TEACHING EFFECTIVENESS OF INSTRUCTORS DURING LECTURES IN PRE-CLINICAL PHASE, SCHOOL OF MEDICAL SCIENCES, USM

MUHD AL-AARIFIN ISMAIL

Dissertation Submitted in Partial Fulfillment of the Requirements of the Master of Science (Medical Education)

UNIVERSITI SAINS MALAYSIA

2016

DECLARATION

This is to certify to the best of my knowledge, this dissertation is entirely the work of the candidate, Muhammad Al-Aarifin Ismail.

_____

DR MUHAMAD SAIFUL BAHRI BIN YUSOFF

Main Supervisor

Medical Education Department

School of Medical Sciences

Universiti Sains Malaysia

ACKNOWLEDGEMENT

Praise to Allah s.w.t, The Almighty God, the most compassionate and most merciful, whose blessing has helped me through the entire completion of this dissertation. I would like to express my deepest gratitude and thanks to all the following individuals who have helped me during the preparation and the completion of this dissertation for my Master in Medical Education in School of Medical Sciences, Universiti Sains Malaysia.

1. My parents, wife, children and all my family members for their great support, motivation, pray, encouragement from the beginning to the end of my study.
2. My university (Universiti Sains Malaysia) for all support including financial grant for me to pursue this master program.
3. My sincere appreciation and gratitude to my main supervisor, Dr Muhamad Saiful Bahri Yusoff from Medical Education Department, School of Medical Sciences, Universiti Sains Malaysia for his continuous support of my study. His guidance had truly helped me in all the times throughout my research and the writing on this thesis.
4. My co-supervisors, Dr Nik Mohd Rizal Mohd Fakri & Dr Mohd Zarawi Mat Nor for their concern, help, and support throughout the process of completing this study.
5. My dean and deputy dean of School of Medical Sciences, USM for their permission and full co-operation for me to conduct this study.
6. All lecturers, staff and colleagues in Medical Education Department for their kind support.
7. Last but not least to all medical students who participated in this study as student evaluators.

TABLE OF CONTENT

LIST OF TABLE

LIST OF FIGURES

ABBREVIATIONS

USM	Universiti Sains Malaysia
SMS	School of Medical Sciences
TBI	Teacher Behavior Inventory
SPSS	Statistical Package for Social Sciences
MQA	Malaysian Qualification Agency
SEEQ	Students' Evaluations of Educational Quality
MD	Doctor of Medicine
Assoc. Professor	Associate Professor
SPICES	Student-centered, Problem-based learning, Integrated, Community-based, Elective and Systematic learning

KEBERKESANAN PENGAJARAN BAGI PENSYARAH-PENSYARAH SEWAKTU KULIAH DI FASA PRA-KLINIKAL, PUSAT PENGAJIAN SAINS PERUBATAN, USM

ABSTRAK

PENGENALAN

Pengajaran adalah salah satu daripada tugas-tugas asas sebagai pensyarah perubatan. Pengajaran yang berkesan akan menhasilkan pembelajaran lebih baik bagi pelajar yang akhirnya dapat mempertingkatkan hasil-hasil pendidikan. Penilaian pengajaran adalah sesuatu yang penting dalam mana-mana kurikulum. Dengan sistem maklum balas yang baik, ia akan menyebabkan kemahiran pengajaran akan meningkat dalam kalangan pensyarah perubatan.

OBJEKTIF

Tujuan kajian ini adalah untuk menilai keberkesanan pengajaran sewaktu kuliah bagi pensyarah dalam fasa pra-klinikal di Pusat Pengajian Sains Perubatan, USM melalui penilaian oleh pelajar dengan menentukan tahap kelakuan-kelakuan mengajar. Ia juga untuk menentukan faktor-faktor (jantina, kumpulan etnik, kepakaran dan pangkat) yang berkait dengan keberkesanan pengajaran oleh pensyarah.

METODOLOGI

Kajian ini menggunakan kaedah kajian keratan rentas. Pensyarah di Pusat Pengajian Perubatan, USM yang terlibat dengan kuliah untuk pelajar-pelajar tahun pra-klinikal diambil sebagai subjek kajian. 30 orang pelajar pra-klinikal telah dilatih untuk

menilai tahap kelakuan mengajar mereka dengan menggunakan alat yang telah disahkan. Kebenaran dari sudut etika telah perolehi daripada jawatan kuasa etika dan pihak pusat pengajian sebelum memulakan kajian. Pengumpulan data telah dilakukan dalam masa empat bulan dan telah dianalisa menggunakan *Statistical Package for Social Sciences* (SPSS) versi 22.

KEPUTUSAN

Sejumlah 55 orang pensyarah perubatan telah terlibat. Mereka mempunyai tahap kelakuan mengajar yang baik dari segi *organisasi* dan *pengucapan & kelajuan* sebagaimana ditunjukkan oleh skor purata masing-masing 4.02 dan 4.15. Untuk aspek *kejelasan, kesungguhan, interaksi* dan *hubungan*, mereka mempunyai tahap kelakuan pengajaran yang boleh diterima sebagaimana yang telah ditunjukkan oleh skor purata dari 3.10 sehingga 3.59. Sementara itu, kelakuan pengajaran yang mendapat penilaian paling rendah adalah *keterbukaan* dimana skor puratanya adalah 2.20. Tidak ada hubungan yang signifikan antara kelakuan pengajaran dan semua factor-faktor yang dilihat (*jantina, kumpulan etnik, kepakaran* dan *pangkat*).

KESIMPULAN

Pensyarah pra-klinikal, Pusat Pengajian Sains Perubatan, USM telah dinilai secara positif dalam aspek *organisasi* dan *pengucapan & kelajuan*. Untuk kej*elasan, kesungguhan, interaksi* dan *hubungan*, kelakuan-kelakuan ini dinilai sebagai aspek di bawah kawasan untuk dipertingkatkan. Manakala, *keterbukaan* pula dinilai sebagai kawasan yang perlukan perhatian. Kelakuan pengajaran tidak mempunyai kaitan dengan *jantina, kumpulan etnik, kepakaran* dan *pangkat.* Walaubagaimanapun, secara relatifnya pensyarah bukan klinikal dinilai sebagai lebih baik dalam beberapa aspek kelakuan pengajaran yang bersifat lebih spesifik.

TEACHING EFFECTIVENESS OF INSTRUCTORS DURING LECTURES IN PRE-CLINICAL PHASE, SCHOOL OF MEDICAL SCIENCES, USM

ABSTRACT

INTRODUCTION

Teaching is one of the major roles as a medical teacher. Effective teaching means better students' learning and ultimately will improve educational outcomes. Teaching evaluation is a vital activity in any curriculum. With proper feedback system, it will lead to improvement of teaching skill among faculty members.

OBJECTIVES

The purpose of this study was to evaluate lecturers' teaching effectiveness during lectures in pre-clinical years USM medical school through medical students' ratings by determining level of teaching behaviors. It was also to determine factors (gender, ethnic groups, specialties and designation) that associate with the lecturers' teaching effectiveness.

METHODOLOGY

This study utilized a cross-sectional design. Faculty members in School of Medical Sciences, USM who has been involved in delivering lecture for pre-clinical year students were chosen as study subjects. 30 pre-clinical year students were trained to rate their teaching behavior by using validated instrument. School and ethical committee clearance were obtained prior to the start of the study. Data collection was

done in 4 months' time and it was analyzed using Statistical Package for Social Sciences (SPSS) version 22.

RESULTS

A total of 55 instructors were involved. They attained good level of teaching behavior in the aspects of *organisation* and *speech & pacing* as indicated by the mean scores of 4.02 and 4.15 respectively. For the *clarity*, *enthusiasm*, *interaction* and *rapport* aspects, they attained acceptable level of teaching behaviors as suggested by mean scores ranging from 3.10 – 3.59. Meanwhile the lowest level was in *disclosure* aspect with mean score of 2.20. There were no significant relationship seen between teaching behaviors and all observed factors (i.e. instructors' gender, ethnic group, specialty and designation),

CONCLUSION

USM medical school pre-clinical instructors were positively perceived in the aspects of *organisation* and *speech & pacing*. The *clarity*, *enthusiasm*, *interaction* and *rapport* aspects were under areas for improvement and the *disclosure* aspect was an area of concern. Teaching behaviors were not associated with gender, ethnic group, area of expertise and designation. However, non-clinical instructors were better perceived than the clinical instructors in several low-inference teaching behaviors.

.

CHAPTER 1

INTRODUCTION

1.1 Title

Teaching Effectiveness of Instructors during Lectures in Pre-Clinical Phase, School of Medical Sciences, USM

1.2 Back ground of the study

All medical lecturers in USM, School of Medical Sciences are considered as experts in their field. To be a faculty member in this school, one must pass their master program and being gazetted by respective accreditation bodies and some of them (particularly for DS) must have PhD. For that reason, their expertise in their field are not questionable. Apart from practicing their knowledge and skills in hospital setting or doing researches, one of their major roles as faculty members is teaching. They are expected to teach their undergraduate as well as post graduate students. However, most of them receive little or no training on effective teaching.

Teaching expertise was previously assumed to be a part of the individuals' content expertise. Whereby, someone who has acquired the knowledge in any disciplines, he or she could inherently be a good teacher in that particular discipline. As the saying goes *"If you understand your discipline, then you should be able to teach it"* and also *"See one, do one, teach one"* (Irby, 1994). However overtime it has been found that teaching skills and content expertise were somehow separate kind of attributes. It is a skill associated with but separate from, content expertise (Wilkerson and Irby,

1998). In other words, being a content expert doesn't necessarily mean one is expert in teaching even in his/her expert area. Pedagogical skills are transferable skills that could be and should be taught and learned.

Malaysian Ministry of Education in year 1997 has made a call for a study on the needs of lecturers to undergo teaching training. It is very important to empower our lecturers in Malaysia higher education institutions with regards to their teaching expertise. However the call has generated more resistances than support from various professionals (Loh, 2008). Just like other faculty development programs, teaching training are also constantly battling with the issue of faculty resistance and it happen all over the world (Ahmed, 2013). The main argument against it was undergoing a teaching course is a waste of time. As lecturers, they need to spend their time to update their knowledge and skills particularly in their own discipline. Probably our lecturers in medical field have more strong reason since they have a very limited time as compared to others. Apart from teaching their students, they have to focus on providing a good service to society and not to forget their role as a researcher.

Most of the literatures relating to the development of medical teachers focuses on the acquisition and improvement of teaching or pedagogical skills by having a formal, generally short courses. Formal courses, however, may have limited impact (MacDougall and Drummond, 2005). Looking into USM context, the university has made some important policies with regards to the improvement of teaching skills among its lecturers. All new faculty members must attend a one-week teaching and learning short course (Kursus Pengajaran dan Pembelajaran) and it is compulsory for every new lecturer to undergo the course before their status as university lecturer

could be confirmed. This initiative is important for medical school as a measure to produce competent medical teachers.

Effective teaching by medical lecturers means a better learning experience for learners during medical training. This, in turn, would be translated into better care for the community in the future. In spite of the fact that, they have undergone the teaching and learning course, the main concern that might be important to consider is that, does the short course really effective to produce good teachers? To date, there was no proper research done to evaluate effectiveness of the course to produce good teachers. Perhaps the evaluation of teaching will provide some useful data to chart future direction of actions to improve the teaching and learning activities in the medical school.

1.3 Context of the Study

1.3.1 School of Medical Sciences (SMS), University Sains Malaysia

The School of Medical Sciences, Universiti Sains Malaysia was established in 1979. The enrolment of the first batch started in 1981. The school initially operated in the main campus in Pulau Penang, Malaysia. Beginning in 1983, the school moved in-stages to the new branch health campus in Kubang Kerian, Kelantan. By 1990, the whole medical school was based in Kubang Kerian Health Campus. The Health Campus is fully equipped with up-to-date teaching, research and patient care facilities. This is in accordance with the primary aims for its establishment to produce doctors and medical practitioners to meet the nation's needs as well as to upgrade the medical services of the country.

In addition to the undergraduate medical (MD) programme, the medical school also offers Masters of Medicine (M.Med), M.Sc and PhD in most of the medical related specialties.

The School of Medical Sciences has three main functions, which are:

(a) Teaching

(b) Research

(c) Patient-care service

1.3.2 Teaching Evaluation in SMS, USM

Evaluation is crucial for improving quality of teaching and learning. The main focus of evaluation in School of Medical Sciences, USM is evaluating students' performances. However, there were no proper and systematic teaching evaluation program done to evaluate quality of teaching or teaching effectiveness of lecturers, be it in lecture setting, clinical teaching, etc. Concerted efforts must be taken by all parties in the school to improve quality of teaching in order to give a better learning environment for our student to learn. However as described earlier, it might be very difficult to convince medical lecturers to undergo specific and continuing teaching training. Perhaps, by conducting teaching evaluation to evaluate their current teaching performances would be a precursor to trigger their thinking and hopefully, it would ease future movement to a have structured and systematic faculty development program particularly in improving their teaching skills.

1.4 Benefit of Study

"Transforming today's medical teacher for sustainable tomorrow's medical doctors"

This research focused on the evaluation of teaching behaviors in the lecture setting. It would provide information about strengths and areas for improvement related to the faculty members' teaching skills in term of teaching behavior from students' perspective. It would give clearer view about areas of concern that should be remedied for further improvement. By improving the faculty members' teaching skills, it will ultimately contribute to produce better graduates which in turn will provide the highest quality of patient care to the society in the future. This research can also be used as a basis for future research in other aspects of teaching modes, for example clinical teaching, problem based learning, and etc.

1.5 Justification of the Study

Teaching and learning is one of the core businesses of faculty members in the medical school. Accreditation standards for the Malaysian Qualification Agency (MQA) also concerns about the effectiveness of faculty members as a teacher in higher education institution. Every faculty member should be able to teach effectively so that students as the main stakeholder would learn better. To date, in USM School of Medical Sciences, there were no proper and systematic evaluation programs done to evaluate the effectiveness of faculty members' teaching skills. Are the faculty members qualified as teachers? What is the evidence to prove that their teaching is effective? These kind of questions would definitely spark positive and negative reactions from faculty members. However these are very important questions to ponder as we encourage people to use evidence based in practicing teaching and learning.

Therefore research in this aspect would provide invaluable data to the school, faculty members and students as a basis for future direction to enhance faculty development program particularly related to teaching aspects.

1.6 Operational Definitions

Students Rating

Student rating is one method or approach to evaluate teaching quality or teaching effectiveness among teachers.

Student Evaluation vs Student Rating

The word *"evaluation"* has a definitive and terminal connotation of determining worth. "Ratings" on the other hand, refers to data that need interpretation (Benton and Cashin, 2012).

High-inference Teaching Behavior

High-inference teaching behavior refers to characteristics of behavior which are global and abstract traits and more subjective such as *"interaction"* or *"rapport"* (Murray, 2007)

Low-inference Teaching Behavior

Low-inference teaching behavior refers to characteristics of behavior which are specific and concrete, such as *"encourage students to ask question"* and *"talk with students before or after class"* (Murray, 2007).

Instructors

Refers to faculty members or lecturer. We used the word of *instructor* instead of *lecturer* in this research to avoid confusion since oone of the category under variable of designation is “lecturers”.

1.7 Research Objectives

1.7.1 General Objective:

To evaluate instructors’ teaching effectiveness during lectures in pre-clinical years USM medical school through medical students’ ratings.

1.7.2 Specific Objectives:

1. To determine level of instructors’ teaching behaviour during lectures in pre-clinical years USM medical school as measured by Teacher Behavior Inventory.
2. To compare level of teaching behaviors among the instructors by the selected factors, i.e. gender, ethnic groups, specialities and designation.

1.8 Research Hypothesis

1. Level of instructors’ teaching behaviour is more than 50% of total mean score.
2. There is significant association between identified factors and level of teaching behaviour

CHAPTER 2

LITERATURE REVIEW

2.1 Evaluation of Teaching Performance

Evaluating the teaching performance of teachers is critical to ensure a high quality education for students (Kozub, 2008). There are two primary purposes of teaching evaluations which are formative and summative (McKeachie, 1997). The main purpose of formative evaluation is to improve and shape the quality of teaching, whereas the main purpose of summative is to "sum up" overall performance to make decision about something such as promotion, tenure, etc. (Berk, 2005). There are a lot of methods to evaluate teaching effectiveness. The most commonly used is a student rating and will be elaborated in the subsequent subchapter. Apart from student ratings, there are several other approaches being practiced by higher education institutions in evaluating effectiveness of teaching among their faculty members. There are so many literatures discussing about it, providing evidences to support their effectiveness, reliability, validity, feasibility, etc. However, it is beyond the scope of this study to discuss in detail for each of those teaching evaluation approaches. Some brief descriptions about each of them follow:

2.1.1 Peer Review

Peer review is defined as an intentional process of observation in which a faculty members sits in on a teaching session of their colleague with the express intention of offering feedback as a 'critical friend' (Kinchin, 2005). It has emerged in higher education for a variety of reasons (Lomas and Nicholls, 2005), possibly because of some teaching skills could not simply be rated by other approaches such as students

rating (due to their unqualified individual in certain area such as, knowledge of the field, quality of course content, how well faculty member meet expectation for teaching, etc.) (Nelson, 1998). For these aspects, their counterparts are the one who have the capacity to give formative as well as summative evaluation.

2.1.2 Self-Evaluation

Other source of teaching evaluation is self-evaluation. It demonstrates faculty knowledge about their own teaching and perceived effectiveness in the classroom (Knapper and Cranton, 2001). Research on this approach is inconclusive. Its validity & reliability are not proven (Berk, 2005). By comparing it with student ratings for example, studies have found that faculty rate themselves higher than (Aleamoni, 1999) or lower (Gentry and Pratt, 2003) than what their students rate them. Self-evaluations also have lack of validity and objectivity necessary for summative evaluation. However, some institutions have found self-evaluation helpful (Kahn, 1993) especially for reflective learning among faculty members. Self-evaluation is frequently associated with one emerging teaching evaluation which is teaching portfolio, an approach to teaching evaluation that many institutions are currently using (Kahn, 1993).

2.1.3 Teaching Portfolio

A teaching portfolio is defined as a teacher-compiled collection of artifacts, reproductions, testimonials, and productions that represents the teacher's professional growth and abilities (Riggs and Sandlin, 2000). It includes documents and materials that collectively suggest the scope and quality of a faculty's teaching performance (Seldin *et al.*, 2010). It is perhaps the most promising teaching evaluation since it could give positive impact on the improvement of teaching

(Saroyan and Amundsen, 2001) by encouraging reflective learning among faculty members. It can be used for both formative or summative purposes (Seldin, 2000)

2.1.4 Other Approaches of Teaching Evaluation

Other than above mentioned approaches, there are a few more approaches could be used for teaching evaluation such as employer ratings, administrator rating, students interviews, exit and alumni ratings, teaching scholarship teaching awards and learning outcome measures (Berk, 2005), however they are not widely used by higher education institutions and the reason might be due to their limitations.

2.2 Student Ratings

Students are the main stakeholder in our educational system and their opinions are considered as a vital source of information concerning the quality of teaching (Kozub, 2008). As mentioned earlier, student ratings are the most common source of teaching evaluation. In fact, student ratings have been used as the primary measure of teaching effectiveness since a few decades ago (Seldin, 1999). It can provide a measure of overt teaching skills and their students' perceptions concerning the effect of these skills on their learning experience (Saroyan and Amundsen, 2001). There are considerable evidences that student ratings can provide reliable and valid information about the quality of teaching among faculty members.

However some people still skeptical about student ratings as an evaluation tools for teaching effectiveness probably due to some misconceptions such as students cannot make consistent judgments, it's just popularity contests, students will not appreciate good teaching until they graduate, student feedback cannot be used to help improve instruction, emphasis on student ratings has led to grade inflation and a lot more

(Benton and Cashin, 2012). Aleamoni (1999) has addressed 16 most common myths with regard to student ratings by providing evidences to prove that the myths was only built on perception without having evidences to support it. These myths have in fact ignored more than half decade of credible research on the validity and reliability of student ratings (Benton and Cashin, 2012). The myths persist, probably due to ignorance of the research findings, personal biases, suspicion, fear, and general hostility toward any evaluation process (Feldman, 2007) .

2.3 Validity of Student Ratings

There are several ways used by researchers for validity studies of student ratings as one of robust teaching evaluation. It could be done by correlating student ratings with students' performance, correlating student ratings with other teaching evaluation approaches (peer-review, self-evaluation, etc.), examining possible sources of bias, manipulating administrative procedures and analyzing the underlying dimensions of ratings (Ory and Ryan, 2001). Evidence from all such studies affects the meaning and interpretation of student ratings or their construct validity (Messick, 1995).

2.3.1 Correlating student ratings with students' performance

The best indicators of effective teaching is student learning (Benton and Cashin, 2012) and the best evidence of student learning could probably be seen in their examinations performance. Studies have been conducted on multiple teachers who teach different sections of the same course. The instructors use the same syllabus and textbook and, most importantly, the same external final exam. There were correlation between student ratings on course and teachers and final exam scores (Benton and Cashin, 2012). Feldman (2007) has reviewed several studies of correlation between

final exam scores with various student ratings items (Feldman, 2007). Such research findings support the correlation between student ratings and student learnings and therefore support the validity of student ratings.

2.3.2 Correlating student ratings with other teaching evaluation approaches

If student rating correlate positively with other approaches of teaching evaluation (as discuss above), it would suggest the validity of student rating. Feldman has reviewed 19 studies and reported an average correlation of 0.29 between teacher self-ratings and student ratings (Feldman, 1989b). Marsh and Dunkin (1997) found a correlation of 0.45 between teacher self-ratings and student ratings on nine scale scores. Such findings again support the validity of student ratings.

2.3.3 Examining possible sources of bias

This would be the main concern of faculty members if student ratings are being used to evaluate their effectiveness of teaching. Age and teaching experience of teachers, their gender, race, personal characteristics and their research productivity (number of publication, etc.) would probably lead to bias. However researches have shown little or no relationship to student ratings (Benton and Cashin, 2012). Age of students, their gender, level of the students (senior vs junior), students' GPA as well as students' personality type might also be possible sources of bias. However, they also have little or no relationship to student ratings (Benton and Cashin, 2012)

2.3.4 Manipulating administrative procedures

Certain procedure used would affect the result of student ratings. For example, non-anonymous ratings versus anonymous rating (anonymous ratings would produce higher ratings and vice versa), teacher present or absent while students complete

ratings (presence of teacher would produce higher ratings) and purpose of the ratings being informed or not to the students (Benton and Cashin, 2012). However, these so called administrative procedures could be controlled with some standard operating procedure. The validity of student ratings depends on standardization of the procedure. In other word, to make sure the validity of the evaluation, the procedure for standardisation must be taken into account.

2.3.5 Analysing the underlying dimensions of ratings

There are a lot of student rating instruments used to evaluate teaching effectiveness. Different instrument has different dimensions of evaluation (teachers' attributes/skill to be evaluated). For example, Teacher Behavior Inventory (TBI) by Harry G Murray has eight dimensions namely clarity, enthusiasm, interaction, organisation, pacing, disclosure, speech and rapport (Murray, 1983). Students' Evaluations of Educational Quality (SEEQ) form by Herbert W Marsh has nine dimensions: learning/value, enthusiasm, organisation, group interaction, individual rapport, breadth of coverage, exams/grades, assignments, and workload (Marsh, 2007). There were no consensus among scholars about dimensions of evaluation to be included in teaching evaluation instrument. However, (Centra, 1993) has identified six factors commonly found in student-rating instrument as listed follows; course organisation and planning, clarity & communication skills, teacher student interaction (rapport), course difficulty (workload), grading and examinations and student self-rated learning. Every dimension used in any instrument has its own validity evidence and no single student ratings dimension is useful for all purposes (Benton and Cashin, 2012).

2.4 Reliability of student ratings

Student ratings of teaching are reliable measures (Cashin, 1995; Marsh and Dunkin, 1992). By definition, reliability refers to the consistency, stability, and generalizability of measurement data (Benton and Cashin, 2012). Usual concerns about student ratings are *consistency* of the ratings and interrater agreement. Reliability varies depending upon the number of raters. Hoyt and Lee have proven that more raters will produce more reliable result (Hoyt and Lee, 2002). *Stability* is concerned with agreement between student raters over time. Marsh and Hocevar (1990) have done 13 year period longitudinal study of the same teachers by using SEEQ instrument. The result was remarkably stable. The mean ratings for the cohort of 195 teachers showed almost no systematic changes over this period. They have concluded that teaching effectiveness as perceived by students was stable (Marsh and Hocevar, 1990). *Generalizability* refers to how interpretations of meaning of the data taken, accurately reflect the instructor's general teaching effectiveness, not just how effective he or she was in certain teaching scenario (similar course, same student they teach, etc). Benton et al. have quoted the work of Marsh (1984), when he had addressed this question by categorizing student ratings data from 1,364 classes into four categories (Benton and Cashin, 2012): (1) the same teachers teaching the same course but in different semesters; (2) the same teachers teaching a different course; (3) different teachers teaching the same course; and (4) different teachers teaching different courses. The aim of the study was to see effects of the teachers and the course. The teacher-related correlations were higher for the same teacher, even when teaching a different course.

2.5 Limitation of Student Ratings

In spite of the fact that there are a lot of studies to support the effectiveness of using students rating as a robust evaluation tool to evaluate teaching, some limitation need to be considered while using it for teaching evaluation. One of it is that, student ratings could only focus on a certain aspect of teaching, which is delivery of instruction (perceived by their students) and to some extent, the instructional design planned by the faculty which precedes it and the evaluation of learning which follows it. The underlying unobservable processes of teaching remain unexplored (Saroyan and Amundsen, 2001). Student ratings is a necessary source of evidence of teaching effectiveness for both formative and summative decisions, but not sufficient if it is used alone. Because of that, scholars currently suggest of combination of student ratings with other approaches of teaching evaluation (Berk, 2005)

2.6 Student Ratings and Teaching Effectiveness

As described above, students instructional rating is the most frequently used criterion measure in teacher effectiveness research. It provides direct measure of student satisfaction with instruction as well as indirect measure of outcome variables such as student learning and student motivation (Murray, 2007). Evidence that student ratings are suitable or appropriate as a direct or indirect measure of teacher effectiveness includes the following:

1. Significant correlation with more objective indicators of teaching effectiveness, such as student achievement (Cohen, 1981; Feldman, 1989a).

2. High retest and interrater reliability (Murray, 2007).

3. Moderate to high agreement with evaluations of the same instructors by other independent judges (Murray, 2007).

4. Relatively unaffected by a variety of variables that might lead to potential biases, such as grading leniency, class size, workload, etc. (Marsh, 2007).

5. It's useful in improving teaching effectiveness when coupled with appropriate consultation (Marsh, 2007)

2.7 Teachers' Factors and Teaching Effectiveness

As stated in chapter I, one of the objectives of this study is to determine demographic factors (gender, ethnicity, specialty and designation) which influence the lecturers' teaching effectiveness. Therefore it is important for us to discuss finding of other studies in other institution with regards to this issue. However, while discussing about the teacher's effects, we must also consider bias effects that would directly influence our judgment in determining which factors are real and which one need to be interpreted cautiously.

2.7.1 Gender of Instructors

Discussions of the effect of gender of instructors on student evaluations of their teaching appears to be quite inconsistent (Wachtel, 1998). Meta-analysis done by Feldman in 1992 reports that the majority of studies reviewed showed no difference in the global evaluations of male and female teachers (Feldman, 1992). However, some studies found that student ratings are biased against women instructors (Basow, 1994). Some researchers have found that female teachers are rated lower than their male teachers (Potvin *et al.*, 2009; Sandier, 1991) and surprisingly some other studies

found that college students rated female teachers significantly higher than male teacher (Bachen *et al.*, 1999; Feldman, 2007; Tatro, 1995). An interaction effect was also found, whereby students tended to rate same-gender teachers slightly higher than opposite-gender teachers (Wachtel, 1998).

2.7.2 Ethnicity

Some studies showed there were significant different between one ethnic to another. For example, in one study, student ratings of overall teaching ability, male faculty identified as "others" received the highest mean score, black male faculty received the lowest mean score. Meanwhile black female faculty received the second lowest mean scores. (Smith, 2009). In Malaysia context, there was no similar study published comparing teaching effectiveness with different ethnic groups.

2.7.3 Designation

There is a substantial evidence that, teachers become more effective over the first few years of their careers (Boyd *et al.*, 2008; Goldhaber and Hansen, 2010; Rivkin *et al.*, 2005). Comparing professors and teaching assistants performance, professors are rated more highly (Marsh and Dunkin, 1992). On average, students of first year teachers learnt less than students of more experienced teachers (Boyd *et al.*, 2008) and student performance of one year experience teachers have significantly lower than teachers with ten to fifteen years of experience (Kane *et al.*, 2008; Rockoff, 2004). Among those studies which found a significant (though weak) relationship, nearly all described that instructors of higher rank received more favorable ratings (Wachtel, 1998).

2.7.4 Age & Years of Experiences

Majority of studies found no significant relationship between age/experience and student ratings (Wachtel, 1998). However, among those studies in which a significant relationship was found, nearly all found an inverse relationship. In other words, instructors with older age and instructional experience received lower ratings. (Wachtel, 1998)

2.8 Conceptual Framework

Based on the literature review, summary of teaching evaluation research was summarized in figure 2.5.1

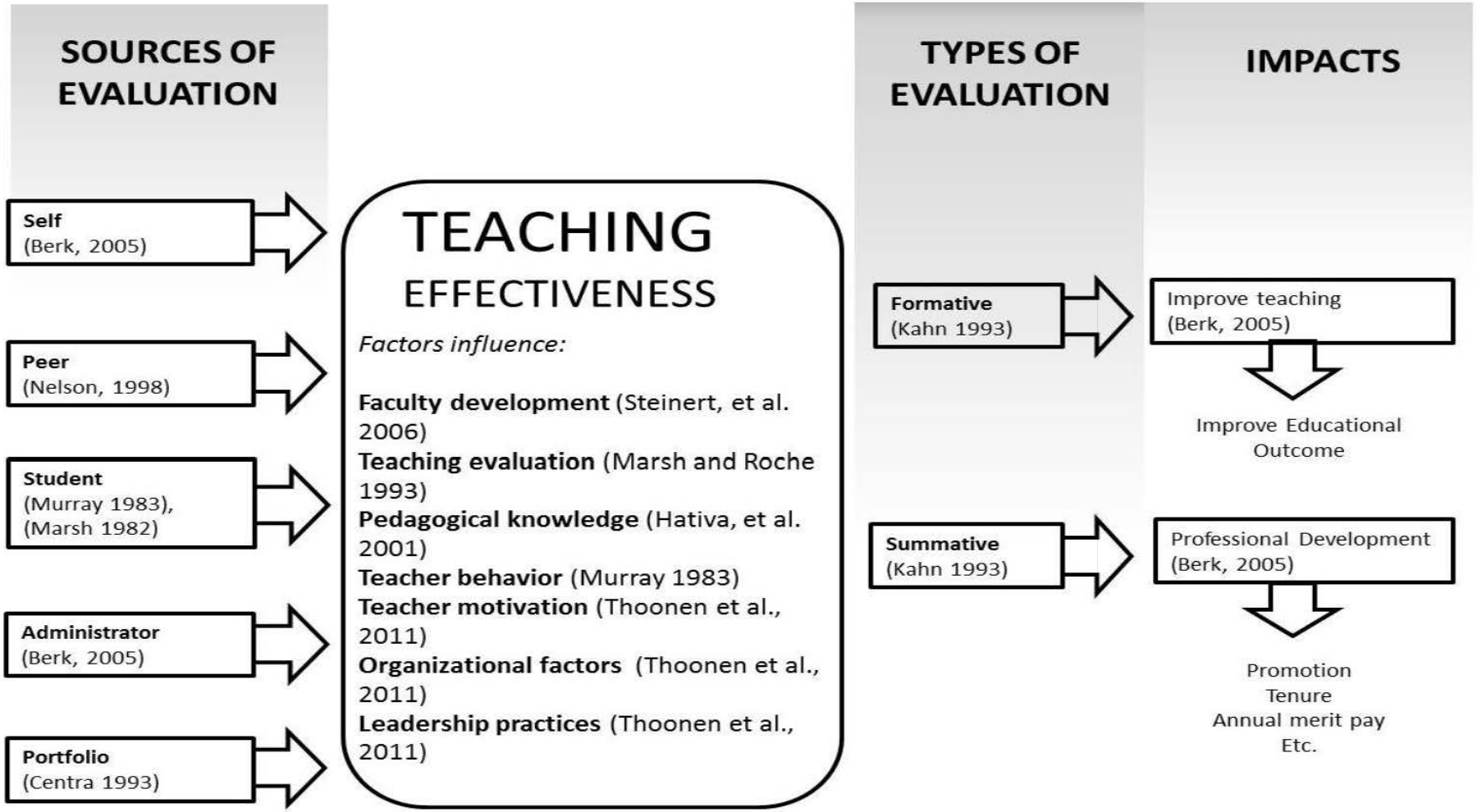

Figure 2.5.1: Conceptual framework on teaching evaluation

CHAPTER 3

METHODOLOGY

3.1 Study design

This study employed a cross-sectional design.

3.2 Subject of Study

Pre-clinical year's faculty members who involved in giving lecture for pre-clinical year students in School of Medical Sciences, USM.

Inclusion creteria:

1. Instructors in School of Medical Sciences, USM who had been appointed to deliver lecture for pre-clinical years.

Exlusion creteria:

1. Trainee instructors

3.3 Duration of Study

This study was conducted within 10 months period, from June 2015 until March 2016.

3.4 Study population

The study population was the instructors in the pre-clinical phase, School of Medical Sciences (SMS), Universiti Sains Malaysia (USM). The medical school is located in Kubang Kerian, Kelantan. Since its inception, it adopts SPICES model as the curriculum strategy with three phases of 5-year medical course. However, since 2014, its medical course was reviewed and changed to the two phase system. The

first phase (Phase I) consists of first and second year medical students (i.e., pre-clinical phase), and the second phase (Phase II) consists of third, fourth and fifth year medical students (i.e., clinical phase). In Phase I, the medical students learn basic science subjects with the integration of clinical application (particularly related to pathogenesis) based on the body systems. In Phase II, they learn clinical sciences through various clinical department rotations (e.g., internal medicine, general surgery, orthopedic) based on the apprenticeship system (PPSP, 2016).

In total, there are 324 instructors (faculty members) from all specialties. 229 (70.7%) of them are clinical instructors, the rest (94 (29%)) are non-clinical instructors (table 3.1.2). All professional rank of faculty members from VK5 professor until newly appointed faculty member DS45, DU51 and DS51 lecturers were shown in table 3.1.1. About one third of faculty members are DU54 lecturers.

Table 3.1.1: USM Medical School faculty members based on professional rank & designation

No	Professional Rank	Designation	Frequency (n)	Percent (%)
1	VK5	Professor	3	0.9
2	VK6	Professor	3	0.9
3	VK7	Professor	41	12.7
4	DU54	Assoc. Professor	61	18.8
5	DS54	Assoc. Professor	6	1.9
6	DU53	Assoc. Professor	1	0.3
7	DS53	Assoc. Professor	3	0.9
8	DU54	Lecturer	104	32.1
9	DUF54	Lecturer	1	0.3
10	DU53	Lecturer	20	6.2
11	DS53	Lecturer	1	0.3
12	DU52	Lecturer	2	0.6
13	DS52	Lecturer	2	0.6
14	DU51	Lecturer	35	10.8
15	DS51	Lecturer	37	11.4
16	DS45	Lecturer	4	1.2
Total			324	100.0

Table 3.1.2: Number of clinical and non-clinical instructors

Speciality	**Frequency (n)**	**Percent (%)**
Clinical	229	70.7
Non-Clinical	94	29.0
Total	324	100.0

3.4 Sample size

The sample size was estimated based on the parameters reported by a previous study (Hadie et.al, unpublished) on the psychometric evaluation of 60-item Teacher Behavior Inventory (TBI-60) developed by Murray (2007). TBI-60 consists of eight attributes. Confidence interval was set at 95% and margin of error (α) = 0.05.

Table 3.4.1: Calculation of sample size for each construct using SPCC software$^{\gamma}$ based on the pilot study data (Hadie et.al, unpublished)

Construct	Mean (SD)	Precision of finding* (significance level, $\alpha = 0.05$)	Sample size
Clarity	4.00 (0.55)	0.20	**29**
Enthusiasm	3.82 (0.52)	0.19	**29**
Interaction	3.64 (0.47)	0.18	**26**
Organisation	4.31 (0.57)	0.22	**26**
Pacing	3.88 (0.59)	0.19	**37**
Disclosure	3.86 (0.75)	0.19	**60**
Speech	4.39 (0.54)	0.22	**23**
Rapport	3.73 (0.75)	0.19	**60**

*Pecision of finding based on the pilot study data

$^{\gamma}$Formula used, $n = (Z\sigma/\Delta)^2$,

n = sample size

σ = population standard deviation

Δ = precision

Z = Z-score at significance level

Table 3.4.1 summarises the estimated sample size by the eight attributes. The largest sample size (60 participants) was selected for the research purpose. After considering 30% dropout rate, sample size needed was 86 instructors.

3.5 Sampling method

Simple random sampling was applied to select instructors who involved in delivering lectures to Phase I medical students between October 2015 and Febuary 2016. Figure 3.5.1 illustrates the sampling process.

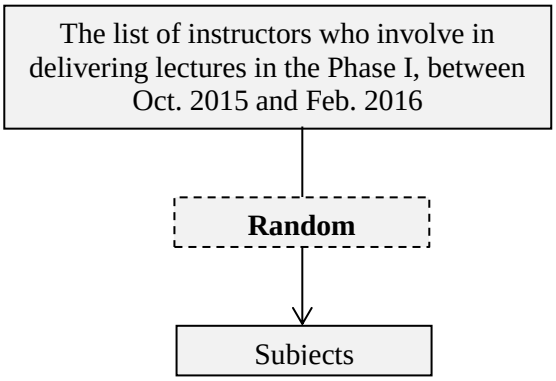

Figure 3.5.1: Sampling method

The student time table was used to find the suitable lecture session for rating. The time table was provided by the academic office, SMS, USM and they were issued course by course. List of all instructors involved in lecture session were determined. Invitation letter was sent to them together with consent and refusal form (see Appendix 3,4 and 5). For those who were willing to join the study they have sent the filled consent form, whereas those who refused to join sent back the filled refusal form. Instructors who did not reply were considered as had agreed to participate as clearly stated in invitation letter.

Lecture session to be rated were then determined. For the instructors who have multiple lecture sessions with the students, we have chosen their last lecture session. The reason for this was to allow better jugdments of the student evaluators on the teaching behaviors of the instructors. Perhaps, this approach will increase the reliability of the rating since they have a couple session with the instructors before the lecture session that would be rated.

3.6 Research instrument

Teacher behavior inventory (TBI) was originally developed by Harry G Murray and contains eight attributes with 60 items. The instrument was not copyrighted and can be reproduced (Benke and Hermanson, 1992). This study use the 32-item Teacher Behavior Inventory (TBI-32) as recommended by Hadie, et. Al, (unpublished) – refer to Appendix 1 for the details. TBI-32 measures seven attributes of teaching behaviors as summarized by Table 3.6.1. For TBI-32, there are two negative items that require reversed scoring for the analysis purpose (items 27 and 28).

TBI is an instrument used for measuring low-inference and high-inference teaching behaviors of an instructional design (Murray, 2007). Low-inference behaviors refer to specific, concrete teaching behaviors, such as *"Use headings and subheadings to organise lectures"* and *"States objectives of each lecture"* that can be recorded with very little inference or judgment on the part of a classroom observer. Conversely, high-inference behaviors refer to global, abstract traits such as *"organisation"* and *"rapport"*. In comparison to high-inference teaching behaviors, low–inference teaching behaviors were proven to be more validly rated by students and had positive impacts on students' learning, students' performance and overall teaching effectiveness (Murray, 2007).