EXPLORING MEDICAL STUDENTS' ASSESSMENT EXPERIENCE IN UNIVERSITI PUTRA MALAYSIA AND MAASTRICHT UNIVERSITY FROM THE PROGRAMMATIC ASSESSMENT PERSPECTIVES

SITI KHADIJAH BINTI ADAM

UNIVERSITI SAINS MALAYSIA

2020

EXPLORING MEDICAL STUDENTS' ASSESSMENT EXPERIENCE IN UNIVERSITI PUTRA MALAYSIA AND MAASTRICHT UNIVERSITY FROM THE PROGRAMMATIC ASSESSMENT PERSPECTIVES

by

SITI KHADIJAH BINTI ADAM

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

September 2020

ACKNOWLEDGEMENT

All praises to Allah swt for His guidance and strength given to me to complete this thesis. This study may not be completed without His blessing.

I would like to express my deepest gratitude to my main supervisor, AP Dr. Muhamad Saiful Bahri bin Yusoff, for his consistent guidance and advice throughout this study. This study would not have happened had it not been his idea. Also, my sincere gratitude goes to my co-supervisors, Dr. Ahmad Fuad bin Abdul Rahim and Dr. Nik Mohd Rizal bin Mohd Fakri for the knowledge and support.

Furthermore, I would like to express my deepest appreciation to Prof. Dr. Cees van der Vleuten for his guidance and support especially in the data collection process in Maastricht University. He is truly an inspiration. Not to forget, Ms Lilian Swaen, Prof. Dr. Sylvia Heeneman, AP Dr. Marjan Govaerts, Ms Lubberta de Jong, Ms Suzanne Schut and Dr. Maryam Asoodar for their advice and assistance. *Dank je wel!*

My sincere appreciation to Universiti Putra Malaysia for the financial support that allowed me to travel to Maastricht University. I am grateful to the Dean of FPSK UPM, Prof. Dr. Zamberi bin Sekawi; the Deputy Dean of Academic, AP Dr Norhafizah bt Mohtarruddin; and the head of JAM, Prof Dr. Sharida bt Fakurazi; for supporting me in pursuing my passion in medical education. I also would like to acknowledge the lecturers and staff in the Medical Education Department USM for the knowledge and help, especially Dr. Hanis and Dr. Anisa for sharing their experience in qualitative research. My special gratitude goes to the programme coordinator, Dr. Jamilah, for being so understanding and patient with us. To my FlexiMEd clan, Kak Wati, Kak Murni, Ati, Faiza, Jahan, Syaz, Evelyn as well as Aimen and Amal, I cannot thank you all enough for the amazing time and wonderful support. Also, I genuinely appreciate the support from my MERIU colleagues; Dr. Rafidah, Dr. Faridah, Dr. Puteri and Prof. Dr. Syafinaz.

My special thanks to Diyana, Kak Mona, Danial Akmal, Kak Fatim, Farisha and Amyra for the great help during the data collection and data analysis. To my family and friends, and those who directly and indirectly contributed to this research, thank you very much!

To my husband, Mohammad Khairil bin Arishah and my mother, Hamidah bt Hj Md Baki, I am most grateful for the love, patience, understanding and support. Both of you have made me strong and moved me forward every time I was down. Last but not least, in memory of my late father, Adam bin Mat Waseh, for instilling the love of learning, this is for you Abah. May Allah forgive him and have mercy on him. Ameen.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

A-KO	Arts-Klinisch Onderzoeker (Physician-Clinical Investigator)
A-KOfolio	The portfolio system for AKO students
AEQ	Assessment Experience Questionnaire
CanMEDS	Canadian Medical Education Directives for Specialists
EPASS	The portfolio system for Maastricht University students
MCQ	Multiple-choice question
MEQ	Modified essay question
OSCE	Objective structured clinical examination
OSPE	Objective structured practical examination
PBL	Problem-based learning
SAQ	Short answer question
SD	Standard deviation
UM	Maastricht University
UPM	Universiti Putra Malaysia

EXPLORING MEDICAL STUDENTS' ASSESSMENT EXPERIENCE IN UNIVERSITI PUTRA MALAYSIA AND MAASTRICHT UNIVERSITY FROM THE PROGRAMMATIC ASSESSMENT PERSPECTIVES

LIST OF APPENDICES

MENEROKA PENGALAMAN PELAJAR PERUBATAN BERKAITAN PENILAIAN DI UNIVERSITI PUTRA MALAYSIA DAN UNIVERSITI MAASTRICHT DARI PERSPEKTIF PENILAIAN PROGRAMATIK

ABSTRAK

Penilaian programatik berbeza dengan pendekatan penilaian tradisional dari segi pemerhatiannya yang subjektif dan penekanannya terhadap penilaian formatif yang berkepentingan rendah, maklum balas, refleksi dan penggunaan maklum balas untuk membimbing pembelajaran yang seterusnya. Penilaian ini bersifat individu dan melibatkan banyak data penaksiran untuk menilai pelajar secara holistik. Kajian ini dijalankan untuk meneroka pengalaman pelajar Universiti Putra Malaysia (UPM) dan pelajar Universiti Maastricht (UM) berkaitan penilaian programatik semasa latihan program perubatan. Persamaan dan perbezaan berkaitan pengalaman mereka mengenai pendekatan penilaian telah dikaji. Reka bentuk penyelidikan kaedah campuran telah digunakan di mana borang soal selidik *Assessment Experience Questionnaire (AEQ)* versi 4.0 diedarkan kepada 318 pelajar UPM dan 119 pelajar UM untuk mendapatkan persepsi mengenai pengalaman mereka berkaitan kaedah penilaian. Persepsi mereka diterokai dengan lebih mendalam melalui temu bual separa berstruktur secara individu bersama 27 pelajar UPM dan 16 pelajar UM. Perbandingan skor AEQ dianalisa dengan ujian t bagi min tak bersandar dengan menggunakan perisian statistik IBM SPSS versi 23. Data kualitatif dianalisa dengan analisis tematik dengan menggunakan perisian ATLAS.ti 8 untuk menghasilkan kategori dan tema yang terhasil daripada temu bual tersebut. Berdasarkan analisis, didapati bahawa para pelajar dapat memenuhi keperluan penilaian dan mempunyai dorongan intrinsik dalam pembelajaran. Pelajar UPM cenderung mempunyai usaha pembelajaran yang lebih tinggi dan pendekatan pembelajaran yang lebih mendalam berbanding pelajar UM. Sebaliknya, pelajar UM cenderung untuk lebih menghargai nilai penilaian formatif dan mempunyai pemahaman yang lebih tinggi, walaupun secara sederhana, mengenai standard tugasan yang diperlukan berbanding pelajar

UPM. Pelajar UPM menekankan bahawa mereka jarang mendapat maklum balas, tetapi mereka lebih menghargai maklum balas yang diterima berbanding pelajar UM. Kajian ini juga mendedahkan bahawa para pelajar dari kedua-dua amalan penilaian berasa bimbang dengan kesubjektifan dan keadilan sistem penilaian mereka. Pendekatan penilaian programatik dihargai oleh pelajar kerana pembelajaran yang diperibadikan dan relevan dengan amalan profesional. Sebaliknya, pendekatan penilaian tradisional mungkin kekurangan peluang maklum balas, kemahiran reflektif serta penilaian yang holistik dan autentik terhadap prestasi pelajar. Tambahan pula, jumlah tugasan, maklum balas dan refleksi yang berlebihan dalam suatu sistem penilaian mungkin menyebabkan kelesuan pelajar. Hasil dapatan dari kajian ini menunjukkan bahawa penilaian menyebabkan kesan yang signifikan terhadap pembelajaran pelajar dan persepi mereka adalah penting ketika merancang dan menilai sistem penilaian.

EXPLORING MEDICAL STUDENTS' ASSESSMENT EXPERIENCE IN UNIVERSITI PUTRA MALAYSIA AND MAASTRICHT UNIVERSITY FROM THE PROGRAMMATIC ASSESSMENT PERSPECTIVES

ABSTRACT

Programmatic assessment differs from the traditional approach of assessment in terms of its subjective observation and emphasis on low-stake formative assessments, feedback, reflection and the use of feedback to guide future learning. The assessment approach is tailored to an individual learner and involves many assessment data points to provide a holistic judgement of the learner. Thus, this study was carried out to explore the experience of Universiti Putra Malaysia (UPM) students and Maastricht University (UM) students in relation to programmatic assessment during medical training. The similarities and differences in their perceived experience pertaining to the assessment approach were investigated. A mixed-method research design was adopted in which the Assessment Experience Questionnaire (AEQ) version 4.0 was administered to 318 UPM students and 119 UM students to gain their perceived assessment experience. Their perceptions were further explored in individual semi-structured interviews conducted with 27 UPM students and 16 UM students. The comparison of AEQ score was analysed using independent t-test in IBM SPSS Statistics software version 23. Qualitative data were analysed using thematic analysis in ATLAS.ti 8 software to generate categories and themes emerged from the interviews. Based on the analysis, it was found that the students were able to cope with their assessment demands and had intrinsic motivation to learn. UPM students were inclined to put higher learning effort and a deeper approach to learning compared to UM students. In contrast, UM students tend to be more appreciative of the value of formative assessments and have higher understanding, albeit moderately, on the standards of work required than UPM students. UPM students

highlighted that they rarely received feedback, yet, they appreciated the feedback more than UM students. This study also revealed that the students from both assessment approaches were concerned with the subjectivity and fairness of their assessment system. The programmatic assessment approach was appreciated by the students for its personalised learning and relevance for professional practice. On the other hand, the traditional approach of assessment may be lacking in terms of feedback opportunities, reflective skills as well as a holistic and authentic judgement on students' performance. Additionally, excessive volume of tasks, feedback and reflection in an assessment system may cause exhaustion to the students. The findings from this study demonstrated that assessment can cause a significant impact on the students' learning and that their perceptions are important when designing and evaluating an assessment system.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Assessment is the process of documenting the level of learners' knowledge, skills and attitude and its purpose is to make judgement and decision about student's learning against a certain standard or benchmark (Downing and Yudkowsky, 2009). The term 'summative assessment' refers to assessments given to learners periodically at a particular time, which are used as a part of the grading process in learning. On the other hand, the term 'formative assessment' refers to assessments that take place throughout the teaching and are employed primarily to provide feedback to the teaching and learning method.

In most medical schools worldwide including Universiti Putra Malaysia (UPM), student assessments are both formative and summative. Formative assessments are usually carried out in class throughout the course to guide learners to improve. Meanwhile, summative assessments are carried out at the end of the course to evaluate the level of learner's understanding. The results of summative assessments are therefore used to determine if a learner can proceed to the next semester or to retake the course. This is also known as the 'assessment of learning'. A higher stake examination such as Professional Examination is held at the end of the preclinical phase and in the final year to determine whether or not the learners can progress to clinical years or graduate, respectively. This traditional approach of assessment is assumed to be teacher-centred in which the teacher is the centre and expert in teaching and learning activities. For the past several years, there have been a lot of effort to re-addressing the process of assessment to become more learner-centred (Duncan and Buskirk-Cohen, 2011). The purpose is for learners to play an active role in defining their learning strategy and develop intrinsic motivation for learning with the support from teachers.

In recent years, a new approach to assessment known as 'programmatic assessment' has increasingly been implemented in medical schools around the world (Schuwirth et al., 2017). It was first implemented at Maastricht University (UM) in 2006. This approach has shifted away from focusing on individual assessment tools or episodes, towards continuous collection and analysis of learner's competence and progress using a variety of assessment instruments (Schuwirth and Van der Vleuten, 2011). This is because all assessment tools or methods have their strengths and weaknesses. For example, some assessment tools are valid, authentic and robust for high-stakes decision making but have low reliability; some are less feasible, while others are less valid but have higher reliability (Wilkinson and Tweed, 2018). With programmatic assessment, each time a learner is assessed, the focus is put on the strengths and weaknesses of that particular learner. In each assessment, the learner receives feedback and recommendations for further learning instead of a decision of 'pass' or 'fail'. Once there is a sufficient amount of information from a variety of assessment episodes outlining the learner's progress, a final evaluation takes place (Van der Vleuten et al., 2012; Wilkinson and Tweed, 2018). This approach optimises the 'assessment for learning', subsequently making a credible and trustworthy decision about the learner's progress.

1.2 Justification of the Study

It is widely known that the educational culture in Asian countries is examination-oriented (Koh and Luke, 2009). In most medical schools in Malaysia, summative assessment plays a dominant role in the medical programme to determine learner's progress. In recent years, the value of formative assessment in enhancing learning experience has been the attention in educational literature. To date, there are a lot of studies done to explore the influence of assessment; both summative and formative, on learning experience and learner's behaviour. The findings however, were inconsistent. While some studies found that formative assessment promotes

positive learning behaviour than summative assessment (Al Kadri et al., 2009; Mitra and Barua, 2015), others found conflicting results (Al Kadri et al., 2011; Jessop et al., 2014). Consequently, it has been suggested that a combination of formative and summative assessments is needed to maximise the educational impact of assessment (Al Kadri et al., 2011; Lau, 2016). It is therefore critical to ensure that appropriate and timely assessment is done since it has a huge impact on the quality of learning.

A radical approach to assessment known as programmatic assessment has been developed by Van der Vleuten et al. (2012) to optimise the learning experience. They redefined the traditional dichotomy between formative and summative assessment as a continuum of stakes, ranging from low to high-stakes decision. The intention was to foster the culture of 'assessment for learning' and improve decision making on learner's progress. The programmatic assessment focuses on multiple low-stake assessments, followed by rich and narrative feedback with learner's reflection throughout the study period. In UPM medical programme, some elements of programmatic assessment such as low-stake (formative) assessments and feedback are being practised. Therefore, this study was designed to explore the impact of the current assessment system on the learners' experience and behaviour. By comparing the findings between these two medical schools, the study also provided a theoretical insight into the interaction of different cultures and contextual factors on learners' experience. These findings do not only describe how the assessment in UPM and UM influence students' learning, but also provide valuable information to evaluate teacher's performance and the quality assurance system. Additionally, emerging data retrieved from this study will support the improvement of the current assessment approach in the respective university to maximise learning experience.

1.3 Objectives of the Study

1.3.1 General objective

To explore students' assessment experience during medical training in relation to programmatic assessment.

1.3.2 Specific objectives

1. To explore the student's perceived assessment experience from the programmatic assessment perspectives in UM and UPM.

2. To compare the perceived assessment experience of UM and UPM students from the programmatic assessment perspectives.

1.4 Research Questions

1. How do students describe their assessment experience in both medical schools?
2. What are the similarities and differences in the perception of assessment experience in both medical schools?

CHAPTER 2

LITERATURE REVIEW

2.1 Assessment

According to the Standards for Educational and Psychological Testing, assessment is defined as, “Any systematic method of obtaining information from tests and other sources, used to draw inferences about characteristics of people, objects, or programs” (Downing and Yudkowsky, 2009). Assessment is an integral part of teaching and learning in medical education. It is mainly used to assess and make judgement whether or not students have met the learning outcomes over a course of study. Eventually, assessment plays a role in certifying safe and competent doctors who can serve the public. Assessment can be broadly categorised into either ‘formative’ or ‘summative’.

2.1.1 Formative assessment

Formative assessment refers to “all activities that are undertaken by teachers and students in assessing themselves which provide information to be used as feedback to modify teaching and learning activities” (Black and Wiliam, 2010). It is also referred to as the ‘assessment for learning’. Formative assessment is usually conducted continuously throughout a course or learning module. Its main purpose is to provide feedback on learners’ performance, guide their learning process and improve their performance as a preparation prior to summative assessment. It can be done formally or informally. However, it is not used for deciding on students’ progression, hence is considered as a low stake.

Formative assessment can occur spontaneously during class for example, when question-and-answer sessions are conducted during a lesson. Another example is when a teacher reads misunderstanding in the body language of students during a class session and queries the student about her understanding (Dixson and Worrell, 2016). Examples of planned and formal formative assessments are quizzes,

assignments and workplace-based observations. These activities are usually followed by feedback on the learners' performance.

2.1.2 Summative assessment

Summative assessment refers to "the cumulative assessments, that usually occur at the end of a unit or topic coverage to capture what a student has learned, or the quality of the learning and judge their performance against some standards" (National Research Council, 2001). It is also referred to as the 'assessment of learning'. The main purpose of summative assessment is to make a judgement on learners' performance. It has major consequences to learners; therefore, it is considered as a high stake. Additionally, it is typically done formally and allows teachers to measure whether or not the learners are competent to progress. The examples of summative assessment are the examinations that occur at the end of courses or semesters. Also, summative tests for graduation or licensure.

2.2 Traditional Approach of Assessment

Majority of medical schools in Malaysia are practising a traditional approach of assessment. This refers to the conventional method in which summative assessment is dominant and highly focuses on grades to determine learners' progress. This approach comprises of tests such as multiple-choice questions and essays given to learners to measure how much they have learned. It is influenced by the strong examination-oriented educational cultures especially in Asian countries (Koh and Luke, 2009).

The traditional approach is modular, with a standardised end-of-course assessment a learner has to pass (Van der Vleuten et al., 2017). It is mostly a single-occasion test, which measures what learners can do at a particular time (Quansah, 2018). It is assumed that when learners passed an examination, they have mastered the knowledge or skills they learned. There have been arguments stating that the

traditional approach of assessment may not be authentic as it relies heavily on testing lower-order thinking skills of the learners (Koh and Luke, 2009). Often, it only requires a learner's ability to memorise and reproduce knowledge during an examination. Moreover, since the assessment is conducted on a single occasion, it is inadequate for learners to demonstrate how they would perform in real-life situations (Quansah, 2018) thus, its validity may be questionable (Abeywickrama, 2012).

Another drawback of the traditional approach is the lack of post-assessment feedback and undervalues the benefits of feedback (Flores et al., 2015). In many practices, test scores and grades are being emphasized. In a study by Harrison et al. (2015), it was found that even though the feedback was provided after the summative assessment, learners tended to overlook its relevance and focused more on the grades. This is supported by a recent study which proved that learners' attention and effort tended to focus only on getting good grades, which may induce poor learning behaviour (Hattingh et al., 2019). Even though grading is a form of feedback, it does not provide enough information on how to improve future performance (Schinske and Tanner, 2014).

It is crucial to note, however, that the traditional approach of assessment is not entirely bad despite the reasons mentioned above. The dominance of summative assessment in the traditional approach also has its positive impacts on learning (Bennett, 2011; Lau, 2016). Due to its standardised manner of testing, it is more practical to implement the traditional approach of assessment in medical schools with a large class size (Duncan and Buskirk-Cohen, 2011). It is also more objective to determine learners' progress based on the various assessment tools (Quansah, 2018). However, in recent years there has been a shift from a traditional approach to a longitudinal assessment with an emphasis on self-directed learning and post-assessment feedback. There is in fact, no perfect and 'one-size-fits-all' assessment system. Importantly, it is vital to have a well-designed and feasible approach that meets the educational objectives of each medical programme.

2.3 Feedback

Feedback is closely related to assessment and both are an integral part of learning. Constructive feedback from a teacher informs the performance of a learner and guides improvement for future performance. Feedback is important for learners in providing them with a basis to recognise their deficiencies and guiding them in correcting any mistakes. It also helps learners to clarify their learning goals and have insight into actual performance and consequences (Krackov, 2013).

Feedback effectiveness is influenced by several factors including the environment in which feedback takes place and the teacher-learner interaction (Perron et al., 2016). Feedback should be timely, concise and focuses on the areas of improvement. It is more effective when learners receive it immediately than when it is delayed. The content should be clear, mutually understood and relevant to promote positive changes. Besides, Eva et al. (2012) highlighted that the feedback-giver, which is the teacher, should be credible to provide feedback and establish a good relationship with the learners to make the situation comfortable.

It is also important that the learners understand the purpose of feedback and utilise it for learning. Feedback alone may not be sufficient to guide further learning (Hattie and Timperley, 2007) thus, reflection and follow-up on feedback are also essential to improve outcomes (Sargeant et al., 2009). Consequently, the feedback process can be tedious. Sometimes, the highly-structured hierarchical environment especially in the traditional approach of assessment, promotes a one-way flow of information from teacher to learners. Therefore, a change of culture that values effective feedback and fosters mutual trust in constructive improvement is inevitable.

2.4 Programmatic Assessment

The concept of programmatic assessment was introduced by Lambert Schuwirth and Cees van der Vleuten emphasising the idea of ‘assessment for learning’ (Schuwirth and Van der Vleuten, 2011). It is an approach in which routine information

about the learners' competence and progress are continually collected, analysed and complemented with additional assessment information. This information are useful for the students and their respective mentors, besides allowing high-stakes decisions at the end of a learning phase (Schuwirth et al., 2017).

2.4.1 Principles of programmatic assessment

Van der Vleuten et al. (2017) formulated some theoretical principles of programmatic assessment. First, each assessment is considered as a single data point since there is no single assessment that is perfect and able to cover all elements of Miller's pyramid. Therefore, a large number of assessments are required for a reliable judgement (Van der Vleuten and Schuwirth, 2005). These assessments need to be longitudinally aggregated to a competency framework. In the traditional approach, assessments are aggregated cross-sectionally by the method. For example, an objective structured clinical examination (OSCE) is used to gather different attributes such as communication skills and physical examination techniques across all its stations. This can be an inappropriate compensation for the pass or fail decision of the examination, such as good communication skills compensating for poor physical examination (Wilkinson and Tweed, 2018). Programmatic assessment, on the other hand, aggregates attributes across methods to meaningful entities. For example, communication-related information are gathered not only from OSCE, but also from a set of mini-CEX and multisource feedback assessment (Van der Vleuten et al., 2017). There can be a mixture of assessment methods depending on its appropriateness and purpose for that particular point of time.

Next, the distinction between 'formative' and 'summative' is not very useful, considering that any assessment should be formative and summative, only to a varying degree. Therefore, a continuum from low to high stakes seems more useful in the programmatic assessment. In low-stake assessment, the results have minor consequences to the learners in terms of certification or promotion, while high-stake

assessment may cause remarkable consequences (Van der Vleuten et al., 2012). A pass or fail decision is not given on a low-stake single data point. Whereas, the decision in high-stake assessment shall be based on many data points and more robust. The higher the stakes of the decision, the more robust information are needed to make the decision. Intermediate review based on several data points is done to inform the learners of their progress. This is considered as diagnostic (How is the learner progressing?), therapeutic (What remedies are needed?) and prognostic (What might happen to the learner?) (Van der Vleuten et al., 2017).

Next, the programmatic assessment approach is developmental, guided by feedback and self-directed learning. Each data point is optimised for learning, generating meaningful feedback to the learners. Each time a learner is assessed, the focus is given on the strengths and weaknesses of the learner. The feedback can be given by teachers, peers as well as patients. The purpose is to promote good learning strategies. All assessments and feedback information are discussed with a mentor or supervisor. Mentors may probe, stimulate learners' self-reflections and discuss remediation plans. However, the learners must be pro-active and self-direct their learning by determining their learning objectives. Indirectly, the programmatic assessment involves personalised learning experience that supports individual learning priorities.

Finally, expert judgement is needed to combine information across various data points during high-stakes decisions (Van der Vleuten et al., 2012). The information is both quantitative and qualitative, and it is impossible to simply averaging it. Therefore, it should be done credibly and transparently, using a holistic approach. These experts should be trained and are independent of the learning process to avoid bias and conflict of interest (Van der Vleuten et al., 2015). This is to ensure the trustworthiness of the decision-making process. All decisions are not limited to just pass or fail, but ideally to be justified to inform the learners of their progress.

2.4.2 Model of programmatic assessment in action

Van der Vleuten et al. (2012) proposed a generic model for programmatic assessment in the educational context based on their previous empirical research and the principles of programmatic assessment. This model was proposed with the assumptions that the training programme is learner-centred, favouring holistic approaches to learning and deep-learning strategies.

In a single period of training activities, learners undergo various learning tasks (denoted by the small circles) such as lecture, problem-based learning (PBL) sessions, individual or group assignment, clinical sessions, small group discussion and self-study (Figure 1). Learners may produce some learning artefacts (denoted by the large circles) during training activities, for instance, logbook, project report and presentations. Simultaneously, learners are also assessed in many forms, such as written tests, clinical examinations and peer evaluation. Each small triangle in Figure 1 represents an assessment method. Some assessments can be an evaluation of the artefacts from the training activities, including a presentation of a case report (denoted by the dotted ellipse). Each assessment is considered as a single data point and is a low stake. However, it shall be developmental and maximally meaningful to learning in which the assessors are able to provide rich and extensive feedback to the learners. At certain times, there can be some mastery-oriented kind of assessment, certifying whether or not the learners have mastered certain skills (denoted by the grey triangles). An example would be intravenous injection skill to a manikin, which has to be drilled until mastery is achieved. Even though these assessments do not determine the passing or failing of the learners, they should be traceable, if needed during later judgement and should encourage desirable learning behaviours. In this period too, there are supporting activities that consist of reflective activity and social interaction between the learners and mentors/supervisors or with their peers (denoted by the connected small circles). Learners interpret the feedback received and plan new

learning tasks or goals. Continuous self-directed learning is emphasised throughout the supporting activities (Van der Vleuten et al., 2017).

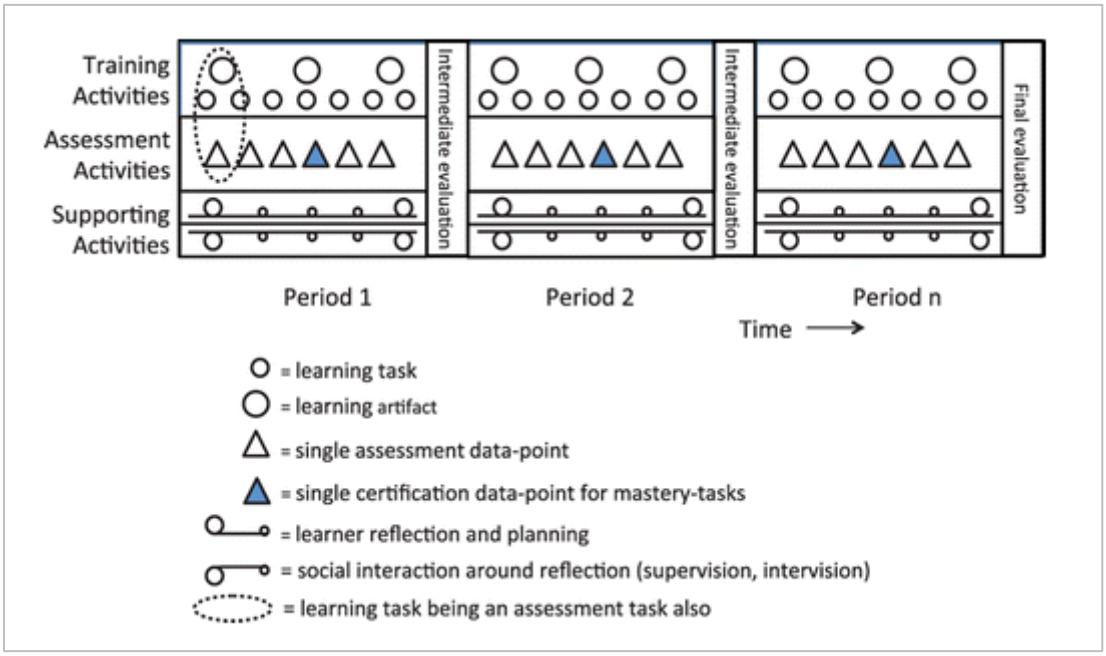

Figure 2.1 Model for programmatic assessment in action.
(Adapted from Van der Vleuten et al., 2012)

At the end of the period, all artefacts, assessment data and selected information from the supporting activities are assessed in an intermediate evaluation. This evaluation is done by an independent and authoritative group of assessing committee. The assessors are the experts who are trained to make decision during the evaluation. They provide information on areas of strength and improvement (diagnosis) as well as suggest remedial steps to help the learners achieve the expected learning outcomes (therapy). They also may predict the learners' performance outcomes later in the training programme (prognosis). The provision of feedback and recommendation of further learning during this point of time is contributing to the credibility of the final decision (Van der Vleuten et al., 2017).

Accordingly, this one period (or cycle) of training, assessment and supporting activities can be repeated depending on the nature of the programme or availability of resources. After an appropriate number of cycles, the final evaluation is carried out. This is a high-stake decision to determine learner's progress. Hence, the process is more robust, involves numerous data points from the previous cycles and have major consequences to the learners. It is done by the same assessing committee during intermediate evaluation. The decision is not limited to a pass or fail decision, but also indicates distinctive excellence of performance. In certain cases, the committee may provide suggestions for remediation before allowing the learners to progress or graduate.

This model depicts a certain learning period, consisting of repeated cycles of training, assessment and supporting activities, ending with a decision on the learner's progression. The learning period can be repeated as many cycles as appropriate to complete the curriculum. The duration of each cycle does not have to be equal, as long as it optimises learning value across the assessment programme. Nevertheless, the model is limited for the programme in action and not the other elements of the programmatic assessment framework, namely programme support, documentation, improvement and justification (Dijkstra et al., 2010).

Based on this model, it can be stated that programmatic assessment is labour-intensive. The assessors need to undergo training to provide effective feedback to the learners. They should take the assessment tasks seriously, by taking the time to provide feedback or record a narrative on a form. Poorly trained assessors may give ineffective feedback that is valuable to the learners to guide their learning process. The assessing committee is required to make a holistic and trustworthy decision so if the decision is challenged, it can be defensible and justified, even in a court of law (Van der Vleuten et al., 2012).

A strong and stable system is vital to handle the massive learner's information for implementing programmatic assessment. For example, the medical programme at Maastricht University, the Netherlands utilises an electronic portfolio as a repository of all learning and assessment information as well as formal and informal feedback. This facilitates administrative and logistical aspects of the process and enables a quick overview of aggregated information (Van der Vleuten et al., 2015).

2.4.3 Comparison between programmatic assessment and traditional approach of assessment

The traditional approach of assessment is typically modular, with summative decisions and grades at the end of modules. The module is considered completed when learners passed the module. If failed, learners are required to resit the assessment or repeat the whole module as a remediation. In contrast, programmatic assessment is developed based on constructivist learning theories, in which learners construct new knowledge and skills based on their current/previous knowledge (Schuwirth and Van der Vleuten, 2011; Vygotsky, 1997). The assessment focus is rather developmental, which emphasises feedback, reflection, use of feedback to optimise learning and tailored to individual learning priorities (Van der Vleuten et al., 2015).

The decision on learner's performance in the traditional approach is standardised and is rather deductive, which is made based on limited information of students' performance. It is usually done on a single-occasion test, which measures what the learners can do at a particular point in time (Quansah, 2018). Often, quantitative strategies are used to aggregate information sources, for example, averaging test scores to reach the pass or fail decision. On the other hand, programmatic assessment guides learning and the decision is made on a wide array of evidence. It is highly narrative and information-rich, which needs qualitative judgement from a group of experts. The programmatic assessment reflects a move

away from objective measures to subjective observations on learners on integrated tasks in various contexts.

The comparison between programmatic assessment and traditional approach of assessment is summarised in Table 2.1.

Table 2.1 Comparison between programmatic assessment and traditional approach of assessment

Characteristics	Programmatic assessment	Traditional approach
Purpose of assessment	▪ For learning	▪ Of learning
Focus	▪ Learner-centred ▪ Formative-inclined	▪ Teacher-centred ▪ Summative-inclined
Decision making	▪ Inductive ▪ A group of panels to decide the fate of students based on triangulation of enriched assessment data	▪ Deductive ▪ Based on numerical data of single-occasion test
Method of judgement	▪ Longitudinal	▪ Cross-sectional
Perspective of judgement	▪ Both objective and subjective	▪ Objective

2.5 Impact of Assessment on Learner's Behaviour

It is a well-known fact that assessment drives learning and study behaviour. These include what learners focus their attention on, how much they study, their quality of engagement with learning tasks, as well as their understanding and future learning following feedback (Gibbs and Simpson, 2003; Wanner and Palmer, 2018). Previous studies have shown how assessment affected positive and negative behaviour in learning.

A study by Al Kadri et al. (2009) has found that the stressful situation during summative assessment led to the surface approach of learning while the learners preferred formative assessment due to its attached feedback to improve their learning strategies. The findings were similar to a recent study revealing that learners tended to

only memorise without having to understand the content to pass the summative examination (Hattingh et al., 2019). They also thought that examinations have no long-term benefits and admitted that they would forget materials after an assessment has ended. Similarly, Cilliers et al. (2012) discovered that summative assessment is able to promote strategic learning as the main aim is only to pass the assessment. This somehow induced poor learning behaviour. Iannone and Simpson (2017) emphasised the fairness of summative assessment, which had cause anxiety to the learners and impacted their performance. Interestingly, the learners regarded that a variety of assessment methods such as oral examination, project and presentations are fairer than a single summative assessment and more valuable for their learning. In a recent study done in a medical school in Australia, the learners admitted that they put more effort into tasks and assessments that they thought are relevant for their future practice (Preston et al., 2020). They also acknowledged the value of formative assessments in addressing their knowledge gaps, providing motivation for them to work hard and directing future learning for improvement.

Lau (2016) suggested that summative assessment, which extrinsically motivates students learning, should be aligned with formative assessment and opportunities to improve. The reason is to enhance positive learning behaviour and possibly a better performance outcome. This is supported by a previous study mentioning that repeated formative assessment attached with feedback was positively correlated with learner's performance in summative assessment (Mitra and Barua, 2015). Meanwhile, Broadbent et al. (2018) utilised an innovative approach by providing low-stake tasks and audio feedback through an online system. They found that their learners were engaged and motivated to learn. The faculty must keep in mind however, that the weightage of formative and summative assessments may also affect learning behaviour. The presence of high summative and low formative assessment leads to learners' orientation towards grading with a focus only on achieving good grades

(Jessop and Tomas, 2017). Jessop et al. (2014) also observed that learners tended to give fewer priorities to formative assessment since it was not graded. In the same study, it was also revealed that the higher the volume of formative assessment in a programme, the less it fostered a deep learning approach among learners.

Based on the above literature, it is interesting to note how learning behaviour is mostly affected by the purpose of assessment, be it formative or summative. While some literature portrayed that formative assessment as ‘good’ whereas summative assessment as ‘bad’ towards learning behaviour, it would be inappropriate to diminish summative assessment in a learning programme. Any assessment systems can have positive impacts on learning and learner’s outcome if carefully designed.

A system with a high volume of feedback, emphasises the use of feedback and consists of appropriate assessment with clear aims and standards, may enhance positive impacts on learning (Jimaa, 2011). Programmatic assessment, for example, fosters feedback culture and self-directed learning. There are some studies have investigated learners’ behaviour in medical schools with programmatic assessment setting. A study by Heeneman et al. (2015) demonstrated that programmatic assessment caused supporting and inhibiting effects on learning. Learners developed deep and continuous learning strategy and they valued the benefit of feedback and reflective practice. On the other hand, some learners did not acknowledge the formative features of assessment and became stress due to the packed schedule and the quality of feedback received. They were less motivated to learn due to the lack of individual decision moments. Learners who are used to summative-dominant culture may have difficulties at first in adjusting with programmatic assessment (Harrison et al., 2016). Nevertheless, they became less dependent on grades, yet were driven to aim for excellence and focused more on preparation to practice.

2.6 Measuring the Perceptions of Assessment Experience

Currently, there are limited instruments used to measure learner's perception on assessment and its impact on their learning, especially for higher education. The most common instrument to determine student's assessment experience is the Assessment Experience Questionnaire (AEQ). It was initially developed by Gibbs and Simpson (2003), which focuses on understanding how assessment influences learning. AEQ is used to evaluate the effectiveness of an assessment system and improve that assessment. Based on their study, there were several conditions for assessment to support student learning. These conditions were clustered under five headings; (i) quantity and distribution of student effort, (ii) quality and level of student effort, (iii) quantity and timing of feedback, (iv) quality of feedback and (v) student response to feedback (Gibbs and Simpson, 2003). Since then, AEQ has been revised by researchers to suit its context and include additional ideas based on an extensive literature review (Hattingh et al., 2019).

Currently, AEQ is being used in Transforming the Experience of Students through Assessment (TESTA) process in the United Kingdom (UK) universities as well as in Australia, India and the United States of America (TESTA, 2017). TESTA is a triangulated research method which gathers data from programme leaders about their assessment and feedback, as well as students about their experience of assessment and feedback using AEQ and focus group discussion. The data will later suggest how assessment and feedback in the whole programme could be modified to improve student learning (Jessop and Tomas, 2017). The latest version of AEQ, AEQ version 4.0 developed by Wu and Jessop (2018), is currently being used for TESTA process in the UK.

Another example is the Students' Perceptions of Assessment Quality Questionnaire (SPAQQ) to evaluate assessment quality from higher education students' perspectives (Gerritsen-Van Leeuwenkamp et al., 2018). SPAQQ consists of

39 items covering the six assessment quality variables, namely the effects of assessment on learning, fairness of assessment, conditions of assessment, interpretation of test scores, authenticity of assessment and credibility of assessment. However, since SPAQQ is only recently developed, it has not been widely used and further evidence of validity is needed to support its interpretations.

On the other hand, Student Assessment for Learning Questionnaire (SAFL-Q) which consists of 28 items, focuses more on the assessment for learning practices in classrooms namely scaffolding activities as well as feedback and self-monitoring (Pat-El et al., 2013). Although SAFL-Q is widely being used, it is however, limited to only secondary education students. Therefore, at present, only AEQ is the established tool that is widely used to measure learners' perceptions on their assessment approach, particularly in higher education.

2.7 Conceptual Framework

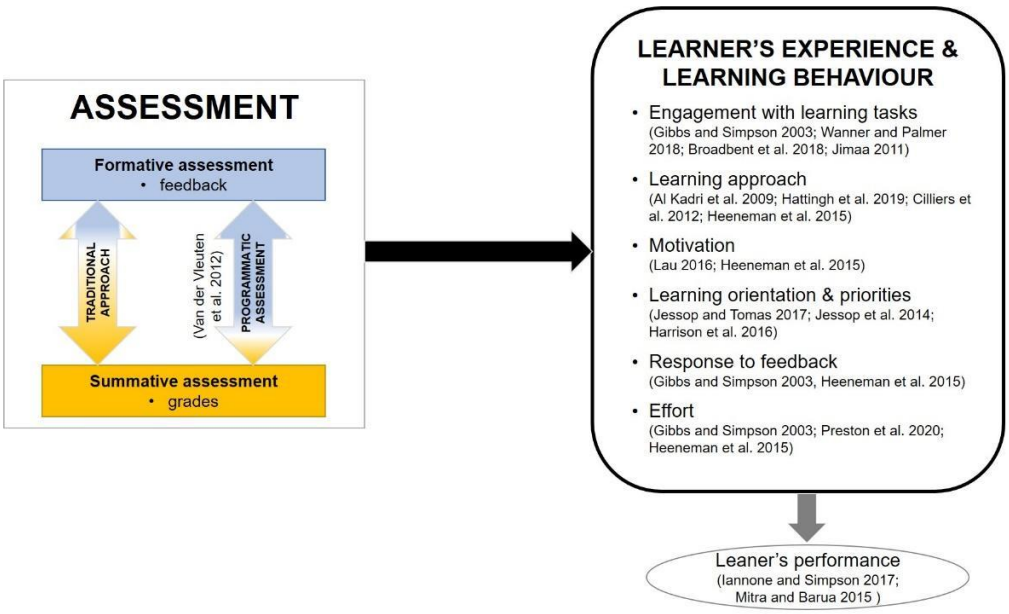

Figure 2.2 Conceptual framework on assessment and its impact on learner's experience, behaviour and performance

Assessment is categorised into formative assessment and summative assessment. The traditional approach of assessment is dominated by summative assessment that emphasises grading, while the programmatic assessment emphasises low-stake formative assessment and the use of feedback in learning. The type of assessment influences students' learning and their learning behaviour such as engagement with learning tasks, learning approach and learning effort. These will eventually have an impact to their academic performance.

CHAPTER 3

METHODOLOGY

3.1 Research Design

This study used a concurrent mixed method design, which integrated quantitative and qualitative data in a single study to allow for more complete analysis and deeper understanding of the participants behaviour. It employs triangulation design to obtain different (quantitative and qualitative) but complementary data on the same topic. The intent of using this design is to bring together the differing strengths and non-overlapping weaknesses of quantitative methods namely large sample size, trends and generalisation, with those of qualitative methods namely small sample size, details and in depth (Creswell and Plano Clark, 2018). For quantitative study, the study design is cross-sectional, while qualitative study adopts a phenomenological study design and employs semi-structured interview method for data collection.

3.2 Participants of the Study

The study was conducted in the Faculty of Medicine and Health Sciences, Universiti Putra Malaysia (UPM) and Faculty of Health, Medicine and Life Sciences, Maastricht University (UM), the Netherlands. The study participants were the undergraduate students of Doctor of Medicine (MD) program in UPM who experienced the traditional approach of assessment, and Physician-Clinical Investigator (A-KO) program students in UM who experienced programmatic assessment approach.

Target participants were based on the following criteria:

Inclusion criteria:

1. UPM MD or UM A-KO students
2. Had completed the first year of their study.

Exclusion criterion:

1. Students who were on semester break or in the examination period during the data collection period.

The inclusion criteria were decided based on the research objectives in which students were required to provide their perception of their assessment environment in their institution. Therefore, they needed to have at least completed their first year to experience all assessment components in a particular year. On the other hand, the exclusion criterion was decided to ensure only available students were able to commit to participating in the study. Hence, this study did not interfere with students who were on their semester break or having an examination.

3.3 Programme Structure

3.3.1 MD programme in UPM

UPM MD is a 5-year undergraduate program with about 100 students for each cohort. The curriculum consists of two phases namely preclinical and clinical phases which run for 2 and 3 years, respectively. The curriculum is integrated, which emphasises on problem-based learning and student-centred learning. All teaching and learning activities are conducted in English. The program adopts the traditional approach of assessment where summative assessment is dominant in their curriculum. For the preclinical phase, teaching and learning activities are based on the systems of the body with the introduction of clinical experience. For each system, the students are assessed through a mid-course assessment that contributes 15% to the total score of the end of semester examination. They are also required to complete a variety of ungraded tasks such as seminar presentations, quizzes and laboratory reports. At the end of the semester, the students are assessed through written and practical examinations including multiple-choice questions (MCQs), short answer questions (SAQs), objective structured practical examination (OSPE) and objective structured

clinical examination (OSCE). The students need to submit their research dissertation and must pass the end of the preclinical phase examination to proceed to the clinical phase.

For the clinical phase, the students in smaller groups rotate the different clinical postings supervised by clinical supervisors. The students are required to engage with patients independently, complete their logbook, document the cases they encounter, attend seminar presentation and submit written assignments including reports and case write-up. They are assessed at the end of each posting through a few assessment methods such as MCQs, modified essay questions (MEQs), OSCE, long case and short case examinations depending on the posting. In their fifth year, they need to sit and pass the clinical phase examination consisting MCQs, MEQs, OSCE, long case and short case examinations before being awarded with the degree of medicine. Although not compulsory, lecturers and/or mentors are encouraged to provide post-assessment feedback to the students. Remediation is provided for students who do not achieve the minimum passing mark.

3.3.2 A-KO programme in UM

UM A-KO is a 4-year graduate-entry programme with 50 students per year. The curriculum is competency-based and adopts the CanMEDS framework (Frank et al., 2005). The programme is fully conducted in Dutch and is divided into four phases. In the first two phases, the students are exposed to modules with similar themes with different approaches. During Phase 1, the focus is on basic knowledge of medicine portrayed by written PBL problems while in Phase 2, the emphasis is on clinical reasoning, diagnosis, treatment and prevention using a real-patient based PBL approach (Godefrooij et al., 2010). The students begin their clinical internship in Phase 3 focusing on patient care, whereas in Phase 4, they are expected to conduct scientific research alongside clinical work. Throughout the programme, the students sit for written assessments in certain topics and progress test comprising vignette-based

MCQs for four times a year (Heeneman et al., 2017). Besides, they are required to complete several written assignments such as literature review, critical appraisal and research proposal.

The programme emphasises on competence-oriented training in which the students are required to work on their development of several competencies based on the CanMEDS framework, namely medical expert, communicator, collaborator, manager, health advocate scholar and professional (Frank et al., 2005). The programme assessment adopts programmatic assessment approach, that is based on the overarching combination of feedback, judgements, evaluations and tests. The students are responsible for collecting information about their learning, analysing feedback and making personal reflection following their personalised learning goals with guidance from their counsellors. All information about their learning goals, assignments, assessments, feedback from peers and teachers and reflection are collated in an electronic portfolio known as A-KOfolio. The portfolio is assessed at the end of each year by a group of the examination committee. Students who do not meet sufficient competencies are required to undergo remedial sessions.

3.4 Quantitative Study

3.4.1 Sample size determination

The number of participants required in each setting was calculated based on the two means formula by Aday and Cornelius (2006). Therefore, the minimum sample required from UM was 119 and from UPM was 318, with 95% confidence interval and 80% power, after considering 0% dropout rate.

3.4.2 Sampling method and participants recruitment

A simple random sampling method was used to recruit participants. The name list of the students who met the inclusion criteria was obtained from the academic office of the faculty. Each student was tagged with a number as an identification. Random