DEVELOPMENT AND VALIDATION OF ENTRUSTABLE PROFESSIONAL ACTIVITY (EPA) ENCOMPASSING EXAMINER COMPETENCIES IN THE MALAYSIAN ANAESTHESIOLOGY PROGRAMME

NOORJAHAN HANEEM BINTI MD HASHIM

UNIVERSITI SAINS MALAYSIA

MAY 2021

DEVELOPMENT AND VALIDATION OF ENTRUSTABLE PROFESSIONAL ACTIVITY (EPA) ENCOMPASSING EXAMINER COMPETENCIES IN THE MALAYSIAN ANAESTHESIOLOGY PROGRAMME

by

NOORJAHAN HANEEM BINTI MD HASHIM

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science (Medical Education)

May 2021

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ABA   American Board of Anaesthesiology

ACGME   The Accreditation Council for Graduate Medical Education

CBME   Competency-based Medical Education

EPA   Entrustable Professional Activities

FCAI   Fellowship of the College of Anaesthesiologists of Ireland

FGD   Focused Group Discussion

FRCA   Fellowship of the Royal College of Anaesthetists

MMC   Malaysian Medical Council

MQA   Malaysian Qualification Agency

UIAM   Universiti Islam Antarabangsa, Malaysia

UITM   Universiti Teknologi MARA

UKM   Universiti Kebangsaan Malaysia

UM   Universiti Malaya

UPM   Universiti Putra Malaysia

USM   Universiti Sains Malaysia

LIST OF APPENDICES

x

DEVELOPMENT AND VALIDATION OF ENTRUSTABLE PROFESSIONAL ACTIVITY (EPA) ENCOMPASSING EXAMINER COMPETENCIES IN THE MALAYSIAN ANAESTHESIOLOGY PROGRAMME

ABSTRAK

Latihan pemeriksa adalah sangat perlu untuk memastikan kebolehpercayaan proses dan keputusan peperiksaan. Kajian ini bermatlamat menghasilkan dan mengesahkan satu siri *entrustable professional activities* (EPA) untuk digunapakai dalam melatih pemeriksa Bahagian Pertama program latihan pakar anestesiologi di Malaysia. Menggunakan kerangka EPA untuk menghuraikan tugas pemeriksa adalah suatu konsep yang baru. Penyelidikan ini telah dijalankan dalam tiga peringkat (pengenalan masalah dan analisa keperluan kajian, pembangunan dan validasi EPA), di mana keperluan latihan, tugas hakiki dan kompetensi pemeriksa telah dikenalpasti melalui semakan dokumen, perbincangan kelompok sasaran dan kajiselidik Delphi. Seterusnya, tugas pemeriksa dihuraikan dalam kerangka EPA sebelum disahkan oleh pakar. Tujuh EPA penting telah dikenalpasti dan diterangkan dalam kerangka EPA. EPA tersebut adalah penyediaan soalan dengan skema jawapan dan rubrik mengikut tatacara sediaada, menyemak soalan, penilaian calon semasa peperiksaan pengendalian ujian lisan, penglibatan membina semasa mesyuarat pembuatan keputusan, penulisan laporan pemeriksa dan pemakluman prestasi calon kepada universiti. Kompetensi yang dikenalpasti adalah kompetensi berorientasikan tugas (kepakaran dalam bidang, teknik peperiksaan, pengetahuan dalam pendidikan perubatan, pengetahuan berkenaan proses dan sistem peperiksaan, komunikasi dan kerja berpasukan) dan kompetensi profesional (agensi, kebolehpercayaan, integriti, rendah diri dan murah hati). Aktiviti yang perlu untuk penauliahan pemeriksa adalah aktif dalam menjalankan tugas klinikal, menghadiri latihan tugas pemeriksa dan

pemerhatian proses peperiksaan. Kesemua tujuh EPA mencapai kriteria penerimaan peringkat instrumen, 0.80. Kajian ini telah mengesahkan keperluan latihan pemeriksa Bahagian Pertama anestesiologi. Tujuh EPA yang menjadi tugas hakiki pemeriksa telah dikenalpasti dan dihuraikan. Maklumat ini boleh digunapakai dalam melatih pemeriksa Bahagian Pertama bagi program latihan pakar anestesiologi di Malaysia.

DEVELOPMENT AND VALIDATION OF ENTRUSTABLE PROFESSIONAL ACTIVITY (EPA) ENCOMPASSING EXAMINER COMPETENCIES IN THE MALAYSIAN ANAESTHESIOLOGY PROGRAMME

ABSTRACT

Examiner training is essential to ensure the trustworthiness of the examination process and results. This study aims to develop and validate the Malaysian Primary Anaesthesiology examiner's core entrustable professional activities (EPA) to facilitate examiner training. We describe the novel concept of using the entrustable professional activities (EPA) framework for examiners. In a three-staged (problem identification and general needs assessment, EPA development and EPA validation) mixed-method study, examiner training needs, essential examiner tasks and competencies required to perform them were identified through triangulation of information from document review, focus group discussions and three rounds of Delphi questionnaires. These tasks were then described in the EPA framework and were validated by a panel of experts. Seven EPAs were identified as essential and were described using the EPA framework. These EPAs were: constructing questions with answer key and rating rubrics according to guidelines, vetting examination questions, rating candidates' performance, conducting a viva examination, constructive participation in decision-making meetings, writing an examiner report and giving feedback regarding candidates' performance to respective universities. The competencies required to perform them are classified as task-related and professional competencies. Task-related competencies consist of subject matter expertise, examination technique, medical education knowledge, knowledge of the examination system and process, communication and teamwork. Professional competencies include agency, reliability, integrity, humility and benevolence. Activities that would support examiner

entrustment and credentialing include active clinical practice, attending task-specific examiner training, and observing the examination process. All seven EPAs met our criteria for scale-level content validity index acceptance of 0.80. Our findings confirmed the need for examiner training for the Malaysian Anaesthesiology Primary examination. We identified and described seven EPAs performed by Primary Anaesthesiology examiners, with their essential competencies and features. This information can be used for examiner recruitment and training by the Malaysian Anaesthesiology Specialty Conjoint Board to the quality of the Primary examination.

CHAPTER 1

INTRODUCTION

1.1   Background

The Anaesthesiology programme in Malaysia is a 4-year Clinical Master programme aimed at training anaesthesiologists for the country. The course is offered by five local universities: Universiti Malaya, Universiti Sains Malaysia, Universiti Kebangsaan Malaysia, Universiti Putra Malaysia and Universiti Islam Antarabangsa, Malaysia. The course content is similar in all universities, the summative examinations are held as a Conjoint Examination. The course delivery may differ amongst the universities, in alignment with the University rules and regulations.

This clinical training comprises three stages: basic (Year 1), intermediate (Years 2 and 3) and advanced (Year 4). There are two summative examinations, Primary, comprising of Physiology & Clinical Measurement and Pharmacology at the end of the first year of training; and Final Examination, comprising of Clinical Anaesthesiology, at the end of training. The examinations are held as a conjoint examination, where all the universities offering the programme take turns hosting and organising the examination. Examiners include lecturers from all universities, specialists from the Ministry of Health and external examiners from other Anaesthesiology training programmes.

The Anaesthesiology Conjoint Board, includes members from the five universities, the Ministry of Health and College of Anaesthesiologists, Academy of Medicine, Malaysia. This Board ensures that the curriculum content, programme delivery, trainee selection, programme accreditation is aligned to the needs of the specialty, university rules and regulations and Ministry of Health (MOH)

requirements. The Board also oversees and sets the standards of the Conjoint Examination.

Currently, a national level curriculum review is in progress. As part of the curriculum review, the assessment system is being evaluated. An identified gap is examiner competence which is not explicitly documented. Current examiner eligibility criteria for the Primary examination are Clinical Qualification (Master in Anaesthesiology, or equivalent), active in clinical work and being appointed as programme trainer. (Examiner Term of Reference, Anaesthesiology Conjoint Board, 2017).

Gaps in the definition of examiner competence have also been identified by Khera et al (2005). The issues highlighted are the development and content of training of examiners, suggestions of examiners competencies and the gaps in available research on the reliability and validity of examinations.

Tavakol and Dennick in their AMEE guide (2017) share that "assessment is a systematic process that collects and interprets information from examination data to legitimise examination content and student marks". Therefore, examiners need to participate in the examination process, ask relevant and fair questions, and rate the trainees' responses so that trainees' marks reflect their actual knowledge and skills.

We plan to ensure our examination trustworthiness by identifying and describing the tasks to be performed by the Primary Anaesthesiology examiner, the competencies required to perform them, and worksheets to facilitate operationalisation.

1.2   Problem statement

Examiner training is essential to ensure the trustworthiness of the examination process and results. (Blew et al, 2010; Iqbal et al, 2010; Tekian and Yudkowsky,

2020). Current Primary Examiner training includes being a written examination (short answer question, SAQ) marker and oral examination observer. The examiners do not receive formal training on basic concepts of the utility of assessment, the role of blueprinting, examination technique and do not receive formal feedback on their performance. At the beginning of the examination, an examiner guide containing examination subjects and components, criteria for passing, description of examination methods, question contribution, board of examination, confidentiality clause and terms of reference of examiners role, is provided by the host University. The Chief Examiner will brief the examiners on the examination process (Conjoint Primary Examination Guide for Examiners 2018).

The shift of medical education towards competency-based training has introduced the concept of entrustable professional activities, (EPA), where the competency framework can be translated to clinical practice. EPAs are descriptors of work; therefore, this concept can be translated from clinical training to clinician-educator training. This idea is advocated by Dewey et al (2017) in using EPA for clinical teachers. Placing clinical educators training in the EPA framework would facilitate training, manage expectations, promote self-directed professional development and improve the quality of clinical educators. The key question to ask when one adopts this frame is "can we trust this educator to perform this task?" The idea of utilising EPA as a tool for examiner training and accreditation is novel.

As there is no formal examiner training and the potential benefits of EPA, we propose a needs and EPA-directed, competency-based approach to develop, implement and evaluate a programme for Primary examiner training to ensure effective examiner behaviour. This study will focus on the needs assessment and the development and validation of Primary Examiner EPA.

1.3   Justification

This study is important to ensure the data collected on examinee performance only relates to their knowledge and skills and is not caused by construct under-representation, for example, under-sampling of the syllabus, and construct-irrelevant variance, for example, insufficient cues, ambiguous questions, or unclear rating rules. (Lineberry, in Yudkowsky, Park and Downing, 2020).

The research output will be used for examiner recruitment and training in the national Anaesthesiology specialist training programme.

1.4   Research questions

1. What are the competencies required for Anaesthesiology trainers to effectively function as examiners for Primary Anaesthesiology examination?

2. Do the EPAs developed represent the competencies of an effective Primary Anaesthesiology examiner?

3. Are the EPAs developed aligned to the concept of EPAs?

1.5   Research objectives

The <u>general objective</u> of this project is to develop and validate a series of Core Entrustable Professional Activities of the Primary Anaesthesiology Examiners.

The <u>specific objectives</u> are:

1. To determine the examiner competencies required for Anaesthesiology trainers to effectively function as examiners for the Primary Anaesthesiology examination

   a. Through analysing gaps in the current examiner preparation

   b. Literature review

   c. Document review

2. To construct a framework of Entrustable Professional Activity (EPA) encompassing examiner competencies for Primary Anaesthesiology examination

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This study aims to develop an EPA for examiners so that it can be used for examiner training and serve as a guide for examiners to ensure validity and fairness of the examination.

The importance of assessment on learning is emphasised by many authors. According to Downing & Yudkowsky (2009), healthcare professionals have the serious responsibility of caring for patients, therefore, the decisions made regarding their education has an impact on health delivery outcomes. It must therefore be supported by the best evidence available. Apart from the serious impact on care delivery, Khera et al (2005) shared that since assessment drives learning, the analysis of its content and process will allow for the development of a robust training programme for paediatricians. Tavakol and Dennick (2017) opined a valid and reliable assessment ensures the appropriateness and accuracy of the information on student learning gathered during the assessment process. To ensure we can train the best anaesthesiologists for the country, taking into account the field's manpower shortage and impact on healthcare delivery outcomes, paired with the educational impact of assessment, examiners must be trained to ensure the decisions made are valid and fair.

2.2 Examiner training

Training of examiners is important, as written by Holmboe (2011): it is not the tool used for assessment, but rather the person using the tool. As it is difficult to design the 'perfect tool' for performance tests and redesigning a tool only changes 10% of the

variance in rating, educators must now train faculty in observation and assessment. It is not irrational to extrapolate this effect on written and oral examinations. Holmboe also shares the reason to have a training programme for assessors. They are changing curriculum structure, content and delivery and emerging evidence regarding assessment; building a system reserve; training programmes are opportunities to identify and engage change agents and allows faculty to form a mental picture of how changes will affect them and improve practice. Enlisting a respected faculty member during training will promote the depth and breadth of change.

Khere et al (2005) shares this opinion. They also shared the paediatric training programme's current situation and their concerns regarding the content, process, and validity of the assessment. They shared the work done by the Royal College of Paediatrics and Child Health in defining examiners competencies, selection process and components of a training programme. They include principles of assessment, examination design, writing questions, interpersonal skills and professional attributes, managing diversity, and assessing the examiners' skills. They believe these contents will ensure the assessment is valid, reliable and fair. As our faculty members have different knowledge levels and experiences, it is crucial to assess their learning needs and provide them with appropriate learning opportunities.

2.3   The concept and ingredients of trust and entrustment

The concept of EPA is based on the element of trust (ten Cate 2005, Hauer 2013, Savoldelli 2016, ten Cate 2017). The definition of trust and its related concepts was outlined by ten Cate et al. in 2016. Trust is defined as "confidence in or reliance on some quality or attribute of a person or thing". It is a result of complex environmental (interactions, context and situations) and personal (information processing, thoughts

and motivations) (Hauer et al. 2013). To entrust is "to confide the care or disposal of [a thing or person] or the execution of [a task] to or with a person". Entrustment is "the action of entrusting or the fact of being entrusted." Putting things into the clinical context, the object of care is the patient, and the task is the professional activity executed in the care of the said patient. The decision being made by the <u>supervisor</u> is to allow the <u>transfer of responsibility</u> of patient care to the learner (ten Cate 2018). Factors that determine trust are supervisor, supervisee, supervisor-supervisee relationship, task and context (Hauer et al. 2013, ten Cate et al. 2016). It is advocated to weigh the supervisee (learner) traits compared to the other factors (ten Cate 2016). We extend this concept of trust in the examination context, where the object of 'care' is the examination integrity and validity. The task is performing as an assessor or examiner, and the context is the Anaesthesiology Primary Examination; the learner and supervisors are the examiners.

The authors (ten Cate et al. 2016) also elaborate the different models of trust in a supervisor-supervisee relationship (presumptive, initial and grounded trust), which explains current tacit practice. Presumptive trust is credential-based trust, a decision made by supervisors before meeting the trainees. It is made based on the certification trainees already possess or based on the recommendation from previous supervisors. In clinical practice, this is seen annually when new trainees enrol in the postgraduate training programmes. Initial trust is based on first impressions and pattern recognition of trainee traits similar to previous trainees. It is vulnerable to biases from the halo and self-fulfilling prophecy effects. In the clinical setting, a supervisor trusts a trainee based on their past experiences with trainees having similar traits. Grounded trust is based on essential and prolonged exposure to the trainees in the confined specific conditions. Grounded trust is preceded by sufficient observation and data to qualify trainees to be entrusted with clinical tasks. In the Anaesthesiology Primary Examination context, the

presumptive (credential-based) trust system is in place, as per the Examination Term of Reference: clinical qualification of at three to five years (depending on the task to be performed), active in clinical and academic duties, credentialed as a trainer in the programme. We would like to enrich this presumptive trust system by providing opportunities for training and collecting evidence for credentialing. We also hope that by having this evidence, we can move towards a grounded trust system.

Ten Cate et al. (2016, 2018) proposed two categories of entrustment decisions based on how these decisions are made. Ad hoc decisions occur on a daily (or nightly) basis. It is made with the estimated trustworthiness of the trainee, the estimated risk of the situation, suitability of this task at the moment for this learner. They may not be generalised to future entrustment decisions in a different context. An example of this is when a trainee is allowed to assess a critically ill patient unsupervised (or with limited supervision). Summative entrustment decisions are grounded in sufficient evaluation, made by programme directors or a competency committee based on planned and accepted standards leading to a privilege to perform the same procedure in a different context, with increasing responsibility and autonomy to prepare the trainees for independent practice. The aim of EPA is for supervisors to be able to make such decisions. We hope, with this study, we can provide the examination board with the opportunity for making summative entrustment decisions when credentialing examiners.

The outcome of entrustment decision is the determination of the level of supervision a trainee requires (ten Cate 2016). When sufficient ability in an EPA is demonstrated, the trainee can be trusted to perform this activity unsupervised (Savoldelli et al. 2016). In this study, we aim to develop a shared understanding of

standards required of examiners to help examiners prepare for independent examination duties and programme providers to identify, recruit and train examiners.

In a thematic analysis, ten Cate and Chen (2020) described trainee features that are important to help supervisors make entrustment decisions. These features are agency, reliability, integrity, capability and humility. They advocate using these five features when making entrustment decisions. These five features provide the basis for describing the knowledge, skills, attitude and behavioural examples in the EPAs developed.

The AMEE Guide No. 140 by ten Cate and Taylor (2020) recommends including these sections when describing an EPA: a title, specifications and limitations, potential risks in case of failure, most relevant competency domains, required knowledge, skills, attitudes and experiences to allow summative entrustment, information sources to assess progress and support entrustment, entrustment level at stages of training and period to expiration if not practised. Our EPA worksheets include these sections.

The widespread use of the term EPA has seen its inclusion into many programmes and has led to items that do not fit into ten Cate's definition of EPA (2005, 2013). One example is the Association and American Medical Colleges' 13 core EPAs that would support entry into residency programmes. Tekian (2017) highlighted four EPAs that do not fit into ten Cate's definition of EPA, related to these problems: they are not discrete tasks, therefore unsuitable for focused entrustment decisions, educational objectives written as EPAs, includes adjectives that refer to proficiency levels, too broad, EPAs that are unsuitable for entrustment decisions.

This issue has led to the development of at least two instruments to evaluate the quality of EPA, QUEPA (Post et al., 2016) and EQUAL (Taylor et al., 2017).

QUEPA aimed to help programmes create high-quality EPAs for milestones-based assessment. The literature review identified the salient domains of an EPA: focused, observable, clear intention, realistic, articulates trustworthiness, generalisable across rotations and integrates multiple competencies. For each domain, the team agreed upon three items, structured on a five-point scale of agreement. The instrument was pilot tested on ten locally developed EPAs, leaving five domains with acceptable inter-rater reliability: focus, observable, realistic, generalisable and integrates multiple competencies. This tool, named QUEPA, was validated using 46 locally developed EPAs and the Alliance for Academic Internal Medicine End of Training EPAs. Statistical analysis showed significant associations between QUEPA scores and ACGME competencies, activity type, and practice locations. QUEPA ratings also showed good to excellent inter-rater reliability in the End of Training EPAs.

Despite the good inter-rater reliability, QUEPA had several limitations. The final domains did not completely align with ten Cate's defining qualities of EPA, and it did not have descriptive anchors for each scale (Taylor 2017). The limitations of QUEPA led Taylor et al. to develop EQual. The team developed a rubric with descriptive anchors and rating scales to reliably measure how much an EPA meets its purpose. They reviewed the literature to identify the elements of EPA definition and common misconceptions. The constructs of EPA were reorganised into three categories: discrete units of work, entrustable and essential tasks of the profession and education focused. Descriptive anchors were created for each category across a five-point scale. The validation process was by applying the rubrics to 31 EPAs developed for residency training in Canada by five experts with experience in EPA development and programme directors. A generalisability study using EPAs as objects of measurement was

performed to evaluate the overall reliability of EQual. Reliability and discriminatory values for each item were also calculated.

As these tools are available and show high reliability in measuring the quality of EPA, we apply them in validating the EPAs developed in this study and identifying strengths and weaknesses in the content of each EPA.

2.4   Entrustable Professional Activity as the basis for faculty development

Assessment requires competent faculty (Holmboe, 2011), and the rate-limiting step in competency-based medical education is faculty development. There is an ongoing interest in the concept of entrustable professional activities (EPAs) as a bridge between the competency framework to the workplace. It is defined as "a unit of professional practice, a task or responsibility to be entrusted to a trainee once sufficient specific competence is reached to allow for unsupervised practice" (ten Cate 2017). Dewey et al. (2017) triggered the question of using EPAs for teachers in medical education. Their paper explores the concept of EPAs for specific aspects of clinical teachers' roles. As the curriculum heads towards an EPA based one, we believe the same principles would apply and facilitate faculty learning and training to perform a specific unit of practice, i.e., assessment.

Iqbal and Al-Eraky (2019) discussed the use of EPA in assessing teaching competencies and evaluating the effectiveness of a faculty development programme. They advocate assigning the levels of expertise to participants (novice to experts). Once mastery is achieved in a specific task, the participant may be entrusted to perform that task independently. The same concept can be applied for remediation purposes when a participant did not perform optimally. Benefits of EPA-based faculty development programmes include well planned, structured and outcome-based training activities. It

supports the assessment of competence levels, transfer of training into the workplace, facultymembers' confidence, competent faculty (ensuring institutional development), and promotion of professional development of faculty members. The authors also described barriers in designing and implementing such programmes due to resource limitations: lack of expertise, time and resistance from faculty members. We believe the benefits of an EPA based faculty development programme outweigh the barriers in its implementation.

Gruppen et al. (2016) experimented with the concept of EPA when developing their competency-based education (CBE) curriculum for the University of Michigan Master of Health Professions Education. This CBE aims to encourage personalised learning that emphasises the development of skills and abilities rather than a time-based outcome. They identified twelve educational competencies and 20 EPAs that directed learning and assessment in the programme. They also defined the roles of faculty members to facilitate personalised learning. They found that EPAs and competencies can provide an alternative solution to traditional courses and facilitates assessment.

Iqbal et al. (2020) developed an EPA framework for small group facilitators through a participatory design approach. Using EPA was based on the understanding that EPAs can serve as a framework to plan, structure and evaluate faculty development programmes. First, they developed a framework of nine tasks and 12 competencies for longitudinal training and entrustment of small group facilitators. Then, they reframed the nine tasks into three EPAs with nine competencies. These EPAs were described to facilitate operationalisation and were validated by 31 international medical educationalists.

Van Dam et al. (2021) utilised the EPA framework to redefine and describe the features for high-quality bedside teaching via triangulation of literature review, focus

group discussion and open-ended email questionnaires. They included participants who performed bedside teaching at various stages of their education career. The findings from this research were presented according to ten Cate's recommendation of EPA description and will be used to guide faculty development and certification of the clinical teacher.

2.5   Document review as data collection method and analysis

A document is written or produced by someone in a specific context for a specific purpose. However, these contents do not necessarily become facts or truths (Matthews & Ross, 2010). The advantages of document review include that it is readily available and informative and useful for data triangulation. In addition, it is helpful to provide information that is available for direct observation and questioning. Possible disadvantages or limits to its usefulness are changes in definition over time and administrative boundaries; and the possibility of error by the original authors' understanding of key concepts or ambiguity of the contents.

The data collected must be based on the research questions and operational definitions. Apart from the content, it should include the authors, typology (medium, origin and purpose), context, and impact. The content analysis may be quantitative, e.g., the frequency of a particular word or phrase mentioned or qualitative, e.g., ideas expressed, underlying approach.

We used official documents to identify legal requirements of assessment procedures and desirable examiner or assessor criteria. These requirements formed the basis of the needs assessment phase.

2.6   Focus group discussion

Focus group discussion was the method used to explore trainees' perspectives of examiners competencies. This discussion aims to understand the trainees' understanding of the examination process, their opinion regarding the examiners' competencies and the circumstances that led them to form these opinions (Stalmeijer et al., 2014, Matthews & Ross 2010). In addition, trainees' opinions on the usefulness of the examination, perceptions on the efficacy of exam methods, the format of the examination and requirements for examiner training (WHO Instructions for using Focus Group Discussion) are also explored. The findings from this FGD were used to formulate the second round of the Delphi questionnaire. The group was moderated by a facilitator unknown to the trainees to encourage and stimulate the participants to share their ideas (Stalmeijer et al., 2014). In addition, each group was homogenous to minimise power differentials amongst the participants, minimising the effect on group dynamics (Stalmeijer et al., 2014).

The advantages of FGD are that it allows exploration of ideas and theme generation (Matthews & Ross, 2010), generation of rich information and data in trainees' own words, and when used in combination with other methods, may enhance findings in helping to answer the research questions. In addition to the advantages above, we chose this method due to logistics reasons, as it was convenient and required shorter sessions than individual interviews.

2.7   Delphi methodology

To develop the EPA, we triangulate data from three sources: A review of current documents, a Delphi study and a focus group discussion.

According to Humphrey-Murto et al., 2017 the Delphi method is a systematic method to measure and develop consensus for problem-solving or recommendations from experts. It consists of conducting anonymous iterative questionnaires and collating expert feedback. The iterative process is repeated until a pre-determined number of rounds is reached, or consensus is reached. The Delphi method straddles qualitative and quantitative methodologies. We chose this method for this project due to its benefits of anonymity, avoidance of undue dominance of certain parties, and potential for including diverse participants. However, we are aware of the risk of diminishing response after a prolonged process and the limitation on discussion (Humpre-Murto et al., 2017).
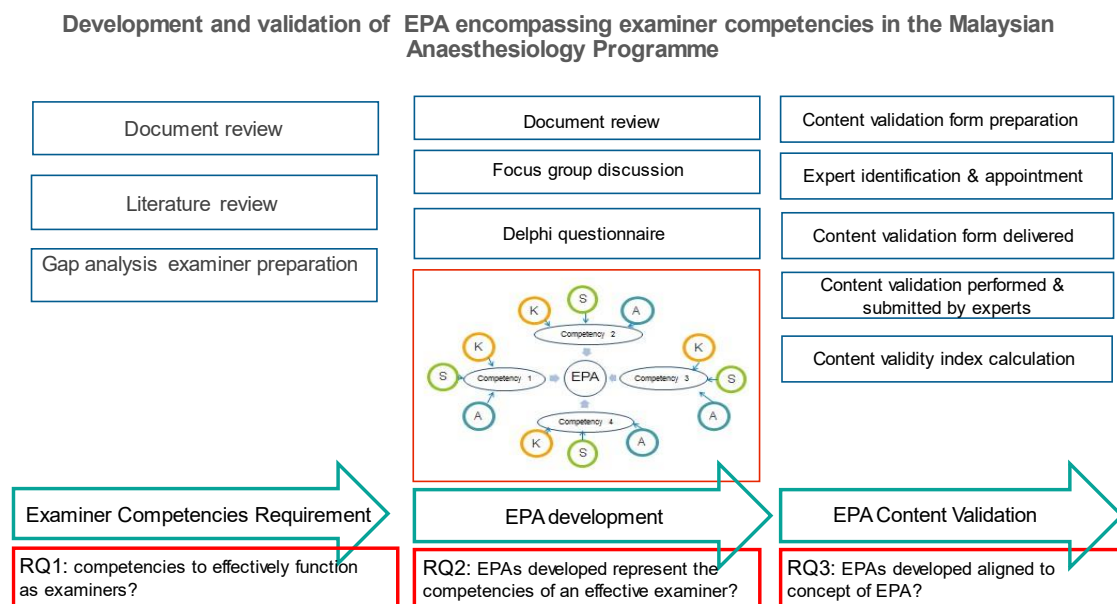
2.8   Conceptual framework

Figure 2.1   The conceptual framework shows how the three research stages (blue arrows) answer the research questions (red text boxes). The blue text boxes depict the methodology. The competencies of examiners, consisting of their knowledge, skills and attitudes, is constructed in the EPA framework (radial diagram).

CHAPTER 3

METHODOLOGY

3.1 Introduction

This research is a 3-staged mixed-method study to develop and validate a series of Primary Anaesthesiology Examiner Entrustable Professional Activities (EPAs). The subsequent sub-sections explain the details of each stage: problem identification and general needs assessment, EPA development and EPA validation. The researchers decided to choose this study design as it provides data sources from multiple stakeholders efficiently. In addition, the data collected through each stage and method were complementary to each other (Tavakol & Sanders, 2014). The study setting was at the Principal Investigator's home institution, Universiti Malaya.

The study population included examiners, trainers and trainees of the Malaysian Anaesthesiology programme.

The reference and target populations were Anaesthesiology programme trainers, examiners and trainees. Therefore, the sampling pool was Anaesthesiology programme trainers and examiners from Universiti Malaya, Universiti Sains Malaysia, Universiti Kebangsaan Malaysia, Universiti Putra Malaysia, Universiti Islam Antarabangsa, Universiti Teknologi MARA, Malaysia and Universiti Malaya Anaesthesiology trainees.

The sampling frame for examiners were examiners involved in the Anaesthesiology Primary examination; trainers are all trainers in the programme registered to the sampling pool, and trainees are Universiti Malaya Master of Anaesthesiology candidates who have passed their Primary examination.

The sub-sections below describe the sampling methods and criteria for inclusion and exclusion.

3.2   Stage 1: Problem identification and general needs assessment

This stage focused on problem identification and general needs assessment to identify examiner qualifications, experiences, expected professional behaviours, knowledge on exam content, process, basic principles of assessment, evidence and threats to validity. Data collection was done by document review.

This stage aimed to identify examiners tasks, competencies required to perform them and gaps in the current training.

The expected outputs of the stage were the qualifications of the examiner/ assessor: definition of criteria and competencies; the knowledge, skills and attitude required to achieve these competencies and the training required to achieve these competencies.

Data was collected from documents with various typologies (medium, origin, and purpose of the documents) and included the content and contexts of each document. Key questions revolved around the tasks to be performed by examiners for written and oral examinations, the qualifications and competencies required to perform them, the training required to perform these tasks, the role of assessors, and the examiner training for written and viva examinations, and clinical examination.

Documents reviewed included documents by licensing agencies (The Malaysian Qualification Agency documents and Malaysian Medical Council's Standards for Medical Specialists training on assessment standards, procedures and assessor training); examination related documents by Universiti Malaya; Conjoint Board of Anaesthesiology documents related to the Primary Examination (Terms of References,

Examiner Guides, Candidate Guides); examination related documents by other programmes and review of chapters in assessment textbooks (Yudkowsky and Downing) and Tekian and Norcini's chapter Faculty development in assessment: What the faculty need to know and do. In P.F. Wimmers & M. Mentkowski (eds)assessing competence in professional performance across disciplines and professions, innovation and change in professional education (13, pp 353-374).

Content analysis was by a qualitative approach, answering the questions above.

3.3   Stage 2: Development of EPAs

This stage focused on developing a series of EPAs for written and oral examiners using triangulation of data from document review (Stage 1), Focus Group Discussions (FGD) and Delphi questionnaires.

3.3.1   Focus group discussion (FGD)

This step aimed to gain insights from trainees in identifying the competent examiner's knowledge, skills, attitudes, and behaviours, identifying examiner training priorities, and relate these to the examination process and experience.

The expected output from this step was the description of the knowledge, skills and behaviours expected from the competent examiner and areas for examiner training.

The sampling method for this FGD was purposive sampling from Universiti Malaya Master of Anaesthesiology candidates who have passed their Primary examination. The estimated sample size was 25, according to Cresswell et al. (1998) and Bertaux et al. (1981), to be 5 to 25 and 15, respectively. All potential participants received personal WhatsApp (WhatsApp Inc. (Facebook, Inc,) (2020)) messages from the Principal Investigator requesting their voluntary participation. These messages were sent only once each to ensure participants were not pressured to participate.

The FGD sessions were conducted in groups of four to five; the expected number of participants to achieve saturation is 20-25. Saturation was achieved when no new information was achieved during the FGD.

A Medical Education lecturer from USM, who is unknown to the participants and their age, moderated the sessions. In addition, the Principal Investigator was not present during the sessions to prevent any influences on the participants' answers.

The researchers developed an interview guide with crucial questions via brainstorming, based on Rowan University, Trinity College of Arts and Sciences, Krueger (2002, University of Minnesota) and FirstWork's guides, used for the FGD. The questions revolved around the competencies expected of a Primary Examiner, the examination process, and the usefulness of the Primary examination and examiner training requirement. The guide also included a section on managing difficult situations, e.g., a dominating participant, different levels of participation and participants having a side conversation.

Steps taken to prevent bias in ensuring participant safety, anonymity and confidentiality included establishing ground rules concerning these issues at the beginning of each session. In addition, to ensure comfortable and optimal group dynamics, the participants were allowed to choose their sessions and groups. Despite this, each group still had different demographic profiles to provide different experiences and perspectives.

As the data collected were relatively small, thematic analysis was manual via MS Word 2016. The researchers chose thematic analysis as the analysis method because it allowed description and interpretation of data, and non-linear analysis, integrating context and content, and not requiring peer checking (Vaismoradi et al., 2013). Analysis was performed based on Vaismoradi et al. and Kiger and Varpio's 2020 AMEE Guide

No. 131 (Thematic analysis of qualitative data): data familiarisation with repeated reading of the FGD transcripts, codes identification and generation, followed by theme generation, review and definition. The final themes were reframed into ten Cate and Chen's A RICH framework (2020).

Transcribed data were read and re-read, initial ideas were highlighted. Next, features highlighted in each data set were collated, compared, and tabulated. Codes that overlapped were re-organised. Finally, data relevant for each code were tabulated.

Themes were constructed by mapping and re-organising collated codes. Finally, these themes were compared and re-organised according to the A RICH framework.

3.3.2   Delphi questionnaires

The participants in the Delphi exercise were redefined as experts. They were current and former examiners of the Conjoint Malaysian Anaesthesiology specialist training programme. They were knowledgeable, representative of the research interest and have practical experience in the examination process and system.

Consensus was considered achieved when 70% of the experts agree or strongly agree to include an item. Items where 70% of the experts disagree or strongly disagree, were discarded. Items that did not meet these criteria were sent to the next round of questionnaires to be re-ranked if still considered relevant by the research team. The second and third rounds of questionnaires included summary statistics and justification for their choices, where relevant.

The sampling method for Delphi experts was purposive sampling. The researchers identified experts from the pool of Anaesthesiology examiners involved in the Conjoint Primary Anaesthesiology examination from 2015 to 2019, from Universiti Malaya,

Universiti Sains Malaysia, Universiti Kebangsaan Malaysia, Universiti Putra Malaysia, Universiti Islam Antarabangsa, Malaysia.

The estimated sample size was 15 (12 or more, based on Humphrey-Murto et al.,2017) and 15 to 30 participants of five to 10 per category from a heterogeneous group (de Villiers et al., 2005). Thirty-three potential experts from the Primary Anaesthesiology examiner pool were identified. They were approached via emails obtained from the Department of Anaesthesiology, University Malaya's database and their respective Heads of Departments. Fourteen who agreed were recruited to form a Delphi panel. This panel comprised examiners from both subjects examined in the Primary Examination.

Online questionnaires for data collection were designed, sent and retrieved using a secure programme (REDCap: Research Electronic Data Capture software), available in Universiti Malaya, accessible only to the Principal Investigator. Panel members and their responses were known to the Principal Investigators, their input was available to all researchers, and they remained anonymous to each other.

Three rounds of questionnaires to be answered by the Delphi panel were sent. The questionnaires were planned to be sent one month apart; each panel member had two weeks to respond. A reminder message was sent to each expert once every three days, three times. The survey remained open until all experts have submitted their responses. The first questionnaire was sent in January 2020, second in December 2020 and third in January 2021 due to expert and researcher unavailability during the COVID pandemic.

Qualitative data were analysed using manual thematic analysis as Kiger and Varpio's 2020 AMEE Guide No. 131. The organisation and final themes were reframed into ten Cate and Chen's A RICH framework.

Quantitative data was exported from REDCap into Microsoft Excel. Descriptive statistics were used to summarise the socio-demographic characteristics of the expert panel members. Items rated 3 and 4 were re-coded as agreement/ relevant, and items rated 1 and 2 were re-coded as disagreement/ not relevant. Consensus was calculated based on the number of experts agreement of each item and is considered reached as per our consensus operational definition. This data will be presented as a percentage in the next chapter.

3.3.2(a)   Round one

The purpose of this round was to identify the gaps in examiner training and preparation and to identify the examiners' perspectives on characteristics of the competent examiner, including the knowledge, skills and attitude required.

The expected output of this round was to identify the definition of a competent examiner, the knowledge, skills and attitude required to be competent, to identify examiner training needs.

A three-sectioned questionnaire was constructed based on the data collected in Stage 1. Section 1 is the experts' profile. Section 2, an open-ended questionnaire, revolved around the experts' perception of the function of the examination, the preparation and training required to be an effective examiner, their description of a competent or effective examiner, and the knowledge, skills and attitudes necessary to perform as an examiner in the Primary Anaesthesiology Examination. Finally, in Section 3, experts were asked to provide their opinion on the relevant knowledge, skills and attitude required to be considered a competent examiner using a 4-point Likert scale (1: not relevant, 2: somewhat relevant, 3: quite relevant, 4: highly relevant).

3.3.2(b) Round two

This round aimed to identify the tasks to be performed by examiners, the description of the tasks, and the competencies required to perform them.

The expected outcomes of this round were to select tasks that meet consensus, describe these tasks and determine the competencies required to perform them.

A three-sectioned questionnaire was constructed based on Stage 1 and thematic analysis of the FGD and data analysis of Round one and further literature review. In Section 1, the experts rated the relevance of tasks to be performed by an examiner in the Primary Anaesthesiology examination using a 4-point Likert scale (1: not a relevant task for an examiner to perform, 2: somewhat a relevant task for an examiner to perform, 3: quite a relevant task for an examiner to perform, 4: a highly relevant task for an examiner to perform). In Section 2, the experts, provided with a list of descriptions of tasks from Section 1, rated the relevance of these descriptions to the tasks provided, again using a 4-point Likert scale. Finally, in Section 3, experts rated the relevance of task-related and professional competencies to perform the tasks presented in Section 1. Each section included open-ended questions for experts to suggest additional tasks, descriptions and competencies.

3.3.2(c) Round three

This round aimed to clarify the findings in Round 2, i.e., to embed elements of tasks that were not discrete into the other tasks, adding descriptions to relevant tasks, re-rank items that did not meet consensus; and identify activities required for credentialing (entrustment) and period of validity of credential.

The expected outcomes of this round were a final list of tasks, along with their descriptions, activities that would support entrustment and the period of entrustment validity.