 T	
q25		0.502 SS	
q16			
q24			
q26			
q20			
q7_r			0.871 SS
q13_r			0.871 SS
q8_r			0.766 RA
q11_r			0.575 RA

Note: RA is Role autonomy; T is Teaching; SS is Social support and _r indicates reverse coding for this question.

in fact correlated most strongly with the perceptions of teaching subscale in the original PHEEM paper (Roff et al. 2005). This factor was responsible for 22% of the total variance in the data. The next factor, Factor Two of eight questions, contained four from the perceptions of role autonomy, three from teaching and one from social support. It correlated most strongly with the perceptions of role autonomy subscale in the original PHEEM paper. It contributed to a further 13% of the total variance in the data. The third factor, Factor Three, contained four questions, two strongest loading from the perceptions of social support and two from role autonomy. It correlated most strongly with the perceptions of social support in the original PHEEM paper. It contributed a further 8% of the total variance. Altogether, these three factors made 43% of the total variance in the data.

Looking at the adequacy of sampling of our data, the KMO test gave a very high value of 0.952. Field (2000) states that a value of greater than 0.5 shows that the sample is adequate, and a value close to one (which this is here – at 0.952) shows that the patterns of correlations are fairly compact and so factor analysis should yield distinct and reliable factors (Field 2000, p. 455).

These results differ from those of Boor et al. (2007) in that their study showed only one factor responsible for 31% of the variance. They did neither state which of the 40 questions of the original PHEEM loaded onto this factor, nor how these questions related to the three PHEEM subscales. In addition, with one factor capturing only 31% of the variance, the remaining 69% of the data is not accounted for in this analysis. We have done somewhat better here, accounting for 43% of the total variance with three factors.

Overall reliability was good, with a value of 0.928 for Cronbach's alpha here, which is very similar to the value of 0.91 for Roff et al. (2005), for the value of 0.921 for Clapham et al. (2007) and the value of 0.93 for Aspegren et al. (2007). So in all these four studies the reliability of PHEEM was good. However, caution is needed in interpreting these values in an instrument such as PHEEM with 40 items, as Cronbach's alpha is dependant to some extent on the number of items. An instrument with 40 items may generate a higher value than one with (say) 10 items with equivalent reliability.

The scores for PHEEM, both in terms of global overall scores and the three subscales did differ between the five

Table 3. Results of the correlations of the three factors and the three original PHEEM subscales scores.

		Role autonomy	Teaching	Social support	Factor One	Factor Two	Factor Three
Role autonomy	Pearson correlation	1	0.802(**)	0.698(**)	0.608(**)	0.632(**)	0.289(**)
	Sigma (2-tailed)		0.000	0.000	0.000	0.000	0.000
Teaching	Pearson correlation	0.802(**)	1	0.681(**)	0.800(**)	0.517(**)	0.172(**)
	Sigma (2-tailed)			0.000	0.000	0.000	0.000
Social support	Pearson correlation	0.698(**)	0.681(**)	1	0.565(**)	0.385(**)	0.535(**)
	Sigma (2-tailed)				0.000	0.000	0.000
Factor One	Pearson correlation	0.608(**)	0.800(**)	0.565(**)	1	0.000	0.000
	Sigma (2-tailed)					1.000	1.000
Factor Two	Pearson correlation	0.632(**)	0.517(**)	0.385(**)	0.000	1	0.000
	Sigma (2-tailed)						1.000
Factor Three	Pearson correlation	0.289(**)	0.172(**)	0.535(**)	0.000	0.000	1
	Sigma (2-tailed)						

Note: **Correlation is significant at the 0.01 level (2-tailed).

subsets of our data. The reasons for this are not obvious to us, and further work will be needed to evaluate the reasons for this. This was not part of our research question, but is reported here as the first multi-national study in which we were able to compare data from several countries in terms of the educational climate in postgraduate medical education. We used both the Kruskal Wallis non-parametric test and the parametric ANOVA, and found significant differences using both tests. Many would say that the Kruskal Wallis test is the more appropriate test to use on Likert data (Field 2000). There has been debate on the appropriateness of using ANOVA for Likert-type data (Pell 2005; Carifio & Perla 2007). This debate has been going on for over 50 years. In particular, Pell points out that '... it is acceptable in many cases to apply parametric techniques to non-parametric data such as that generated from Likert scales ...'. It all depends on the characteristics of the data being used. Also, other statistical techniques we use frequently on such data, such as regression and factor analysis, do use mean scores in their statistical calculations (Pell 2005).

PHEEM, as a 40 item educational climate tool, does seem to us to be a multi-dimensional instrument. There are good significant correlations between the three factors we have derived from statistical testing and the three subscales produced by Roff and her colleagues using qualitative methods grounded theory, focus groups, a nominal group method and a Delphi technique. So we did not find that there was a clash of the two cultures, in terms of educational research methods between the qualitative interpretivist framework and the quantitative positivist framework. Both gave a good degree of agreement in terms of the three subscales derived qualitatively and the three main factors derived using statistical methods.

What about our conceptualisation of the three factors found here, as well as looking at the levels of agreement with the three subscales of PHEEM? Looking carefully at the factors, and the statements which load onto the three factors, we may try to re-conceptualise these three factors.

Factor One has the top four statements (in descending order) as good mentoring skills, regular feedback, good feedback on strengths and weaknesses and seniors using learning opportunities appropriately. So this may be about senior doctor support and teaching skills.

Factor Two has the top four statements (in descending order) as hours of work, protected educational time, workload and contract of employment. So this may be about the conditions of working and time to learn.

Factor Three has the top four statements (in descending order) as the absence of racism, absence of sexism, lack of inappropriate tasks and lack of being bleeped inappropriately. So this may be about the lack of harassment of all kinds.

In addition, we have found that the scores for the three PHEEM subscales do indeed correlate with each other (Table 3). In our view this is not surprising. A good educational climate will contain good scores from all three subscales, and a poor one the reverse. In fact, this finding has not been described before, possibly because no one has tested their data in this way. This finding should be tested and hopefully repeated with other data.

Ethical considerations

Ethical approval was not required. This was an evaluation of the existing data, which had been used to evaluate the educational climate in the various hospital environments. All data contributing for this analysis was anonymised so that D. Wall who did the statistical analysis had no access to individual named data.

Acknowledgements

We gratefully acknowledge and thank Dr K. Boor and colleagues from The Netherlands for the permission to use their PHEEM data, confirmed by Dr Sue Roff from the Centre for Medical Education in Dundee, Scotland.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Notes on contributors

DAVID WALL, MB ChB, MMed, PhD, FRCP, FRCGP is a Deputy Regional Postgraduate Dean and Professor of Medical Education in the West Midlands Deanery, UK.

MIKE CLAPHAM, MB BS, MSc, PhD, FRCA is an Associate Dean for Education in the West Midlands Deanery, UK, and the Director of Medical Education and a Consultant in Intensive Care at University Hospital Birmingham.

ARNOLDO RIQUELME, MD, MMedEd is an Undergraduate and Postgraduate Clinical Tutor in Internal Medicine and Consultant in the Department of Gastroenterology at the Pontificia Universidad Catolica de Chile School of Medicine, Chile.

JOAQUIM VIEIRA, MD, PhD is a Secretary of Coordination for the Centre for Development of Medical Education 'Prof. Eduardo Marcondes' – University of Sao Paulo Medical School and an Assistant Physician in Anaesthesia Division at Hospital das Clinicas, University of Sao Paulo Medical School, Sao Paulo, Brazil.

RICHARD CARTMILL, MB BCh BAO, MMed, FRCOG is a Postgraduate Clinical Tutor at Good Hope Hospital in Sutton Coldfield, UK, and a Consultant Obstetrician and Gynaecologist at Good Hope Hospital, Sutton Coldfield, UK.

KNUT ASPEGREN, MD, PhD has recently retired and was Professor of Medical Education at the Postgraduate Deanery, University of Southern Denmark.

SUE ROFF, BA (Hons), MA teaches in the Centre for Medical Education, University of Dundee, UK. She co-developed the DREEM and the PHEEM. She is a lay member of the General Medical Council Fitness to Practice Committee and the Postgraduate Medical Education and Training Board.

References

- Aspegren K, Bastholt L, Basted KM, Bonnesen T, Ejlersen E, Fog I, Hertel T, Kodal T, Lund J, Madsen JS, et al. 2007. Validation of the PHEEM instrument in a Danish Hospital Setting. *Med Teach* 29:504–506.
- Boor K, Scheele F, van der Vleuten CPM, Scherpbier AJJA, Teunissen PW, Sijtsma K. 2007. Psychometric properties of an instrument to measure the clinical learning environment. *Med Educ* 41:92–99.
- Carifio J, Perla RJ. 2007. Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *J Soc Sci* 3:106–116.
- Cattell RB. 1966. The scree test for a number of factors. *Multivar Behav Res* 1:245–276.
- Chambers R, Wall D. 2000. *Teaching made easy: A manual for health professionals*. Abingdon: Radcliffe Medical Press Limited.

- Clapham M, Wall D, Batchelor A. 2007. Educational environment in intensive care medicine – Use of Postgraduate Hospital Educational Environment Measure (PHEEM). *Med Teach* 29:e184–e191.
- Field A. 2000. *Discovering statistics using SPSS for Windows*. London: Sage Publications Limited.
- Grant CW, Keitz SA, Holland GJ, Chang BK, Byrne JM, Tomolo A, Aron DC, Wicker AB, Kashner TM. 2008. Factors determining medical students' and residents' satisfaction during VA based training: Findings from the VA Learners' Perceptions Survey. *Acad Med* 83:611–620.
- Grix J. 2001. *Demystifying Postgraduate Research*, Birmingham: University of Birmingham Press.
- Hoosen IA, Callaghan R. 2004. A survey of workplace bullying of psychiatric trainees in the West. Midlands. *Psychiat Bull* 28:225–227.
- Keeley PW, Waterhouse ET, Noble SIR. 2004. Not just in macho specialties – Palliative medicine too. *BMJ.com* – rapid responses 17 September 2004 (accessed 25 October 2005).
- Lucas MN, Samarage DK. 2008. Trainees' perceptions of the clinical learning environment in the postgraduate training programme in paediatrics. *Sri Lanka J Child Health* 37:76–80.
- Pell G. 2005. Use and abuse of Likert scales. *Med Educ* 39:970.
- Quine L. 2002. Workplace bullying in junior doctors: Questionnaire survey. *BMJ* 324:878–879.
- Roff S, McAleer S, Skinner A. 2005. Development and validation of an instrument to measure the postgraduate clinical learning environment for hospital based junior doctors in the UK. *Med Teach* 27:326–331.
- Spencer J. 2003. Learning and teaching in the clinical environment. *BMJ* 326:591–594.
- Spencer J, Lennard T. 2005. Time for gun control? *Med Educ* 39:868–869.