

## Research Article

# Stability of DREEM in a Sample of Medical Students: A Prospective Study

**Muhamad Saiful Bahri Yusoff**

*Medical Education Department, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kota Bharu, Kelantan, Malaysia*

Correspondence should be addressed to Muhamad Saiful Bahri Yusoff, msaiful@kb.usm.my

Received 14 March 2012; Revised 28 June 2012; Accepted 5 July 2012

Academic Editor: Terry Stratton

Copyright © 2012 Muhamad Saiful Bahri Yusoff. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Background.* Over the last 15 year, DREEM was applied in various educational settings to appraise educational climate. So far, none of article reported its stability in Malaysian medical students. *Objective.* To determine stability of the DREEM to measure educational climate at different time and occasions using a sample of medical students. *Methodology.* A prospective cohort study was done on 196 first year medical students. It was administered to the medical students at four different intervals. The Cronbach's alpha and intraclass correlation analysis were applied to measure internal consistency and agreement level across the intervals. The analysis was done using SPSS 18. *Result.* A total of 186 (94.9%) medical students responded completely to the DREEM inventory. The overall Cronbach's alpha value of the DREEM at the four measurements ranged between 0.91 and 0.94. The average Cronbach's alpha values of the five subscales ranged between 0.45 and 0.83. The ICC coefficient values for the DREEM total score was 0.67 and its subscales ranged between 0.51 and 0.62. *Conclusion.* This study supported satisfactory levels of stability and internal consistency of the DREEM to measure educational climate over multiple observations in a sample of Malaysian medical students. Continued research is required to optimise its psychometric credential across educational settings.

## 1. Introduction

It is widely agreed among medical educators that optimal educational climate is an important factor for effective learning to occur [1, 2]. Indeed, evaluation of the educational climate has been highlighted as a key to the delivery of a high-quality medical education [1, 2]. Therefore, to conduct such evaluation, a valid and reliable tool is vital.

Over the past 15 years, medical and allied health educators across places and educational settings have widely used the Dundee Ready Educational Environment Measure (DREEM) to appraise their institutions' educational climate [3–9]. This valuable tool was originally designed in English [8] and has been translated into various languages such as Swedish, Greek, and Spanish [10–12]. These facts have proven that DREEM is internationally accepted as a useful tool to provide feedback on strengths and weaknesses of educational climate at particular educational institution. One of important implications of the DREEM is that it provides a standardized way for international comparisons between

medical schools as well as allowing them to benchmark their educational climate [13]. In addition, it may locate areas of concern shared by majority of students that might be unintentionally neglected by educators.

Validity is broadly described as the ability of a measurement to measure attributes that it intended to measure [14]. The initial psychometric evaluation that was carried out by its developer showed that DREEM is a valid tool to measure educational environment. However, four previous studies [11–13, 15] reported that confirmatory factor analysis did not support the five-factor structure claimed by the DREEM developer. These studies concluded that the construct validity of DREEM was not well supported. These psychometric shortcomings do invite further inspection on validity aspects of DREEM [13]. To author's knowledge, one article reported on validity of the DREEM in a sample of Malaysia medical students [15].

Reliability is broadly described as the consistency or reproducibility of a measurement over time and occasions and it can be gauged in the form of internal consistency and

stability [14]. The internal consistency of a tool is commonly measured based on single administration while the stability of a tool is measured based on multiple administrations on different occasions or time [14]. The internal consistency is measured by various ways such as Cronbach's alpha, Kuder-Richardson, and split halves [14]. Stability is measured by the degree of agreement between observations based on multiple administrations in the form of interrater reliability, intrarater reliability, and test-retest reliability [14]. The degree of agreement between multiple observations can be gauged as correlation coefficients such as intraclass correlation coefficient (ICC) and kappa Cohen coefficient [16–18]. The DREEM has been reported to have a high level of internal consistency as the overall Cronbach's alpha coefficient more than 0.7 [8, 10–13, 19]. It was also found to have a high level of stability as the test-retest correlation coefficient more than 0.8 [11]. In addition, an article reported that its reliability coefficients in a sample of Malaysian medical students across years of study ranged between 0.53 and 0.82 [15]. As far as the author's concerns, none of article reported on its stability in Malaysian medical students. Therefore, this study was done to fill in the gap as well as to enrich evidence related to its credential in measuring educational climate in Malaysian settings. Apart from that, the author believes that the DREEM will be a very helpful tool to benchmark the educational climate in Malaysian educational settings since the results obtained can be compared with educational climates in other institutions outside Malaysia. Through this strategy, the Malaysian educational climate can be improved to a better level. From that notion, investigating its stability through multiple administrations on a cohort of medical students would be a worthwhile effort to provide evidence to support its psychometric credential in Malaysian contexts.

This study was designed to answer three questions: (1) what is the internal consistency of the DREEM over multiple administrations?; (2) what is the internal consistency of the DREEM subscales over multiple administrations?; (3) what is the degree of agreement between measurements of the DREEM subscales over multiple administrations? The author hypothesized that the DREEM would demonstrate good level of stability and internal consistency in measuring students' perception on educational environment over multiple administrations. This study will provide stability evidence of the DREEM to measure educational environment across time and occasions.

## 2. Methodology

A prospective study was conducted on first year medical students in a Malaysian public medical school. Purposive sampling method was applied and a total of 196 medical students were selected. They were then followed up at four intervals.

**2.1. Collection of Data.** The DREEM was administered at four intervals; 2 months (time 1), 4 months (time 2), 6 months (time 3), and 8 months (time 4) of the first year medical training. Informed consent was obtained from the

respondents and their matrix number was used as their identity code. They were asked to response to all statements completely and voluntarily. Data was collected by guided self-administered questionnaire during face-to-face sessions in a hall. The questionnaires were immediately returned after they completely filled in. Data was analysed by Statistical Package for Social Sciences (SPSS) version 18.

**2.2. The Dundee Ready Education Environment Measurement (DREEM).** The DREEM inventory was developed as a tool to measure educational climate at educational institutions [8, 20] and was claimed as a "cultural-free" instrument [21]. There are 50 items measuring five aspects of educational environment based on students' perception which include students' perception of learning (SPoL), students' perception of teaching (SPoT), students' academic self-perception (SASP), students' perception of atmosphere (SPoA), and students' social self-perception (SSSP). Each item is rated based on five Likert-scales range between 0 and 4 (0 = strongly disagree, 1 = disagree, 2 = unsure, 3 = agree, and 4 = strongly agree). There are 9 negative items that must be scored in a reverse manner prior to analysis and interpretation; item 4, 8, 9, 17, 25, 35, 39, 48, and 50 [8]. It has been translated in various languages and the reported overall Cronbach's alpha coefficient ranges between 0.89 and 0.93 [10–13, 19]. The original version of DREEM was used in this study.

**2.3. Stability Analysis.** Reliability analysis was applied to determine the internal consistency of the DREEM. Internal consistency of its items was measured using Cronbach's alpha coefficient. The items were considered to represent an acceptable level of internal consistency if the Cronbach's alpha value within 0.5 to 0.7 and good level if the Cronbach's alpha value more than 0.7 [14, 16, 22].

Intraclass correlation (ICC) analysis was done to determine the level of agreement between measurements at four different intervals. The agreement level was represented as ICC coefficient. The ICC coefficient value less than 0.2 was considered as poor agreement, 0.21 to 0.40 was considered as fair agreement, 0.41 to 0.60 was considered as moderate agreement, 0.61 to 0.80 was considered as good agreement, and 0.81 to 1.0 was considered as very good agreement [14, 16, 17].

## 3. Result

A total of 186 (94.9%) medical students responded completely to the DREEM inventory. Majority of the respondents were female (65.1%), Malay (55.9%), and came from the matriculation stream (88.2%) as shown in the Table 1.

Reliability analysis (Table 2) showed that the overall Cronbach's alpha value of the DREEM at the four measurements ranged between 0.91 and 0.94, indicating good level of internal consistency at different time and occasions. The Cronbach's alpha value for the SPoL subscale ranged between 0.81 and 0.84, indicating good level of internal consistency over four different measurements. The Cronbach's alpha

TABLE 1: Demographic profile of participants.

Variable		Frequency (%), (N = 186)
Sex	Male	65 (34.9)
	Female	121 (65.1)
Race	Malay	104 (55.9)
	Chinese	57 (30.6)
	Indian	20 (10.8)
	Others	5 (2.2)
Entry qualification	Matriculation	164 (88.2)
	HSC	13 (7.0)
	A-level	9 (4.8)

value for the SPoT subscale ranged between 0.65 and 0.75, indicating acceptable to good level of internal consistency across the four intervals. The Cronbach's alpha value for the SSSP subscale ranged between 0.36 and 0.51, indicating poor to acceptable level of internal consistency across multiple measurements. The Cronbach's alpha value for the SASP subscale ranged between 0.76 and 0.80, indicating good level of internal consistency levels across time and occasions. The Cronbach's alpha value for the SPoA subscale ranged between 0.79 and 0.85, indicating good level of internal consistency across measurements. The SPoT and SPoA subscales demonstrated the highest average Cronbach's alpha values and the SSSP demonstrated the lowest average Cronbach's alpha value, while the other subscales were in between them. This study highlighted that the SSSP had the poorest internal consistency that might need some attentions.

ICC analysis (Table 2) showed that ICC coefficient values for the five subscales of DREEM ranged from 0.51 and 0.62, indicating acceptable to good level of agreement across the four measurements. The ICC coefficient value for the overall score was 0.61, indicating good level of agreement across the four measurements. This finding provided evidence to support the stability of DREEM at different time and occasions.

#### 4. Discussion

Our data found that overall Cronbach's alpha of the DREEM demonstrated a high level of internal consistency across multiple administrations that were done at different times and occasions as the Cronbach's alpha value more than 0.7 [14, 16, 22]. This indicated that the DREEM had good level of internal stability to reproduce similar results on similar cohort of studied population at different times and occasions. On top of that, our finding was comparable with previous studies that reported that its overall Cronbach's alpha value was more than 0.7 [8, 10–13, 19]. In addition, ICC analysis showed that the overall ICC coefficient was more than 0.6, indicating good level of agreement between measurements that were done at different time and occasions [14, 16, 17]. In other word, the ICC analysis showed that it had good stability across time and occasions. These findings demonstrated that the DREEM was a stable tool to measure educational climate across multiple measurements. This finding supported a previous study reported that the DREEM

had a high level of stability based on test-retest reliability [11].

Our data also demonstrated that the DREEM subscales had good level of internal consistency across multiple administrations as their average Cronbach's alpha values were more than 0.7 except for SSSP subscale which was 0.45 (i.e., ranging between 0.36 and 0.51 as shown in Table 2). This finding suggested that they had a stable internal consistency across occasions and time that reflected reproducibility of a measurement over time and occasions [14, 16]. In addition, this finding is in keeping with a previous study found that the Cronbach's alpha values of the subscales ranged from 0.53 to 0.83 [15]. Apart from that, it appeared the SSSP subscale had average Cronbach's alpha value lower than 0.5, indicating poor internal consistency. Perhaps, this finding suggesting that this domain should be relooked on how its reliability can be improved in future. Previous studies reported similar finding where the SSSP subscale had the lowest Cronbach's alpha values which were 0.40 [11] and 0.55 [13]. One possible reason could be due to the cross-cultural variability of the internal consistency of the DREEM. Another reason could be due to items in the SSSP subscale were not correlated with each other suggesting that they perhaps could contribute better to internal consistency of other constructs. After looking back at each item in the SSSP construct, the author believes that the items were not culturally bias because all the statements were constructed using simple, nonsensitive, and comprehensible English sentences. The author found that the low internal consistency is due to a few items that might poorly represent the SSSP construct such as *"there is a good support system for students who get stressed"* that might be more relevant to the SPoA construct, and *"I am too tired to enjoy the course"* and *"I am rarely bored in this course"* that might be more relevant to the SPoL construct as was also concerned by Hammond et al. [13]. Considering these explanations, the author postulated that the low internal consistency of the SSSP construct is due to a few items that were poorly representing the SSSP construct rather than the cultural variability or culturally bias as was stated by Roff and McAleer (1991) as the DREEM is a tool that "cultural-free". Nevertheless, these findings provided evidence to support satisfactory stability of internal consistency of the subscales to measure educational environment in Malaysia medical student population.

On further analysis, the subscales demonstrated moderate to good level of agreement between measurements done at different times and occasions as the ICC coefficient values ranged from 0.51 and 0.62 [14, 16, 17]. This reflected satisfactory degree of agreement between the subscales' measurements over multiple administrations at different time and occasions. In other word, they were able to produce similar results for similar individual over time and occasions. One implication of this finding was the tendency for respondents to cheat or fake their responses to the items representing the DREEM subscales was low. Rationally, if respondents were faking their responses, the degree of agreement between measurements at different time and occasions will be very poor, but our data showed satisfactory level of agreement across multiple administrations. These

TABLE 2: Internal consistency and ICC values across the four measurements.

DREEM	Cronbach's alpha value					ICC value <sup>a</sup>
	Time 1	Time 2	Time 3	Time 4	Average	
Overall	0.91	0.92	0.93	0.94	0.93	0.67**
Students' Perception of Learning (SPoL) (12 items)	0.81	0.83	0.84	0.82	0.83	0.61**
Students' Perception of Teachers (SPoT) (11 items)	0.65	0.75	0.72	0.73	0.71	0.60**
Students' Social Self-Perception (SSSP) (7 items)	0.51	0.36	0.45	0.47	0.45	0.51**
Students' Academic Self-Perception (SASP) (8 items)	0.76	0.80	0.76	0.78	0.78	0.62**
Students' Perception of Atmosphere (SPoA) (12 items)	0.79	0.79	0.83	0.85	0.82	0.61**

<sup>a</sup> ICC analysis (single measure) between Time 1, Time 2, Time 3, and Time 4.

\*\* $P < 0.001$ .

Average of Cronbach's alpha was referred to mean Cronbach's alpha of the four measurements.

findings clearly demonstrated that the subscales have satisfactory level of stability to measure educational environment over different times and occasions. This finding support a previous study reported that the DREEM was a stable tool to measure educational environment [11].

The reliability analysis has provided evidence of stability and internal consistency of the DREEM in measuring education environment in a sample of Malaysian medical students. Despite of these encouraging findings, this study has several limitations that should be considered for future research as well as for interpretation. The first, this study confined to first year medical students at a medical school, thus, interpretation of the study findings should be made with caution because the findings might differ if was data collected from other years of study. One possible reason could be due to the different level of medical students' maturity as a result of the different phases of medical training that might lead to different perception on the educational climates. Perhaps, a validation study should be conducted in a multicentre, cross-cultural, and cross-phases of medical training setting to verify its stability. The second, this study selected study subjects based on purposive sampling method therefore sampling bias might compromise the accuracy of the reported results. Therefore, random sampling method should be applied in future research to minimise the sampling bias. Considering these limitations, interpretation and any attempt to generalise the result should be done with caution. Continued research is required to optimise psychometric credential of the DREEM across educational settings.

## 5. Conclusion

This study supported a satisfactory level of stability and a high level of internal consistency of the DREEM to measure educational climate over multiple observations. Continued research is required to optimise psychometric credential of the DREEM across educational settings.

## Acknowledgments

The author's special thanks to the School of Medical Sciences, Universiti Sains Malaysia for supporting and allowing him to

undertake this study. Thanks to Human Ethical Committee of Universiti Sains Malaysia for the permission and clearance to conduct this study. The author's appreciation to all first year medical students involved in this study. The author's special thanks also to the support staff of the Academic Office and Medical Education Department staff for their help. This study was made possible under the Short Term Research Grant no. 304/PPSP/6139071.

## References

- [1] J. A. Dent and R. M. Harden, *A Practical Guide for Medical Teachers*, Elsevier Churchill Livingstone, 2009.
- [2] D. Newble, R. A. Cannon, and Z. A. Kapelis, *A Handbook for Medical Teachers*, Kluwer Academic, 2001.
- [3] H. Arzuman, M. S. B. Yusoff, and S. P. Chit, "Big Sib students' perceptions of the educational environment at the school of medical sciences, universiti sains Malaysia, using Dundee Ready Education Environment Measure (DREEM) inventory," *Malaysian Journal of Medical Sciences*, vol. 17, no. 3, pp. 40–47, 2010.
- [4] A. Al-Hazimi, A. Al-Hyiani, and S. Roff, "Perceptions of the educational environment of the medical school in King Abdul Aziz University, Saudi Arabia," *Medical Teacher*, vol. 26, no. 6, pp. 570–573, 2004.
- [5] A. Al-Hazimi, R. Zaini, A. Al-Hyiani et al., "Educational environment in traditional and innovative medical schools: a study in four undergraduate medical schools," *Education for Health*, vol. 17, no. 2, pp. 192–203, 2004.
- [6] N. M. Mohd Said, J. Rogayah, and A. Hafizah, "A study of learning environments in the Kuliyah (Faculty) of Nursing, International Islamic University Malaysia," *Malaysian Journal of Medical Sciences*, vol. 16, no. 4, pp. 15–24, 2009.
- [7] B. S. Thomas, R. R. Abraham, M. Alexander, and K. Ramnarayan, "Students' perceptions regarding educational environment in an Indian dental school," *Medical Teacher*, vol. 31, no. 5, pp. e185–e188, 2009.
- [8] S. Roff, S. McAleer, R. M. Harden et al., "Development and validation of the Dundee Ready Education Environment Measure (DREEM)," *Medical Teacher*, vol. 19, no. 4, pp. 295–299, 1997.
- [9] R. Varma, E. Tiyagi, and J. K. Gupta, "Determining the quality of educational climate across multiple undergraduate teaching sites using the DREEM inventory," *BMC Medical Education*, vol. 5, no. 1, p. 8, 2005.



- [10] A. Riquelme, M. Oporto, J. Oporto et al., "Measuring students' perceptions of the educational climate of the new curriculum at the Pontificia Universidad Católica de Chile: performance of the Spanish translation of the Dundee Ready Education Environment Measure (DREEM)," *Education for Health*, vol. 22, no. 1, p. 112, 2009.
- [11] I. D. Dimoliatis, E. Vasilaki, P. Anastassopoulos, J. P. Ioannidis, and S. Roff, "Validation of the Greek translation of the Dundee Ready Education Environment Measure (DREEM)," *Education for Health*, vol. 23, no. 1, p. 348, 2010.
- [12] U. Jakobsson, N. Danielsen, and G. Edgren, "Psychometric evaluation of the Dundee Ready Educational Environment Measure: Swedish version," *Medical Teacher*, vol. 33, no. 5, pp. e267–e274, 2011.
- [13] S. M. Hammond, M. O'Rourke, M. Kelly, D. Bennett, and S. O'Flynn, "A psychometric appraisal of the DREEM," *BMC Medical Education*, vol. 12, no. 1, p. 2, 2012.
- [14] L. D. Streiner and G. R. Norman, *Health Measurement Scales: A Practical Guide to Their Development and Use*, Oxford University Press, New York, NY, USA, 2008.
- [15] M. S. B. Yusoff, "Psychometric properties of dreem in a sample of malaysian medical students," *Medical Teacher*, vol. 34, no. 7, pp. 595–596, 2012.
- [16] D. G. Altman, *Practical Statistics For Medical Research*, Chapman & Hall, London, UK, 1991.
- [17] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
- [18] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, no. 2, pp. 420–428, 1979.
- [19] J. S. Khan, S. Tabasum, U. K. Yousafzai, and M. Fatima, "DREEM on: validation of the dundee ready education environment measure in Pakistan," *Journal of the Pakistan Medical Association*, vol. 61, no. 9, pp. 885–888, 2011.
- [20] S. McAleer and S. Roff, "A practical guide to using the Dundee Ready Education Environment Measure (DREEM)," *AMEE Medical Education Guide*, no. 23, pp. 29–33, 2001.
- [21] S. Roff and S. McAleer, "What is educational climate?" *Medical Teacher*, vol. 23, no. 4, pp. 333–334, 2001.
- [22] J. C. Nunally, *Psychometric Theory*, McGraw-Hill, New York, NY, USA, 1978.

