

# Determinants of clarification research purpose in medical education research: A systematic review

Wee Shiong Lim<sup>1,2</sup>, Kar Mun Tham<sup>3</sup>, Fadzli Baharom Adzahar, Han Yee Neo<sup>4</sup>, Wei Chin Wong<sup>1</sup> & Charlotte Ringsted<sup>5</sup>

<sup>1</sup>Department of Geriatric Medicine, Tan Tock Seng Hospital, Singapore; <sup>2</sup>Health Outcomes and Medical Education Research, National Healthcare Group, Singapore; <sup>3</sup>Yong Loo Lin School of Medicine, National University of Singapore, Singapore; <sup>4</sup>Department of Palliative Medicine, Tan Tock Seng Hospital, Singapore; <sup>5</sup>Centre for Health Science Education, Faculty of Health, Aarhus University, Denmark

## Abstract

**Background:** Medical education research should aspire to illuminate the field beyond description (“What was done?”) and justification (“Did it work?”) research purposes to clarification studies that address “Why or how did it work?” questions. We aim to determine the frequency of research purpose in both experimental and non-experimental studies, and ascertain the predictors of clarification purpose among medical education studies presented at the 2012 Asia Pacific Medical Education Conference (APMEC).

**Methods:** We conducted a systematic review of all eligible original research abstracts from APMEC 2012. Abstracts were classified as descriptive, justification or clarification using the framework of Cook 2008. We collected data on research approach (Ringsted et al., 2011), Kirkpatrick’s learner outcomes, statement of study aims, presentation category, study topic, professional group, and number of institutions involved. Significant variables from bivariate analysis were included in logistic regression analyses to ascertain the determinants of clarification studies.

**Results:** Our final sample comprised 186 abstracts. Description purpose was the most common (65.6%), followed by justification (21.5%) and clarification (12.9%). Clarification studies were more common in non-experimental than experimental studies (18.3% vs 7.5%). In multivariate analyses, the presence of a clear study aim (OR: 5.33, 95% CI 1.17-24.38) and non-descriptive research approach (OR: 4.70, 95% CI 1.50-14.71) but not higher Kirkpatrick’s outcome levels predicted clarification studies.

**Conclusion:** Only one-eighth of studies have a clarification research purpose. A clear study aim and non-descriptive research approach each confers a five-fold greater likelihood of a clarification purpose, and are potentially remediable areas to advance medical education research in the Asia-Pacific.

**Keywords:** *Research Purpose; Research Approach; Medical Education Research; Asia-Pacific*

## Practice Highlights

- The hallmark of clarification research is the presence of a conceptual framework or theory that can be affirmed or refuted by the study results.
- We should aspire towards clarification studies that address “Why or how did it work?” questions.
- Only one-eighth of studies have a clarification research purpose.
- A clear study aim and non-descriptive research approach are potentially remediable areas to promote clarification studies.

## I. INTRODUCTION

There is much debate about how to ensure that medical education research is not perceived as the poor relation of Biomedical research (Shea, Arnold, & Mann, 2004; Baernstein, Liss, Carney, & Elmore, 2007; Todres, Stephenson, & Jones, 2007). Some have proposed that if medical education were to fulfil its research potential and enjoy academic legitimacy, the discipline must develop a clearer sense of purpose and more rigorously follow the scientific line of enquiry characterized by a cycle of observation; formulation of a model or hypothesis to explain the results; prediction based on the model or hypothesis; and testing of the hypothesis (Cook, Bordage, & Schmidt, 2008a; Bordage, 2009). In particular, medical educators often focus on the first step (observation) and the last step (testing), but omit the intermediate steps (model formulation or theory building, and prediction), and perhaps more importantly, fail to maintain the cycle by building upon previous results. Some authorities attribute this lack of a conceptual framework as a major reason for the paucity of impactful research questions that can illuminate and magnify the body of knowledge to advance the field of medical education (Albert, Hodges, & Regehr, 2007; Cook et al., 2008b; Eva & Lingard, 2008).

Conceptual frameworks represent ways of thinking about an idea, problem, or phenomenon by relating to theories, models, evidence-based best practices or hypotheses (Rees & Monrouxe, 2010; Gibbs, Durning, & Van der Vleuten, 2011). The framework assists in formulating the research question, choosing an appropriate study design, and determining appropriate outcomes to answer the research question. Situating the research question within a conceptual framework elevates and transforms the research purpose from a study which is focused on local issues, into a clarification study of general interest by engendering generalizable knowledge which can be transferable to new settings and future research (Bordage, 2009; Bunniss & Kelly, 2010). Conceptual frameworks are also essential in interpreting the results. Their interdependent relationship is underscored by the fact that results are interpreted in light of the existing theories and conversely, existing boundaries of the theoretical framework may limit interpretation of the findings (Wong, 2016). For instance, in the field of observation-based assessments, there is a gradual theoretical shift from a more psychometric (based on large numbers of random elements) to a more expert judgement (based on fewer observation of well-informed opinions) framework (Hodges, 2013)

### A. Cook's Framework of Research Purpose

To better delineate this problem, Cook et al. (2008a) proposed a framework for classifying the purposes of research, namely description, justification and

clarification. *Description studies* focus on the first step in the scientific method (observation) by addressing the question: "What was done?" *Justification studies* focus on the last step in the scientific method (testing) by asking: "Did it work?" However, without prior model formulation and prediction, the results may have limited application to future research or practice. In contrast, *clarification studies* seek to answer the question: "Why or how did it work?" The hallmark of clarification research is the presence of a conceptual framework that can be affirmed or refuted by the study results (Cook et al., 2008a; Ringsted, Hodges, & Scherpbier, 2011). Such research is often performed using classic experiments, but non-experimental methods such as correlation research, comparisons among naturally occurring groups, and qualitative research, are also applicable (Shea et al., 2004; Cook et al., 2008a).

Applying this framework in a systematic survey of 850 experimental and non-experimental studies on problem-based learning, Schmidt (2005) reported a paucity of clarification studies (7%) vis-à-vis description (64%) and justification (29%) studies. More recently, García-Durán et al. (2011) reported the predominance of description studies (92.8%) with very few justification (6.8%) studies and just one (0.4%) clarification study among research presentations at a medical education meeting in Mexico. These results are consistent with the seminal study of 105 articles describing education experiments in 6 major journals by Cook et al. (2008a), which noted that clarification studies were uncommon (12%) relative to justification (72%) and description (16%) studies. In this study, inter-rater agreement for these classifications was only moderate at 0.48, with disagreements largely occurring in the classification of less clear-cut single-group pre-test/post-test studies; this discrepancy has since been clarified in the revised definitions.

### B. Gaps in Current Knowledge

Cook et al. (2008a) proposed expanding the use of their framework beyond the limited genre of experimental studies to incorporate non-experimental study designs. This input is sorely needed, as studies with a purely descriptive design (which may not qualify as research by some authorities) historically constituted a significant proportion of the literature in medical education (Reed et al., 2008; García-Durán et al., 2011). There is a unifying call for the use of stronger study designs in the field beyond cross-sectional descriptive approaches to enhance the quality of medical education research (Gruppen, 2007; Colliver & McGaghie, 2008).

The opportunity to extend the study of research purpose beyond experimental studies, was afforded by the research compass framework described by Ringsted et al. (2011). Core to the model is the conceptual framework,

which is central to any research approach taken. The compass depicts four main quadrants of research approaches in the conduct of medical education research: (1) Explorative studies, aimed at modelling by seeking to identify and explain elements of phenomena and their relationships; (2) Experimental studies, with the main aim of justification to define appropriate interventions and outcomes; (3) Observational studies, aimed at predicting outcomes by the study of natural or static groups of people; and (4) Translational studies, which focus on implementing knowledge and findings from research in complex real-life settings.

Furthermore, predictors of a clarification research purpose in medical education scholarship have hitherto not been studied. Factors that are associated with better quality of medical education research include number of institutions studied (Reed et al., 2007; Reed et al., 2008), outcomes based on the widely-used hierarchy of Kirkpatrick (1967), and the presence of a clear statement of study intent (Cook et al., 2008b). However, the association between these factors with research purpose has not been previously examined.

### *C. Aims and hypothesis*

In recent years, there is a surge of interest in research scholarship in medical education in the Asia-Pacific region. Concomitantly, regional forums such as the Asia-Pacific Medical Education Conference (APMEC) have emerged for the sharing of medical education research, along with ongoing discussions about how to propel the field forwards in conducting meaningful research that can inform educational practice (Gwee, Samarasekera, & Chong, 2012). Determining the prevalence of research purpose of medical education studies from the Asia-Pacific region would be of immediate relevance in ascertaining whether there is a similar lack of clarification studies and research approaches beyond descriptive study designs.

Building upon the earlier work of Cook et al. (2008a) in experimental studies, we developed an empirical operational model that combined the frameworks of Cook and Ringsted to broaden the evaluation of research purpose to include non-experimental studies. The objectives of our study are: (1) to determine the frequency of research purpose in both experimental and non-experimental studies, and (2) to ascertain the predictors of a clarification research purpose among original research abstracts presented at APMEC 2012. In light of the findings in earlier studies of the relative paucity of clarification studies (Schmidt 2005; Cook et al., 2008a; García-Durán et al., 2011), we hypothesize that the proportion of clarification studies would likewise be comparatively lower.

## II. METHODS

### *A. Study setting*

This review drew from research abstracts submitted to APMEC 2012. This study is part of a larger piece of work that aims to contribute to the research agenda in the Asia-Pacific region by determining the trends in research purpose and approach in the last 5 years (2008 to 2012). The APMEC is an established regional conference held in Singapore that serves as an accessible “clearinghouse” providing a timely and comprehensive snapshot of research in the Asia Pacific region. The theme for the 9<sup>th</sup> APMEC in 2012 was “Towards transformative education for healthcare professionals in the 21<sup>st</sup> century – nurturing lifelong habits of mind, behaviour and action.” The National Healthcare Group Institutional Review Board deemed this study exempt from review.

### *B. Study eligibility*

All original research abstracts from APMEC 2012 were considered. Original research was defined as an educational intervention or trial; implementation of evidence based practice or guidelines; curriculum evaluation with subjective or objective outcomes; evaluation of an educational instrument or tool; qualitative research; and systematic reviews. We excluded abstracts from plenary lectures, workshops, special interest group meetings and discussions. Among 210 eligible abstracts, we excluded 24 that were not original research, thus yielding a final sample of 186 abstracts.

### *C. Data extraction*

We performed a pilot study using randomly selected abstracts from APMEC 2011 conference in order to fine-tune definitions of study variables and to refine the data collection form. Four reviewers were involved in data collection. After training and harmonization in the pilot phase, the four raters achieved good to excellent agreement in the coding (overall percentage agreement: 80-87%; ACI-statistic: 0.73 – 0.82) (Gwet, 1991). For the study proper, each abstract was rated independently and in duplicate. Disagreements were resolved by discussion, and if no consensus was reached, via adjudication by a third independent reviewer.

### *D. Data collection*

#### *1) Study design*

We classified abstracts into 2 broad categories based upon the “research compass” framework proposed by Ringsted et al. (2011): (1) Experimental, defined as any study in which researchers manipulated a variable (also known as the treatment, intervention or independent variable) to evaluate its impact on other (dependent) variables, including evaluation studies with experimental designs (Fraenkel & Wallen, 2003); and (2) Non-experimental, defined as all other studies that do not meet criteria for (1).

Studies using mixed methods (for instance, an experimental design with a qualitative component) were classified according to the methodology that was deemed to be predominant.

Experimental studies were further sub-classified as experimental, quasi-experimental or pre-experimental according to established hierarchies of research designs (Creswell, 2013). We defined experimental studies by the presence of randomization; examples included factorial design, crossover design and randomized controlled trials. In contrast, for quasi-experimental studies, experimental and control groups were selected without random assignment of participants (Colliver & McGaghie, 2008). Pre-experimental studies, namely single group pre-post and post-only designs, did not have a control group for comparison.

Non-experimental studies were further sub-classified as descriptive, qualitative, psychometric, observational (comprising associational, case-control and cohort studies), and translational. Descriptive studies typically provide descriptions of phenomena, new initiatives or activities, such as curriculum design, instructional methods, assessment formats, and evaluation strategies (Ringsted, Hodges, & Scherpbier, 2011). Because pure descriptive study designs may not strictly qualify as research by some authorities, they are ranked by default as lowest in the hierarchy of study designs (Crites et al., 2014). Hence, when two study designs were identified within the same study with one being descriptive, we coded based upon the “higher” non-descriptive study design.

## 2) Research purpose

Research purpose is classified as description, justification or clarification based upon modified definitions of the Cook framework (Table 1). We further sub-classified clarification studies into whether they relate to theory, model/evidence-based practices, or hypothesis. Because the original definitions pertain only to experimental

studies, several modifications were necessary in order to accommodate non-experimental studies in the integrated frameworks of Cook and Ringsted (Table 2). In the process, we were mindful to adhere to the original spirit of the definitions as far as possible. For instance, even though the original definition of justification studies merited a comparison group, we waived this requirement for good quality psychometric studies for which we deemed that there was sufficient rigor in the measures of validity and reliability to answer the question “Does this assessment tool work?” This was motivated by the intention to not “penalize” these studies and spuriously inflate the proportion of description studies in this category. Many validation studies of assessment tools often involve a single group design to determine whether a tool works via implicit comparison with an unknown ‘good enough’ criterion. The dominance of the psychometric views on assessment would also mean that many assessment studies are unlikely to have included an explicit statement of the underlying theoretical (Classical Test Theory, G-theory or Item Response Theory) framework.

Similarly, prompted by the observation that certain categories of approaches would be incongruent with a justification design, such as qualitative and observational studies, we delinked where appropriate the hierarchy of purpose from description to justification. Thus, a well-conducted observational study underpinned by a conceptual framework that explains the relationship between independent and dependent variables, would still qualify as a clarification study. Lastly, in response to difficulties encountered in coding during the pilot phase, we further modified the definition of clarification studies to specify the presence of a conceptual framework that fulfilled 3 crucial elements: 1) A theory, model, or hypothesis that asks “Why or how does it work?” 2) Transferability to new settings and future research; and 3) Confirmed or refuted by the results and/or conclusions of the study.

<b>A) Description</b>	Describes what was done or presents a new conceptual model. Asks: “What was done?” There is no comparison group. May be a description without assessment of outcomes, or a “single-shot case study” (single-group, post-test only experiment).
<b>B) Justification</b>	Makes comparison with another intervention with intent of showing that 1 intervention is better than (or as good as) another. Asks: “Did it work?” (Did the intervention achieve the intended outcome?). Any experimental study with a control group or single-group with pre-post intervention assessment would qualify. Good quality psychometric studies with measures of validity and reliability are exempt from the need for a comparison group, since justification that the tool “works” typically does not involve a comparison group in these studies. Justification studies generally lack a conceptual framework or model that can be confirmed or refuted based on results of the study.
<b>C) Clarification</b>	Clarifies the processes that underlie observed effects. Asks: “Why or how did it work?” Often a controlled experiment, but could also use a case-control, cohort or cross-sectional research design. Much qualitative research also falls into this category. Its hallmark is the presence of a conceptual framework that can be transferable to new settings and future research, and which can be confirmed or refuted by the results and/or conclusions of the study. Further sub-classified into whether the conceptual framework pertains to a theory, model/evidence-based practice, or hypothesis.

Table 1: Definitions of research purposes (modified from Cook et al.<sup>5</sup>)

Study Design*	Descriptive	Justification	Clarification
<b>(I) Experimental</b>			
- Experimental	√	√	√
- Quasi-experimental	√	√ (no randomization)	√
- Pre-experimental	√	√ (pre-post only)	√
<b>(II) Non-experimental</b>			
<b>(1) Explorative</b>			
- Descriptive	√	X	+/-
- Qualitative	√	X	√
- Psychometric	√	√ (validity, reliability)	√
<b>(2) Observation</b>			
- Associative	√	X	√
- Case control	√	X	√
- Cohort	√	X	√
<b>(3) Translational</b>			
- Knowledge creation			
Narrative	√	+/-	X
Quantitative review	√	√	+/-
Realist review	√	√	√
- Implementation	√	√	√
- Efficiency	√	√	√

Table 2: Conceptual framework for possible classifications of research purpose when analyzed by research approach  
 \*Modified from: Ringsted, C., Hodges, B., & Scherpbier, A. (2011). *Medical Teacher*, 33(9), 695-709.

### 3) Other variables

We extracted data on other variables which may affect the quality of medical education research. These included presentation category, topic of medical education addressed, professional group being studied, country of the study population, number of institutions involved, Kirkpatrick's learner outcomes (if applicable), and statement of study intent. We measured learner outcomes on 4 levels based upon Kirkpatrick's expanded outcomes typology, namely learner reactions (level 1), modification of attitudes/perceptions (level 2a), modification of knowledge/skills (level 2b), behavioural change (level 3), change in organizational practice (level 4a) and benefits to patients or healthcare outcome (level 4b) (Kirkpatrick, 1967; Reeves, Boet, Zierler, & Kitto, 2015). If a study reported more than one outcome, the rating for the highest-level outcome was recorded, regardless of whether this was a primary or secondary outcome. Although the validity of hierarchical application of Kirkpatrick's levels as a standard critical appraisal tool has been questioned, it still remains widely used in assessing the impact of interventions in medical education (Yardley & Donan, 2012). The research question is arguably the most important part of any scholarly activity

and is framed as a statement of study intent often in the form of a purpose, objective, goal, aim or hypothesis (Cook et al., 2008b). We therefore collected data on whether there is an explicit statement of study intent, and if present, its quality as judged by correct location in the aims section; representation of study goals as opposed to mere stating of educational objectives; and completeness of information (i.e. whether any important objective was omitted).

### E. Data Analysis

Results were summarized using descriptive statistics. Preplanned subgroup analyses were conducted with Chi-square test or Fischer's exact test using research purpose (description, justification or clarification) as the dependent variable. Significant variables from bivariate analysis ( $P < .10$ ) were included in logistic regression analysis to ascertain which of these factors were associated with a clarification study purpose. All analyses were performed using SPSS for Windows version 17.0 (SPSS Inc, Chicago, Illinois, USA). Statistical tests were two-tailed and conducted at 5% level of significance.



### III. RESULTS

#### A. Abstract characteristics

Our sample of 186 original research abstracts comprised 38 (20.4%) oral communications, 20 (10.8%) best posters, and 128 (68.8%) poster presentations. All abstracts employed the AMRaC (Aims, Methods, Results and Conclusion) format, with the exception of two unstructured abstracts that were presented in the symposiums. The most common topics covered were in the areas of curriculum ( $N=54$ , 29.0%), teaching and learning ( $N=53$ , 28.5%), assessment ( $N=19$ , 10.2%), and e-learning ( $N=14$ , 7.5%). Besides Singapore ( $N=62$ , 33.3%), there was a good mix of abstracts from other countries in the South-East Asian region such as Malaysia, Indonesia, Thailand, Philippines and Myanmar ( $N=19$ , 10.2%), and other parts of Asia ( $N=88$ , 47.3%). Most of the studies involved a single institution ( $N=169$ , 90.9%). Kirkpatrick's learner outcomes were applicable in approximately half ( $N=94$ , 50.5%) of the abstracts, with level one (satisfaction, attitudes and opinions of the

learners) accounting for 50 (53.2%) of eligible outcomes, followed by knowledge/skills ( $N=32$ , 34.0%). An explicit statement of study intent was absent in 29 (15.6%) of abstracts. Among the remaining abstracts with an aims statement, 8 (4.3%) were incorrectly sited in the methods sections, 20 (10.8%) stated educational objectives instead of study goals, and 10 (5.4%) were incomplete.

#### B. Prevalence of research purpose (Table 3)

Description research purpose was the most common ( $N=122$ , 65.6%), followed by justification ( $N=40$ , 21.5%) and clarification ( $N=24$ , 12.9%). The majority of clarification studies pertain to models ( $N=20$ , 83.3%), with the remainder involving theory ( $N=3$ , 12.5%) or hypothesis ( $N=1$ , 4.2%). The prevalence of clarification studies was higher in non-experimental ( $N=17$ , 18.3%) compared with experimental ( $N=7$ , 7.5%) studies. Conversely, for justification studies, the prevalence is higher for experimental ( $N=36$ , 38.7%) compared with non-experimental ( $N=4$ , 4.3%) studies.

Study	N	Nature of abstracts	Descriptive N(%)	Justification N(%)	Clarification N(%)
Lim et al, 2016	186	APMEC 2012 conference abstracts, not limited to particular study type	122 (65.6)	40 (21.5)	24 (12.9)*
Experimental	93		50 (53.8)	36 (38.7)	7 (7.5)
Non-experimental	93		72 (77.4)	4 (4.3)	17 (18.3)
Schmidt, 2005	850	Studies on problem-based learning, not limited to particular study type	543 (63.9)	248 (29.2)	59 (6.9)
Cook et al, 2008	105	Experimental studies from 6 major journals published in 2003-4	17 (16.2)	75 (71.4)	13 (12.4)
Garcia-Duran et al, 2011	265	UNAM 2008 and 2010 conference abstracts, not limited to particular study type	246 (92.8)	18 (6.8)	1 (0.4)

APMEC: Asia-Pacific Medical Education Conference; UNAM: Universidad Nacional Autónoma de México

\*Comprises 83.3% Models, 12.5% theory and 4.2% hypothesis.

Table 3. Comparison of research purpose among various studies

#### C. Relationship of variables with research purpose (Table 4)

There was no significant association between research purpose with research category, professional group, country of study, and number of institutions (Table 4). Learner outcomes of Kirkpatrick's level 2 and above were more likely to have a justification or clarification research purpose ( $\chi^2 [4, N=186] = 67.12, p<.001$ ), as were studies with a clear statement of study objectives ( $\chi^2 [2, N=186] = 10.51, p=.005$ ). Experimental studies were less likely than non-experimental to involve a description purpose ( $\chi^2 [2, N=186] = 29.26, p<.001$ ) even though the frequency of clarification studies was comparatively lower (7.5% vs 18.3%). Non-descriptive studies were more likely to have a justification or clarification purpose ( $\chi^2 [2, N=186] = 71.70, p<.001$ ).

#### D. Logistic Regression (Table 5)

We included in the model three independent variables (statement of study intent, Kirkpatrick's outcome levels and descriptive research approach) which were significant in bivariate analysis. Experimental design was not included due to multicollinearity resulting from high correlation with Kirkpatrick's levels. The Hosmer-Lemeshow test was non-significant ( $\chi^2 [5, N=186] = 1.78, p=.881$ ), indicating goodness of fit of the final model. The presence of a clear study aim [Odds ratio (95% CI) = 5.33(1.17 – 24.38)] and non-descriptive research approach [Odds ratio (95% CI) = 4.70(1.50 – 14.71)] but not higher Kirkpatrick's outcome levels, independently predicted a clarification research purpose.

Characteristic	Description	N (%)	Justification	N (%)	Clarification	N (%)	P
Study category							.765
Poster	88 (68.8)		25 (19.5)		15 (11.7)		
Best Poster	12 (60.0)		5 (25.0)		3 (15.0)		
Orals	22 (57.9)		10 (26.3)		6 (15.8)		
Professional Group							.143
Postgraduate Medical	35 (74.5)		9 (19.1)		3 (6.4)		
Undergraduate Medical, Clinical	49 (62.0)		18 (22.8)		12 (15.2)		
Undergraduate Medical, Basic Science	20 (64.5)		8 (25.8)		3 (9.7)		
Nursing	7 (53.8)		1 (7.7)		5 (38.5)		
Allied Health	2 (40.0)		2 (40.0)		1 (20.0)		
Country of study							.492
Singapore	39 (62.9)		16 (25.8)		7 (11.3)		
South-East Asia, excluding Singapore	11 (57.9)		6 (31.6)		2 (10.5)		
Asia, excluding South-East Asia	61 (69.3)		14 (15.9)		13 (14.8)		
Europe	4 (44.4)		4 (44.4)		1 (11.1)		
North America	3 (100)		0 (0)		0 (0)		
Number of institutions studied							.857
1	111 (65.7)		37 (21.9)		21 (12.4)		
2	4 (57.1)		2 (28.6)		1 (14.3)		
>2	7 (70.0)		1 (10.0)		2 (20.0)		
Kirkpatrick's learner outcomes							<.001
Not applicable	73 (79.3)		5 (5.4)		14 (15.2)		
Kirkpatrick's level 1	40 (80.0)		7 (14.0)		3 (6.0)		
Kirkpatrick's level 2 and above	9 (20.5)		28 (63.6)		7 (15.9)		
Aims statement							.005
Absent or unclear	52 (77.6)		13 (19.4)		2 (3.0)		
Present, clear aims	70 (58.8)		27 (22.7)		22 (18.5)		
Experimental study							<.001
Yes	50 (53.8)		36 (38.7)		7 (7.5)		
No	72 (77.4)		4 (4.3)		17 (18.3)		
Descriptive study							<.001
Yes	94 (92.2)		3 (2.9)		5 (4.9)		
No	28 (33.3)		37 (44.0)		19 (22.6)		

Table 4. Relationship of variables with research purpose

	$\beta$	S.E.	Wald	P	Odds ratio (95% CI)
Clear study aims	1.67	.78	4.66	.031*	5.33 (1.17 – 24.38)
No outcomes <sup>^</sup>	-.17	.73	.06	.815	1.19 (0.29 – 4.94)
K2 outcomes and above <sup>^</sup>	-.10	.82	.02	.900	0.90 (0.18 – 4.52)
Non-Descriptive study	1.55	.58	7.09	.008**	4.70 (1.50 – 14.71)

Table 5. Logistic regression predicting likelihood of clarification research purpose

\* $P < .05$ ; \*\* $P < .01$

<sup>^</sup>Reference group: Kirkpatrick level one outcomes

Nagelkerke R square: 0.197

#### IV. DISCUSSION

The seminal study by Cook et al. (2008a) ushered in a series of studies that examined research quality through the lenses of research purpose. The underlying premise is that situating the research question and the accompanying study design, methods and analysis within a strong conceptual framework, facilitates the conduct of quality research that transcends the local context, allows transferability of findings, and can lead to new programmes of research (Bordage, 2009; Gill & Griffin,

2009; Beran, Kaba, Caird, & McLaughlin, 2014). By integrating the frameworks of Cook and Ringsted, this systematic review of APMEC 2012 original research abstracts contributes to this conversational turn by extending the Cook framework to include non-experimental studies. The strengths of our study include duplicate review at all stages; standardized definitions of coding categories; clear and detailed description of the methods/procedures involved; and high inter-rater reliability among the coders.

The distribution of research purpose in our study is broadly in line with earlier studies. Only around one-eighth of original research studies have a clarification research purpose. Around two-thirds of studies focused on “What was done?” description purposes which are not readily transferable beyond the immediate context of the individual study. Nonetheless, the relatively higher proportion of clarification purpose in our cohort vis-à-vis the 0.4-12.0% reported in earlier studies (Schmidt, 2005; Cook et al., 2008a; García-Durán et al., 2011) is reassuring (Table 3). Similar to Cook et al. (2008a), experimental studies account for the majority of justification studies. This is unsurprising, given the inherent nature of experimental studies in answering “Does it work?” question. Conversely, because research approaches such as qualitative and observational studies tend to ask “Why?” or “How?” questions, non-experimental designs have a higher proportion of clarification purpose compared with experimental studies (18.3% vs. 7.5%). To promote the further development of medical education scholarship in the Asia-Pacific region, we propose tapping upon regional initiatives like the Asia Pacific Medical Education Network (APME-Net), the Asian Medical Education Association, as well as regional journals such as *The Asia-Pacific Scholar*, to emphasize clarification studies that promote the wider application of theory which can be affirmed or refuted by the study results.

Our study also highlighted that a non-descriptive study design, in concert with a clear statement of study aims, each predicted a 5-fold increased odds of a clarification research purpose. Similar to developments in outcomes-based research within the field, we advocate a “design-balanced” approach whereby the best study design is one that best answers the research question within a given context (Lim, 2013). While descriptive study designs retain a role in the sharing of innovations and preliminary ideas, we should encourage the greater use of more rigorous non-descriptive study designs where appropriate (Ringsted et al., 2011). For experimental studies, quasi-experimental designs with a control group and true experimental designs characterized by randomization are less likely to overestimate effect size compared with single group pre-/post-test studies (Cook, Levinson, & Garside, 2011). In non-experimental studies, the plurality of non-descriptive approaches includes qualitative, psychometric, observational and translational research designs (Cheong et al., 2015; Ong et al., 2016).

Given the fundamental importance of the research question, it is disconcerting that around one-sixth of abstracts lack an explicit objective statement of study aims, whilst another one-fifth have an aims statement that is either incorrectly sited, confused with educational objectives, or incomplete. This may be indicative of poor

reporting quality, or more ominously, the lack of a clear research question underpinning the research study (Cook, 2016). A systematic review that evaluates the quality of abstracts of 110 experimental studies reported that essential elements of an informative abstract were often under-reported, especially in unstructured abstracts (Cook et al., 2007b). There is evidence that structured formats improve the quality of reporting of research abstracts (Taddio et al., 1994; Wong et al., 2005; Cook et al., 2007a). Reporting quality is positively associated with superior methodological quality (Cook et al., 2011), which in turn is associated with funding for medical education research (Reed et al., 2007). There is thus a case to be made for the consistent use of structured abstracts with relevant and thoughtful headings beyond the IMRaD (Introduction, Methods, Results and Discussion) format. Where relevant, separate headings for background and aims would neatly cater to the need for both literature review plus an explicit statement of study objectives. In addition, we propose a separate heading for conceptual framework or study hypothesis to spur the development of higher-order clarification studies, and a “limitations” heading to prompt researchers to think about more rigorous study designs and outcomes through consideration of the limitations of their current research (Cook et al., 2007b).

Some limitations are worth highlighting. Our research is based upon conference abstracts, which has a significant word constraint as compared to full-length papers. The validity of our findings is highly dependent on the reporting quality of the abstracts, such that the quality of a research (as judged by research purpose and approach) may be more a reflection of the reporting quality rather than the actual quality of research. Notwithstanding, evidence affirming the positive relationship between reporting and methodological quality lends credence to the validity of assessing conference abstracts as an indirect quality indicator of research (Cook et al., 2011). Moreover, our research involved fairly objective and essential elements of reporting such as study aims and outcomes. Secondly, we choose to focus on specific aspects of quality rather than a more comprehensive evaluation of methodological quality using validated scales such as the medical education research study quality instrument (MERSQI) (Reed et al., 2007). Our approach is compatible with the ongoing debate about what constitutes quality in medical education research, which highlights the pre-eminence of the conceptual framework in framing meaningful research questions that can advance the field (Eva, 2009; Monrouxe & Rees, 2009).

## V. CONCLUSIONS

Taken together, our study identified gaps that will, hopefully, serve to promote further discourse among



medical education scholars in the region about the purpose and approach of their research inquiry. To advance the research agenda of the Asia-Pacific region, we should tap upon regional platforms to promote clarification studies that employ more rigorous research approaches beyond cross-sectional descriptive study designs. The thoughtful use of structured abstracts to facilitate ancillary factors such as a clear aims statement that makes explicit the underlying conceptual framework, and study design, can also pave the way towards addressing gaps in research purpose and approach.

#### Notes on Contributor

W. S. Lim planned and executed the study, performed statistical analysis, and wrote the manuscript. K. M. Tham, F. B. Adzahar, H. Y. Neo, and W. C. Wong contributed to acquisition of data. C. Ringsted contributed to the conception and design of the study. All authors were involved in the critical revision of the paper and approved the final manuscript for publication.

#### Acknowledgements

The study was supported by an educational research grant from the National Healthcare Group Health Outcomes and Medical Education Research office.

#### Declaration of Interest

All authors have no potential conflicts of interest.

#### References

Albert, M., Hodges, B., & Regehr, G. (2007). Research in Medical Education: Balancing Service and Science. *Advances in Health Sciences Education, 12*(1), 103-115.

Baernstein, A., Liss, H. K., Carney, P. A., & Elmore, J. G. (2007). Trends in study methods used in undergraduate medical education research, 1969-2007. *JAMA, 298*(9), 1038-1045.

Beran, T. N., Kaba, A., Caird, J., & McLaughlin, K. (2014). The good and bad of group conformity: a call for a new programme of research in medical education. *Medical Education, 48*(9), 851-859.

Bordage, G. (2009). Conceptual frameworks to illuminate and magnify. *Medical Education, 43*(4), 312-319.

Bunniss, S., & Kelly, D. R. (2010). Research paradigms in medical education research. *Medical Education, 44*(4), 358-366.

Cheong, C.Y., Merchant, R.A., Ngiam, N.S.P. & Lim, W.S. (2015). Case-based simulated patient sessions in Mental-State Examination teaching. *Med Education, 49*(11), 1147-8.

Colliver, J. A., & McGaghie, W. C. (2008). The reputation of medical education research: quasi-experimentation and unresolved threats to validity. *Teaching and Learning in Medicine, 20*(2), 101-103.

Cook, D. A., Beckman, T. J., & Bordage, G. (2007a). Quality of reporting of experimental studies in medical education: a systematic review. *Medical Education, 41*(8), 737-745.

Cook, D. A., Beckman, T. J., & Bordage, G. (2007b). A systematic review of titles and abstracts of experimental studies in medical education: Many informative elements missing. *Medical Education, 41*(11), 1074-1081.

Cook, D. A., Bordage, G., & Schmidt, H. G. (2008a). Description, justification and clarification: a framework for classifying the purposes of research in medical education. *Medical Education, 42*(2), 128-133.

Cook, D. A., Bowen, J. L., Gerrity, M. S., Kalet, A. L., Kogan, J. R., Spickard, A., & Wayne, D. B. (2008b). Proposed standards for medical education submissions to the Journal of General Internal Medicine. *Journal of General Internal Medicine, 23*(7), 908-913.

Cook, D. A., Levinson, A. J., & Garside, S. (2011). Method and reporting quality in health professions education research: a systematic review. *Medical Education, 45*(3), 227-238.

Cook, D. A. (2016). Tips for a great review article: crossing methodological boundaries. *Medical Education, 50*(4), 384-387.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Crites, G. E., Gaines, J. K., Cottrell, S., Kalishman, S., Gusic, M., Mavis, B., & Durning, S. J. (2014). Medical education scholarship: An introductory guide: AMEE Guide No. 89. *Medical Teacher, 36*(8), 657-674.

Eva, K. W., & Lingard, L. (2008). What's next? A guiding question for educators engaged in educational research. *Medical Education, 42*(8), 752-754.

Eva, K. W. (2009). Broadening the debate about quality in medical education research. *Medical education, 43*(4), 294-296.

Fraenkel, J.R., & Wallen, N.E. (2003). *How to design and evaluate research in education*. New York: NY: McGraw-Hill.

García-Durán, R., Morales-López, S., Durante-Montiel, I., Jiménez, M., & Sánchez-Mendiola, M. (2011). Type of research papers in medical education meetings in Mexico: an observational study. Paper presented at the Annual Meeting of the Association for Medical Education in Europe, Vienna, Austria.

Gibbs, T., Durning, S., & Van Der Vleuten, C. (2011). Theories in medical education: Towards creating a union between educational practice and research traditions. *Medical teacher, 33*(3), 183-187.

Gill, D., & Griffin, A. E. (2009). Reframing medical education research: let's make the publishable meaningful and the meaningful publishable. *Medical Education, 43*(10), 933-935.

Gruppen, L.D. (2007). Improving medical education research. *Teaching and learning in medicine, 19*(4), 331-335.

Gwee, M. C., Samarasekera, D. D., & Chong, Y. (2012). APMEC 2013: In Celebration of Innovation and Scholarship in Medical and Health Professional Education. *Medical Education, 46*(s2), iii-iii.

Gwet, K. (1991). *Handbook of Inter-Rater Reliability*. STATAXIS Publishing Company.

- Hodges, B. (2013). Assessment in the post-psychometric era: Learning to love the subjective and collective. *Medical Teacher*, 35(7), 564-568.
- Kirkpatrick, D.L. (1967). Evaluation of training. In *Training and development handbook*, edited by R. L. Craig and L. R. Bittel. New York: McGraw-Hill. Pp. 87-112.
- Lim, W. S. (2013). More About the Focus on Outcomes Research in Medical Education. *Academic Medicine*, 88(8), 1052.
- Monrouxe, L.V., & Rees, C. E. (2009). Picking up the gauntlet: constructing medical education as a social science. *Medical Education*, 43(3), 196-198.
- Ong, Y.H., Lim, I., Tan, K.T., Chan, M. & Lim, W.S. (2016). Assessing Shared Leadership in Interprofessional Team Meetings: A Validation Study. *The Asia Pacific Scholar*, 1(1), 10-21.
- Reed, D. A., Cook, D.A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *JAMA*, 298(9), 1002-1009.
- Reed, D. A., Beckman, T. J., Wright, S. M., Levine, R. B., Kern, D. E., & Cook, D. A. (2008). Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's Medical Education Special Issue. *Journal of General Internal Medicine*, 23(7), 903-907.
- Rees, C. E., & Monrouxe, L. V. (2010). Theory in medical education research: how do we get there?. *Medical Education*, 44(4), 334-339.
- Ringsted, C., Hodges, B., & Scherpbier, A. (2011). 'The research compass': An introduction to research in medical education: AMEE Guide No. 56. *Medical Teacher*, 33(9), 695-709.
- Reeves, S., Boet, S., Zierler, B., & Kitto, S. (2015). Interprofessional Education and Practice Guide No. 3: Evaluating interprofessional education. *Journal of interprofessional care*, 29(4), 305-312. <https://doi.org/10.3109/13561820.2014.1003637>.
- Schmidt, H.G. (2005). Influence of research in practices in medical education: the case of problem-based learning. Paper presented at the Annual Meeting of the Association for Medical Education in Europe, Amsterdam, The Netherlands.
- Shea, J. A., Arnold, L., & Mann, K. V. (2004). A RIME perspective on the quality and relevance of current and future medical education research. *Academic Medicine*, 79(10), 931-938.
- Taddio, A., Pain, T., Fassos, F. F., Boon, H., Ilersich, A. L., & Einarson, T. R. (1994). Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *CMAJ: Canadian Medical Association Journal*, 150(10), 1611.
- Todres, M., Stephenson, A., & Jones, R. (2007). Medical education research remains the poor relation. *BMJ: British Medical Journal*, 335(7615), 333.
- Wong, G. (2016). Literature reviews in the health professions: It's all about the theory. *Medical education*, 50(4), 380-382.
- Wong, H. L., Truong, D., Mahamed, A., Davidian, C., Rana, Z., & Einarson, T. R. (2005). Quality of structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association: a 10-year follow-up study. *Current Medical Research and Opinion*, 21(4), 467-473.
- Yardley, S., & Dornan, T. (2012). Kirkpatrick's levels and education 'evidence'. *Medical Education*, 46(1), 97-106.

---

\*Wee Shiong Lim  
 Department of Geriatric Medicine  
 Tan Tock Seng Hospital  
 11 Jalan Tan Tock Seng, TTSH Annex 2, Level 3  
 Singapore 308433  
 Tel: +65 6359 6474  
 Email: Wee\_Shiong\_Lim@ttsh.com.sg