**Original research paper**

# Development of the clinical learning climate measure for undergraduate medical education

*Danai Wangsaturaka, Sean McAleer*

**Abstract**

This study aimed to develop an instrument to measure the clinical learning climate of undergraduate medical education in Thailand. There were 3 stages of data collection. The first stage employed qualitative inquiry to identify key attributes of the learning environment. Stage 1 output was arranged into the items in the questionnaire for Stage 2 data collection. The validating process in Stage 2 resulted in the list of factors influencing the clinical learning climate with the degree of importance of each factor. The initial version of the learning climate measure was developed based on the findings from this stage. Stage 3 aimed to collect data for further refinement of the instrument. Item analysis and factor analysis were used to explore the constructs of the climate measure. The subscales were established using both statistical and qualitative factor analysis. The final output of this study is the 43-item learning climate measure for clinical phase (10 subscales). Finally, the psychometric properties of the instrument were discussed. The utility of the learning climate measure was recommended.

## Introduction

Many inventories have been developed to measure the learning climate of educational programmes. Nevertheless, during the past ten years, there have been only eight instruments designed specifically for medical education. Of these, only the Dundee Ready Education Environment Measure (Roff *et al*, 1997) and the unnamed inventory developed by Pololi and Price (2000) focus on undergraduate medical education. However, one instrument may not be sensitive enough to measure the learning climate of the entirety of undergraduate medical education as factors influencing medical students' learning climate may differ from one year to another. For example, first year students usually learn through lectures, practical classes or small group learning with little or no clinical input; whereas final year students spend most of their time in wards, outpatient clinics or other clinical settings with very few or no lectures. Therefore, items relating to clinical experience should be important for final year but not for first year students. In other words, it would seem appropriate to have specific learning climate measures for each phase of the undergraduate medical education.

To the researcher's knowledge, there are no studies which examine differences in the composition of the learning environment in each phase of undergraduate medical education. The research was conducted to trace the evolution of the undergraduate learning environment and to further determine the key features of the postgraduate learning environment in order to develop the climate measures for each phase of medical education. This article will describe only the development of the learning climate measure for the clinical phase of undergraduate medical education.

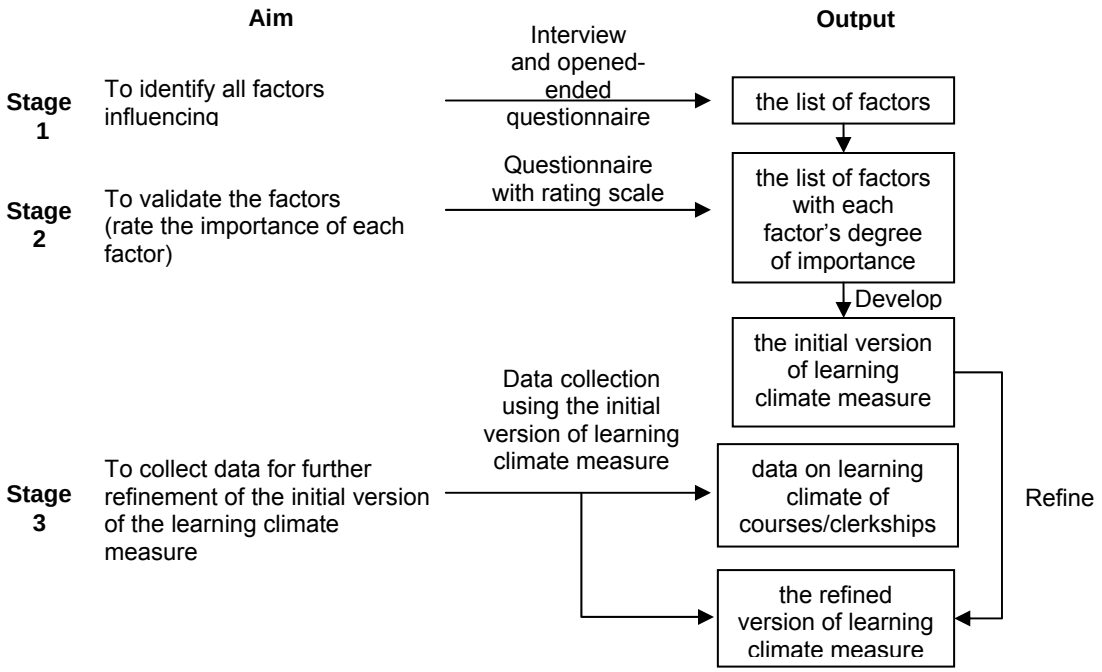
## Methods

***Research Design*** This study was divided into three stages. In Stage 1, qualitative inquiry was used to identify key attributes of the learning environment. The issues generated in this stage were arranged into the items in the questionnaire. In Stage 2, the respondents rated the importance of each item. The output was used to develop the initial version of the learning climate measure. The inventory was then refined using data obtained from administering the preliminary instrument in Stage 3. This three-stage approach is summarized in Figure 1.

[1]*MD PhD, Department of Pharmacology*
*Faculty of Medicine, Chulalongkorn University*
*Rama IV Road, Patumwan, Bangkok, 10330 Thailand*
*Email naidanai@yahoo.co.uk*

[2]*DPhil, Centre for Medical Education, University of Dundee, 484 Perth Road, Dundee, DD2 1LR, Scotland, United Kingdom*
*Email j.p.g.mcaleer@dundee.ac.uk*

***Figure 1: Research design for instrument development***




***Stage 1***

The target populations in this stage were clinical students, faculty, and executives (e.g. the Associate Dean for academic affairs) in six medical schools in Thailand. In each school, one executive and two members of the clinical faculty were purposefully selected and individually interviewed. A group interview was used for collecting data from year 5 students with 6-8 students per school. The technique used in a group interview was a combination of the focus group, the nominal group and the critical incident techniques. At one medical school, an open-ended questionnaire was used in addition to the group interview with the whole class of Year 4 students participating in this data inquiry method. The key questions asked during the data collecting process were:

- What is a good learning environment for clinical students?
- What is a bad learning environment for clinical students?

Additional questions which were asked to prompt further responses were:

- What kind of environment makes clinical students happy?
- What kind of environment makes clinical students unhappy?
- What kind of environment helps clinical students succeed in their study?
- What kind of environment makes clinical students unsuccessful in their study?
- How can the medical school create a better environment for clinical students?
- What might be current problem(s) for clinical students in other medical schools?

The records of both group interviews and individual interviews were transcribed by the researcher. The interview transcripts, the respondents' answering sheets, and the notes taken by the researcher during the interviews was analysed in order to answer the research question – what are the factors influencing the learning climate of clinical phase? The features of the learning environment identified were complied into a list of items. The change of original wording used by the respondents was kept to a minimum. This resulted in the Stage 1 output – a list of the potential factors influencing the clinical learning climate.

***Stage 2***

The target population (n) included 1,028 Fifth year students at 15 medical schools in Thailand. The formulae used to calculate the sample size (n) were:

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

$$n_0 = \frac{Z^2 pq}{e^2}$$

(Israel, 1992)

As the desired level of precision (e) used in this study was .05, the $Z^*$ equaled 1.96. Assuming $p^{**}$ = .5 which was the maximum variability resulted in $q^{***}$ = .5.

* $Z^2$ is the abscissa of the normal curve that cuts off an area α at the tails (1-α equals the desired confidence level)

** p = estimated proportion of an attribute that is present in the population

*** q = 1 - p

The value of $n_0$ was 385 and the sample size (n) was 292 students, shown in the equations below:

$$n_0 = \frac{Z^2 pq}{e^2} = \frac{(1.96)^2 (.5)(.5)}{(.05)^2} = 385$$

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}} = \frac{385}{1 + \frac{385 - 1}{1028}} = 281$$

The calculated sample size was increased 20% to compensate for non-responders resulting in the sample size of 345 students. Stratified sampling was used to ensure that the samples would reflect the proportions of (1) males to females and (2) the number of students in each medical school. In each medical school, both male and female samples were randomly selected within each gender by using the Statistical Package for Social Science (SPSS) for Windows Release 10.1.0 – except at two medical schools where convenience sampling had to be employed.

The questionnaires constructed for this stage were composed of two parts: general information and close-ended questions. In the first part, participants were asked about their gender, route of entry and cumulative grade point average (GPAX). The second part contained 140 statements based on the results of Stage 1. A five-point rating scale, running from not important at all (0), slightly important (1), somewhat important (2), very important (3) and absolutely important (4), was applied to all items. The respondents were asked to rate the items according to the perceived importance to their clinical learning climate.

Data from the questionnaires were entered onto the Microsoft Excel 97 Version 8.0. The data were double-checked. The amended data were analysed using both the Statistical Package for Social Science (SPSS) for Windows Release 10.1.0 and the Microsoft Excel 2002 SP3.

**Results**

Three hundred and seven questionnaires were returned from the 345 sampled in 15 institutions, resulting in a response rate of 89.0%. One hundred and forty-seven respondents (47.9%) were male. The percentage of male participants is slightly higher than the percentage of males in the sample (47.0%).

One hundred and forty statements in the questionnaires were analysed and classified into 9 domains: Teachers (21 items), Human Environment (24 items), Patient Care (5 items), Learning Experience (18 items), Assessment (12 items), Educational Resources (13 items), Physical Environment (17 items), Social and Personal Life (19 items), and Miscellaneous (11 items).

The items' mean scores ranged from 1.35 to 3.74, two-thirds of which were rated very important (mean score ≥ 3.00). The number of items classified by their degree of importance is shown in Figure 2.

Table 1 shows the 10 most important items as perceived by the clinical students. They were categorised into the following domains: Teachers (3), Social and Personal Life (3), Learning Experience (2), Assessment (1) and Miscellaneous (1).

**Instrument development**

It would not be practical if all 140 items were used in the clinical learning climate measure since students would have little attention in completing a long questionnaire. Only the items rated 'very important' (mean score ≥ 3.00) were, therefore, included to increase the validity of the instrument. However, this still resulted in the 94-item inventory. The 65$^{th}$ percentile was, hence, chosen as a cut point, resulting in the 49-item inventory. The numbers of items in each domain of the inventory were as follows: Teachers (11 items), Human Environment (4 items), Patient Care (1 item), Learning Experience (8 items), Assessment (2 items), Educational Resources (4 items), Physical Environment (7 items), Social and Personal Life (6 items), and Miscellaneous (6 items).

.

***Figure 2: Bar chart illustrating percentages of items classified by their levels of importance***

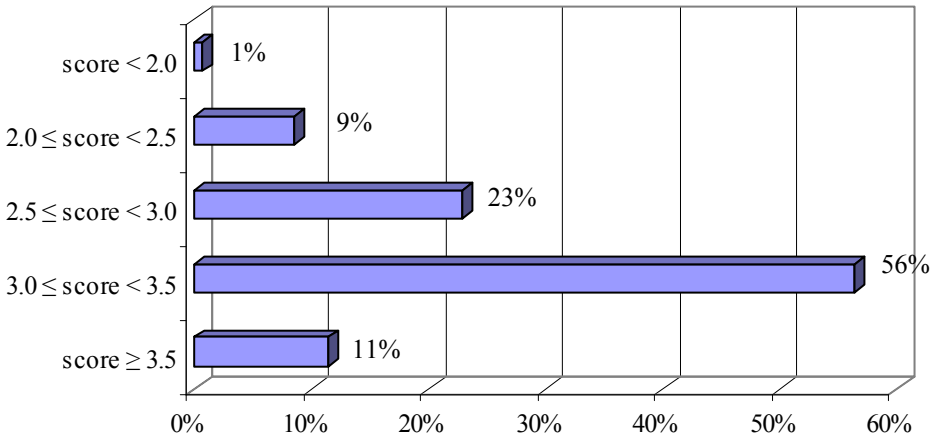



***Table 1. The ten most important items influencing the clinical learning climate***

| Domain | | Items | Mean (SD) |
|---|---|---|---|
| Teachers | 1. | Teachers teaching knowledge which is relevant to patient care. | 3.74 (0.52) |
| Teachers | 6. | Teachers having good teaching skills. | 3.71 (0.50) |
| Social and personal life | 108. | Having supportive family. | 3.70 (0.64) |
| Learning experience | 71. | Having enough variety of cases for learning. | 3.63 (0.59) |
| Social and personal life | 47. | Getting along well with friends. | 3.61 (0.61) |
| Learning experience | 70. | Having opportunities to practise at outpatient clinics. | 3.59 (0.62) |
| Social and personal life | 140. | Having some personal time. | 3.59 (0.66) |
| Teachers | 20. | Teachers being open-minded to students' opinions. | 3.58 (0.62) |
| Assessment | 83. | The content assessed focusing on practical points. | 3.58 (0.65) |
| Miscellaneous | 62. | Feeling eager to learn. | 3.57 (0.60) |

The 49 items were arranged into the initial version of the instrument. The format of the statements was changed from 'phrase' to 'sentence'. Some phrases were modified into personalised sentences so that responses would reflect individual's perception e.g. *'having close friend(s)'* was changed to *'I have close friend(s)'*. A five-point Likert scale ranging from 'strongly disagree' to 'strongly agree' was chosen as the response option for all items in all phases. Answering options were scored as follows: strongly disagree = 0, disagree = 1, uncertain = 2, agree = 3 and strongly agree = 4. After the pilot study, the questionnaire was distributed to all Year 4 and 5 students at Chulalongkorn Medical School and its three affiliated hospitals for instrument field-testing.

Three hundred and ninety-two questionnaires were returned for further analysis. Of the 10 departments and the 3 affiliated hospitals, the Cronbach's alphas ranged from 0.67 to 0.95 with only one department having the alpha less than 0.8. The item analysis using item-total correlation was then performed to establish the validity evidence from internal structure of the test. Any items with the item-total correlation less than 0.3 were removed. The constructs of the inventory were explored using factor analysis. As there were no pre-assumed loadings, it could be classified as 'exploratory factor analysis'. The data were factor analysed using the principal component method of extraction. The Kaiser-Guttman rule (eigenvalue > 1.0) was applied to determine the number of factors to retain for rotation. Varimax rotation was chosen as it would *'maximise the variance explained by each factor'* (Norman and Streiner, 1997).

The data from the ten clerkships and the three affiliated hospitals were combined. Six items were removed before performing factor-analysis as they had the item-total correlation less than 0.3. The factor analysis showed eleven factors with eigenvalues higher than 1, accounting for 63.10% of the variance. Table 2 shows the items with loadings greater than or equal to 0.30 on the eleven factors.

In order to establish the subscales of the learning climate measure, the results from the factor analysis and the classification of the items according to the nine domains were considered together. The combination of these two methods was expected to prevent meaningless statistical grouping of the items while maintaining the construct established by the factor analysis as much as possible.

There were ten subscales established for the clinical learning climate measure: teachers, colleagues, learning experience, ward environment, assessment, educational resources, physical environment, motivation, health and stress, and institutional environment. Table 3 shows the details of each subscale including their internal consistency coefficients.

***Table 2: Factor loadings and factor structure of the clinical learning climate measure with varimax rotation (normalised; n = 376)***

| Items | Components | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 5 | .787 | | | | | | | | | | |
| 1 | .751 | | | | | | | | | | |
| 4 | .735 | | | | | | | | | | |
| 3 | .693 | | | | | | | | | | |
| 2 | .676 | | | | | | | | | | |
| 6 | .568 | | | | | | | | | | |
| 7 | .565 | | | | | | | | | | |
| 9 | .564 | | | | | | | | | | |
| 10 | .458 | | | | | | | | | | |
| 47 | | .752 | | | | | | | | | |
| 45 | | .733 | | | | | | | | | |
| 48 | | .724 | | | | | | | | | |
| 44 | | .701 | | | | | | | | | |
| 49 | | .559 | | | | | | | | | |
| 46 | .355 | .513 | | | | | | | | | |
| 41 | | | .862 | | | | | | | | |
| 42 | | | .834 | | | | | | | | |
| 29 | | | .627 | | | | | .316 | | | .399 |
| 40 | | | .535 | | | | .405 | | | | |
| 19 | | | | .819 | | | | | | | |
| 17 | | | | .771 | | | | | | | |
| 18 | | | | .765 | | | | | | | |
| 20 | .322 | | | .638 | | | | | | | |
| 34 | | | | | .713 | | | | | | |
| 35 | | | | | .685 | | | | | | |
| 37 | | .319 | | | .500 | | .430 | | | | |
| 30 | | | .449 | | .467 | | | .380 | | | |
| 12 | | | | | | .800 | | | | | |
| 11 | | | | | | .719 | | | | | |
| 13 | | | | | | .495 | | | | .469 | |
| 36 | | | | | | | .752 | | | | |
| 38 | | | | | | | .742 | | | | |
| 39 | | | | | .397 | | .491 | | | | |
| 33 | | | | | | | | .770 | | | |
| 31 | | | | | | | | .643 | | | |
| 32 | .321 | | | | | | | .416 | | .365 | |
| 25 | | | | | | | | | .768 | | |
| 26 | | | | | | | | | .761 | | |
| 27 | | | | | | .338 | | | .510 | | |
| 15 | | | | | | | | | | .757 | |
| 16 | | | | | | | | .384 | | .439 | |
| 28 | | | | | | | | | | | .685 |
| 24 | | | | | .344 | | | | .359 | | .368 |

***Table 3: The subscales of the clinical learning climate measure***

| Subscale | Items |
|---|---|
| Teachers<br>(9 items)<br>alpha = .87 | 1. Teachers are enthusiastic in their teaching of students.<br>2. Teachers teach knowledge which is relevant to patient care.<br>3. Teachers have good teaching skills.<br>4. Teachers show that they take good care of patients.<br>5. Teachers understand and care for their students.<br>6. Teachers are patient with students when they do not know about something.<br>7. Teachers are open-minded to students' opinions.<br>9. Teachers are good ethical role models.<br>10. Teachers write handouts which cover all the important content. |
| Colleagues<br>(2 items)<br>alpha = .73 | 11. Residents/house officers/externs are friendly to students.<br>12. Residents/house officers/externs are pleased to supervise students' procedural practice. |
| Ward environment<br>(4 items)<br>alpha = .64 | 24. Rooms for bedside laboratory are clean and well-equipped.<br>25. Classmates help each other complete assignments.<br>26. Classmates are unselfish.<br>27. Performance assessment in the wards is fair. |
| Learning experience<br>(6 items)<br>alpha = .71 | 13. I have opportunities to do procedures which are of minimal requirement.<br>15. There is enough variety of cases for learning.<br>16. I have educational resources at hand when seeing patients.<br>31. I know which books/textbooks are recommended for the course.<br>32. Handouts/textbooks are easy-to-understand.<br>33. I am able to read English textbooks fluently. |
| Assessment<br>(1 item) | 28. The content assessed focuses on practical points. |
| Educational resources<br>(3 items)<br>alpha = .62 | 30. I have a quiet and temperate study area with enough light.<br>34. There are enough new books in libraries.<br>35. Libraries' opening hours are suitable for students. |
| Physical environment<br>(4 items)<br>alpha = .69 | 36. The dormitory is near the working place.<br>37. The environment of the dormitory is pleasant.<br>38. I feel safe in the hospital including between wards and the dormitory.<br>39. The medical school has a lush green environment. |
| Motivation<br>(4 items)<br>alpha = .80 | 17. I enjoy taking care of patients.<br>18. I want to be a doctor.<br>19. I feel eager to learn.<br>20. I like the current department/ward. |
| Health and stress<br>(4 items)<br>alpha = .79 | 29. I have enough time for self-study.<br>40. I feel healthy (on this course).<br>41. I have enough time to rest.<br>42. I have some personal time. |
| Institutional environment<br>(6 items)<br>alpha = .81 | 44. I get along well with my friends.<br>45. Junior and senior students are united.<br>46. The advisor takes good care of me.<br>47. The medical school has a friendly atmosphere.<br>48. I am proud of this institution.<br>49. I am confident that I will not be less competent than graduates from other institutions. |

**The survey of clinical learning climate**

The 43-item instrument to assess the clinical learning climate comprises of ten subscales with items ranging from 1 to 9. A five-point Likert scale ranging from 'strongly agree' to 'strong disagree' is used as the response option which is scored as the followings: strongly disagree = 0, disagree = 1, uncertain = 2, agree = 3, and strongly agree = 4. As a result, it has a maximum score of 172 (100%) and the minimum of 0 (0%). Nevertheless, since the number of the items in the subscales differs from each other, the maximum score of every subscale is adjusted to ten for inter-subscale comparison. This inventory contains no negative statements.

The clinical students participating in the field-testing step were from Chulalongkorn Medical School (CU), Bhumibol Adulyadej Hospital (BAH), Chon Buri Hospital (CBH) and Prapokklao Chanthaburi Hospital (PCH). Their response rates were, respectively, 94.2% (277/294), 95.7% (44/46), 100% (37/37) and 94.4% (34/36). Sixteen questionnaires were excluded as they contained missing values.

The maximum scores of the overall learning climate at these hospitals were 100% (CU), 86% (BAH), 87.8% (CBH) and 87.8% (PCH). The minimum scores were 36.6% (CU), 45.3% (BAH), 50% (CBH) and 50% (PCH). The average scores were 71.3% (CU), 67.4% (BAH), 72.7% (CBH) and 71.9% (PCH).

Table 4 shows that 'teachers', 'colleagues' and 'institutional environment' were the three most pleasant subscales for the clinical students at Chulalongkorn Medical School, Chon Buri Hospital and Prapokklao Chanthaburi Hospital. The three subscales received highest scores from BAH clinical students were colleagues, motivation, and teachers. The scores of 'educational resources' and 'health and stress' subscales were considerably low in all hospitals. So were the ward environments in the affiliated hospitals.

Comparing between the four hospitals, Chon Buri Hospital was rated higher than the others in total score and in five subscales: teachers, colleagues, learning experience, assessment, and institutional environment. This might be due to the characteristics of the CBH students or Chon Buri Hospital due to providing a better learning environment for their students.

Focusing on the Chulalongkorn Medical School, the four major departments were chosen for comparison of their learning environment: Medicine, Surgery, Obstetrics & Gynaecology, and Paediatrics. Table 5 shows that the students rotating at the Department of Paediatrics rated their learning climate higher than the others in most subscales. The 'health and stress' subscale in 3 departments received unsatisfactorily low scores (< 5.0). Furthermore, there was the problem of inadequate educational resources in all departments. Learning experiences at the Department of Medicine and at the Department of Obstetrics and Gynaecology also needed improvement.

***Table 4: The clinical learning climate of the four hospitals***

| Subscale and total scores | Mean (SD) | | | |
|---|---|---|---|---|
| | CU | BAH | CBH | PCH |
| Teachers | 8.0 (1.3) | 7.5 (1.2) | 8.2 (1.1) | 7.6 (1.2) |
| Colleagues | 7.6 (2.1) | 7.9 (1.9) | 8.0 (1.6) | 7.6 (1.4) |
| Learning experience | 6.3 (1.5) | 6.7 (1.1) | 7.1 (1.3) | 7.0 (1.3) |
| Ward environment | 6.9 (1.6) | 6.4 (1.5) | 6.2 (1.5) | 6.2 (1.3) |
| Assessment | 6.9 (2.4) | 6.5 (2.2) | 7.4 (1.8) | 6.9 (1.4) |
| Educational resources | 5.8 (2.2) | 4.1 (2.0) | 5.1 (1.7) | 6.4 (1.4) |
| Physical environment | 7.2 (1.7) | 7.2 (1.8) | 6.9 (1.7) | 7.5 (1.7) |
| Motivation | 7.4 (1.8) | 7.9 (0.9) | 7.7 (1.4) | 7.5 (1.5) |
| Health and stress | 5.6 (2.3) | 4.9 (1.5) | 6.0 (1.6) | 6.3 (1.7) |
| Institutional environment | 8.1 (1.4) | 7.2 (1.9) | 8.6 (1.1) | 7.9 (1.5) |
| Total score | 71.3 (10.6) | 67.4 (10.5) | 72.7 (9.6) | 71.9 (8.8) |

***Table 5: The clinical learning climate of the four major departments Chulalongkorn Medical School***

| Subscale and total scores | Mean (SD) | | | |
|---|---|---|---|---|
| | Pediatrics | Surgery | Ob/Gyn | Medicine |
| Teachers | 8.3 (1.2) | 8.4 (1.2) | 7.9 (1.3) | 7.9 (1.3) |
| Colleagues | 7.3 (2.4) | 8.1 (1.7) | 6.8 (2.0) | 7.8 (2.1) |
| Learning experience | 6.8 (1.4) | 6.4 (1.3) | 5.6 (1.6) | 5.9 (1.5) |
| Ward environment | 6.7 (1.7) | 6.8 (1.5) | 7.4 (1.5) | 6.4 (1.6) |
| Assessment | 7.8 (2.1) | 6.6 (2.0) | 6.8 (2.6) | 6.2 (2.4) |
| Educational resources | 6.3 (1.7) | 5.5 (2.1) | 5.7 (2.4) | 5.6 (2.2) |
| Physical environment | 7.5 (1.5) | 6.8 (1.9) | 7.1 (1.7) | 7.4 (1.8) |
| Motivation | 7.6 (1.6) | 7.4 (1.8) | 6.9 (2.0) | 7.2 (1.9) |
| Health and stress | 6.6 (1.8) | 4.9 (2.1) | 4.8 (2.3) | 4.7 (2.3) |
| Institutional environment | 8.0 (1.2) | 7.8 (1.5) | 7.8 (1.5) | 8.3 (1.5) |
| Total score | 74.0 (9.6) | 70.7 (9.4) | 68.3 (10.7) | 69.2 (10.4) |

It might be generalised that the strengths of Chulalongkorn Medical School's clinical learning environment were the teachers and the institutional environment since these two subscales were rated very high in all four departments. The additional highlight of the Department of Surgery was their residents and externs as the 'colleagues' subscale was rated 8.1 – higher than the other three departments.

## Discussion

### Psychometric properties

According to the classical test theory, the two psychometric properties of any instruments are validity and reliability. There are a variety of opinions on the issue of how high the reliability coefficient must be. Nunnally (1978) recommends the reliability of 0.70 as the minimum for early stages of research, 0.80 as the satisfactory level for basic research, and 0.95 as the desirable standard for the settings where very important decisions are to be made. Similarly, Downing (2004) suggests the minimum reliabilities of 0.70 for minor assessment, 0.80 for major assessment, and 0.90 for very high stake assessment while Hayes (2000) addresses that 'anything less than 0.8 would not really be considered reliable' for a typical psychometric test.

Based on the recommendations above, the reliability of all scores obtained from field-testing the instrument was within an acceptable range. Only the alpha of the scores obtained from clinical students rotating at one department (0.67) did not meet the minimum reliability for basic assessment, which might be caused by a discrepancy of opinions within a small number of cases (n = 9), rather than the property of the inventory.

However, a high alpha does not always guarantee *'unidimensionality'* of a test since a test with 20 items or more will have an acceptable coefficient alpha, although it may comprise two or three *'orthogonal'* constructs (Streiner, 2003a). As the development of the inventory was based on the nine domains, it was possible that it was a multi-dimensional inventory and high Cronbach's alphas might be resulted from the test length (49 items) with partial inter-item correlations across dimensions. Therefore, further exploration of the construct of each inventory was performed as described above.

The American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999) jointly defines validity as *'the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests'*. They state that there are five types of evidences, from which the validity of test-score interpretation can be obtained, which are: (1) test content, (2) response processes, (3) relationship to other variables, (4) internal structure and (5) consequences of testing.

#### (1) Test Content

The content-related evidence can be established by demonstrating that (1) the test content relates to the construct it claims to measure and (2) the test content is well sampled by the test items. In this study, the 'learning climate' construct was not drawn from any particular theories. It was a concept explained to all participants before they were asked to generate the item pools. Adopting the Delphi's iterative process, the items were then selected from the pools based on the criteria described above. This study, thus, can claim that the content-related evidence for the validity is well-established.

Another issue relating to the content-related validity evidence is the correspondence between the purpose and the test content of the inventory (Shea & Fortna, 2002). Many learning climate measures were invented so as to differentiate between medical schools, for example, the College Characteristics Index

(Pace & Stern, 1958) and the Medical Schools Environment Questionnaire (Wakeford, 1981). Thus, only the items which serve this purpose were included in the inventories. However, the aim of the instrument developed in this study is to provide a comprehensive view of the learning climate of the medical school. Hence, the ability of the items to distinguish the climate between each medical school was not included in the study's instrument development criteria.

**(2) Response Process**

The validity evidence from response processes was inapplicable for this study since the learning climate measures were intended to assess the respondents' perceptions or feelings towards the elements of the environment. It did not require *'particular cognitive or evaluative process'* to complete these uncomplicated inventories.

**(3) Relationship to Other Variables**

The inventory developed in this study was intended to assess the elements of the learning environment, some of which had never been measured before by other climate inventories. Since the gold standard of the learning climate measures does not exist, the 'evidence from relationship to other variables' could not be obtained. Van der Vleuten (1996) critically addressing the issue of using correlational research to establish validity without gold standards states that:

> *'Validity research in educational testing contains a plethora of correlational studies, replete with mid-range correlations, which are more like Rorschach tests for the creative researcher to interpret favourably regardless of the outcome (glasses are always half empty or half full anyway)'.*

**(4) Internal Structure**

*'Analyses of the internal structure of a test can indicate the degree to which the relationships among test items and test components conform to the construction which the proposed test score interpretations are based'* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Nevertheless, it is essential to identify if the inventory is intentionally developed as *'a scale'* or *'an index'*.

Streiner (2003b) defines *'a scale'* as a collection of items or questions, which *'are composed of theoretically correlated items'* and defines *'an index'* as an inventory, which comprises unrelated items. The question is 'should a learning climate measure be developed as a scale or an index?'

Considering the learning climate measure as 'a scale', during the developing process, six items were removed due to the low item-total correlations. The removal of these items was corresponding to the psychometric theories since it increased item statistics and a scale does not *'require each item to be intrinsically meaningful, as long as the total scores … are meaningful'* (Schuwirth and van der Vleuten, 2004). Therefore, the refinement of the inventory after the Stage 3 data collection has provided the internal structure evidence for validity of the climate measure.

However, Streiner (2003b) explains that *'the differentiation of questionnaires into scales and indexes represent the two ends of a continuum. In reality, there are many that fall somewhere in between where it is difficult to determine ...'*. If the learning climate measure developed in this study was expected to be in the middle of the scale-index continuum, the elimination of the items would probably have decreased the reproducibility and the validity of the test *'although the numbers will indicate otherwise'* (Schuwirth and van der Vleuten, 2004). It was noticed that all negative statements were removed during the process. The removed items might be the crucial aspects of the learning environment as Streiner (2003a) describes that:

> *'When developing an index, the choice of the specific items is much more important than is the case in the construction of scales. Because the items may be uncorrelated, it cannot be assumed that what is missed if one item is omitted will be covered by the others that remain'.*

There is no absolute true or false answer for this discrepancy as it depends on how the hypothetical construct of the learning climate is – which has yet to be universally agreed. At least, there are two versions of the learning climate measure for the clinical phase of undergraduate medical education in Thailand available for use.

The utilization of factor analysis to explore the construct of the refined inventory confirmed that the learning climate measure was 'multi-dimensional' indeed. The combination of qualitative and statistical factor analyses resulted in the meaningful subscales with the alpha coefficients ranging from 0.62 to 0.87. Applying the concept of the scale-index continuum to the sub-construct-level of the instrument, there is, again, no definite answer of whether the subscales with low internal consistencies need further modification.

**(5) Consequences of Testing**

As the learning climate measure has just been developed and has not been widely implemented, the validity evidence regarding the consequences of test cannot be established yet.

**The utility of the clinical learning climate measure**

Learning climate is students' perceptions towards the elements of learning environment. If a learning climate measure could tap all essential elements of learning environment, it will assess every aspect of a curriculum which is perceived important by students. Hence, a learning climate measure might be used as a single screening test to replace numerous curriculum/ course/classroom evaluation forms and to help prevent the 'questionnairophobia' syndrome. However, there are 2 limitations regarding the use of the learning climate measure: (1) the instrument is nation-specific and (2) only perception can be measured.

**(1) Nation-Specific Issue**

Schwarz & Wojtczak (2002) have introduced the concept of the global minimum essential requirements but they emphasize that the local and the national needs *'must also be taken into account'*. Similarly, the components of learning environment may be divided into the international, the nation-specific and the school-specific elements (Figure 3).

Some elements of the learning environment may be generic to every country in the world while some elements may be crucial to a few countries, for instance, a country with multiracial society may have the issue of racism. The third group of the learning environment is the school-specific elements. This may relate, for example to the private medical school environment or the military medical school environment.

The learning climate measure developed in this study can be utilized to measure the international and the nation-specific elements. Hence, it will not be valid to be used to compare medical schools across countries. However, the instrument, which is non-culturally specific such as the Dundee Ready Education Environment Measure (Roff et al, 1997), will not be able to assess the nation-specific and the school-specific elements of the learning environment. Nevertheless, there has not been any evidence suggesting the proportion of each component yet. If the proportion is similar to the Model A (Figure 4), it will be essential to develop neither the non-culture specific nor the nation-specific, but the school-specific instrument. If the proportion is like the Model C, then there will be no much need for the nation-specific instrument.

***Figure 3. The model illustrating the international, the nation-specific and the school-specific elements of the learning environment***

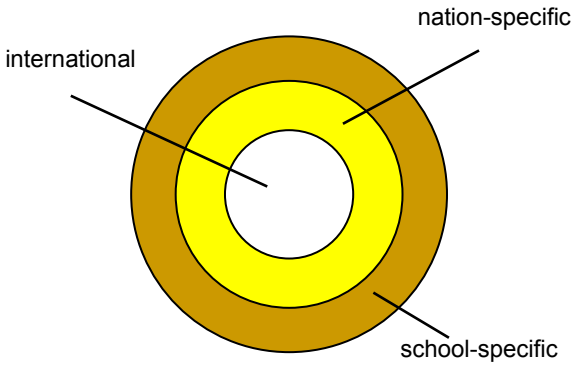
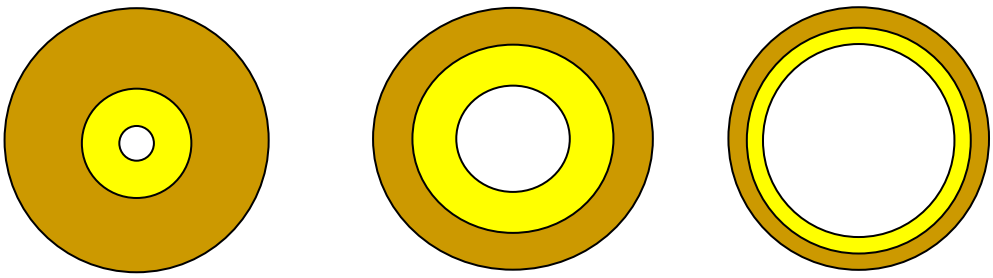



***Figure 4. The model illustrating three possible proportions of three elements of the learning climate***



**(2) Measuring Students' Perception**

Kirkpatrick (1967) describes four levels of evaluation, in which the complexity of the behavioural change increased as evaluation strategies ascended to each higher level. The four levels are:

- First level – evaluation of reaction (satisfaction or happiness)
- Second level – evaluation of learning (knowledge or skills acquired)
- Third level – evaluation of behaviour (transfer of learning to workplace)
- Fourth level – evaluation of results (transfer of impact on society)

Learning climate measures are the instruments intended to assess the first (and the lowest) hierarchy of levels of evaluation. However, it is suggested that evaluators should consider the third or the fourth levels of this hierarchy for their evaluation since they would provide more accountable impact. As a result, assessors should be aware that measuring learning climate alone is not enough. They should consider evaluating the higher levels of Kirkpatrick's model as well.

## Conclusion

This research aims to develop the instrument to measure the clinical learning climate of Thai undergraduate medical education. A 43-item climate measure with validity and reliability evidence was developed. Its ten subscales are: teachers, colleagues, learning experience, ward environment, evaluation, educational resources, physical environment, motivation, health and stress, and institutional environment.

## References


American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999) *Standards for educational and psychological testing.* Washington: American Educational Research Association.

Downing, S.M. (2004) Reliability: on the reproducibility of assessment data, *Medical Education*, 38(9), pp. 1006-12.

Hayes, N. (2000) Doing psychological research: gathering and analysing data. Buckingham: Open University Press.

Israel, G.D. (1992) Fact sheet PEOD-6: determining sample size. Gainesville: Institute of Food and Agricultural Sciences, University of Florida.

Kirkpatrick, D.L. (1967) Evaluation of training. In: R. Craig & I. Bittel (Eds) Training and development handbook, New York: McGraw-Hill.

Norman, G.R. & Streiner, D.L. (1997) PDQ statistics. St. Louis: Mosby.

Nunnally, J.C. (1978) *Psychometric theory*. New York: McGraw-Hill.

Pace, C.R. & Stern, G.G. (1958) An approach to the measurement of psychological characteristics of college environments. *Journal of Educational Psychology,* 49(5), pp. 269-77.

Pololi, L. & Price, J. (2000) Validation and use of an instrument to measure the learning environment as perceived by medical students, *Teaching and Learning in Medicine*, 12(4), pp. 201-7.

Roff, S., McAleer, S., Harden, R.M., Al-Qahtani, M., Ahmed, A.U., Deza, H., Groenen, G. & Primparyon, P. (1997) Development and validation of the Dundee Ready Education Environment Measure (DREEM) *Medical Teacher*, 19(4), pp. 295-9.

Schuwirth, L. & van der Vleuten, C. (2004) Merging views on assessment, *Medical Education*, 38(12), pp. 1208-11.

Schwarz, M.R. & Wojtczak, A. (2002) Global minimum essential requirements: a road towards competence-oriented medical education, *Medical Teacher*, 24(2), pp. 125-9.

Shea, J.A. & Fortna, G.S. (2002) Psychometric methods, In: G.R. Norman, C.P.M. Van der Vleuten & D.I. Newble (Eds) International handbook of research in medical education, Dordrecht: Kluwer Academic Publishers.

Streiner, D.L. (2003a) Starting at the beginning: an introduction to coefficient alpha and internal consistency, *Journal of Personality Assessment*, 80(1), pp. 99-103.

Streiner, D.L. (2003b) Being inconsistent about consistency: when coefficient alpha does and doesn't matter, *Journal of Personality Assessment,* 80(3), pp. 217-22.

Van der Vleuten, C.P.M. (1996) The evaluation of professional competence: developments, research and practical implications, *Advances in Health Sciences Education*, 1(1), pp. 41-67.

Wakeford, R. (1981) Students' perception of the medical school learning environment: a pilot study into some differences and similarities between clinical schools in the UK, *Assessment and Evaluation in Higher Education*, 6(3), pp. 206-17.