

A critical analysis of evidence about the impacts of faculty development in systematic reviews: A Systematic Rapid Evidence Assessment

Abstract

Introduction: Faculty Development (FD) encompasses a range of learning activities undertaken to prepare clinical teachers for their various roles as educators, leaders and scholars. This paper presents the findings of a Systematic Rapid Evidence Assessment (SREA) which aimed to investigate the impact of FD for clinical teachers.

Methods: We searched the published and grey literature for systematic reviews of FD to identify evidence to help inform judgements about which kinds of FD activities are effective for which groups of clinical teachers.

Results: After screening the literature we found seven systematic reviews met our inclusion criteria. Following a critical analysis of these reviews we found that they contained a number of limitations in relation to their use of review methods and reporting of results.

Discussion: On the basis of the analysis presented in this SREA we argue that the included reviews do not provide high quality evidence to effectively support decisions about choices of FD activities, even where the review authors made positive claims about impact. We go on to provide suggestions to improve the quality of systematic reviews in this area.

Key words: Faculty Development; healthcare educators; Systematic Rapid Evidence Assessment; critical analysis

Introduction

Different professional groups have different expectations regarding the preparation and accreditation of clinical teachers.¹ The different kinds of activities undertaken to prepare clinical teachers for their various roles is often referred to collectively as Faculty Development. In this paper, the term Faculty Development (FD) refers to the activities

undertaken by teachers of medical and healthcare professionals (from here on referred to as clinical teachers) to improve their knowledge, skills and behaviors as teachers, educators, leaders, managers, researchers and scholars of health professions education.²

FD activities vary in format, content, duration and setting. As a result, faculty developers have to make important choices when designing and delivering their courses. In these circumstances research evidence about the impacts of FD has a useful role in helping to inform their decisions about which kinds of activities or approaches may be effective for which groups of practitioners in which contexts.³ Systematic reviews of research about impacts can provide a comparatively efficient and rigorous source of evidence for this purpose.⁴

This paper presents the findings and discusses the implications for practice and research of our analysis of a number of systematic reviews of FD. These reviews were identified to investigate the following research question: ‘which kinds of FD lead to what outcomes for what kinds of clinical teacher?’

Methods

It is our view that the systematic reviews of FD that we identified and the studies included in them represent a reasonably comprehensive, view of the research literature on the impact of FD within health professions education. It is in this context that we present a summary of the methods we used.

We employed an approach which has been termed a Systematic Rapid Evidence Assessment (SREA).⁵⁻⁶ This approach follows the principles of systematic review, however, a number of strategies are used to accelerate a more rapid review process. Specific adaptations employed in this SREA included a selective data extraction process, a limited quality assessment process and a simple synthesis of included materials. In addition, we undertook a ‘review of reviews’ method, as opposed to reviewing primary research studies in this SREA. This approach has been used by John Hattie and colleagues in the ‘Visible Learning Project’⁷ and for the Teaching and Learning Toolkit produced by the Education Endowment Foundation.⁸ A SREA approach retains the advantages of transparency and rigor in the

review process when compared to non-systematic literature reviews, however, it reduces the time and resources required when undertaking a comprehensive systematic review of primary research. The specific procedures used in this SREA are detailed below.

Locating systematic reviews

We searched a range of sources including bibliographic databases (ASSIA, BEI, ERIC, MEDLINE, CINAHL, EMERALD); the Internet (Google Scholar) and systematic review organisations, specifically Cochrane EPOC Group, BEME and the Joanna Briggs Institute. A search string comprising a variety of synonyms for 'Faculty Development' were combined using the Boolean term 'OR' the results of which were then combined with 'AND' systematic review.

Selection criteria

The following selection criteria were applied to titles and abstracts of provisionally identified materials to identify relevant reviews: (1) reviews investigating the impact of accredited and non-accredited FD for clinical teachers/educators; (2) systematic reviews (i.e. reviews of research with recognized review methods reported); (3) reviews in English since 2000; and (4) reviews containing primary research studies from clinical educational systems which are broadly comparable to the UK in terms of the organization and delivery of health professions education (e.g. mainland Europe, North America, Australasia). A review had to meet all of these criteria to be included. At the second stage of screening (full papers), these selection criteria were re-applied. Figure 1 provides an overview of the searching and screening process.

INSERT FIGURE 1 ABOUT HERE

Data extraction, analysis and synthesis

Reviews that met the selection criteria were coded for relevant details about contexts, methods and review results/outcomes using a coding tool developed from previous systematic review work.⁹ A narrative synthesis of included reviews was completed. This narrative involved summarizing and combining the descriptive and contextual and outcome

information from the included reviews. The reported outcomes were categorized using a version of Kirkpatrick's outcomes typology adapted by Leslie et al.¹⁰ (Table 1).

INSERT TABLE 1 ABOUT HERE

As Table 1 shows, levels 1, 2a, 2b and 3 cover outcomes which measure impact on the clinical teachers who participated a FD intervention; whereas levels 4a, 4b and 4c measure impact on organizational practices (e.g. curriculum development), and/or the healthcare students of the participating clinical teachers and/or on the quality of patient care.

Quality assurance processes

The search strategy and inclusion criteria for this SREA were developed iteratively and piloted. Screening of abstracts was carried out by one of the authors with a random 10% sample screened by a second author, and any discrepancies discussed and then resolved. Data extraction and coding of key information from the included reviews was undertaken by one of the authors. A second author checked the extracted/coded information from the first two included reviews for consistency. Again, any discrepancies were discussed and then resolved.

Results

Faculty Development characteristics

Following our literature searches, we initially generated 2480 abstracts (Figure 1). After removing duplicates, the titles and abstracts of 2348 articles were screened using the four selection criteria described above. Seven systematic reviews of FD were included in the analysis.¹⁰⁻¹⁶ The largest proportion of primary studies in these reviews originated from North America, with a handful of studies from a range of countries, including, Australia, China, Denmark, Germany, Finland, Japan, Israel, Italy, Pakistan, Slovenia, Sweden, and the UK. See Appendix 1 for summaries of the included reviews (Supplementary Digital Content).

Faculty Development content and process

An overview of FD activities contained in the included reviews is presented in Table 2. In general, the reviews defined FD as a broad activity encompassing the development of

teaching skills/abilities as well as other roles (e.g. administration, scholarship and leadership). However, three of the reviews were more focused: Steinert et al.¹⁵ on FD activities that promoted leadership; Hill et al.¹¹ on FD in the form of ‘resident-as-teacher’ programmes; and Pearce et al.¹² in the form of ‘train-the-trainer’ programmes.

INSERT TABLE 2 ABOUT HERE

While all the reviews reported FD activities that involved participants from medicine, three also included study participants from other professional groups, such as dentistry, nursing, occupational therapy, pharmacy, speech therapy, social work.^{10,12,15} However, even in these reviews the majority of included studies were focused on FD for medical practitioners.

According to the included reviews, the majority of the FD activities were undertaken in classrooms or seminar rooms, with very little FD occurring in workplace/based settings. While all the reviews provided an outline description of the different delivery modes (e.g. workshops, retreats, seminars) involved in their included studies, there was a lack of detailed information on specifics of each of these delivery modes. As a result, it was difficult to discern within and across the reviews the exact nature of the FD activities, and the degree of difference or similarity between them. It was not clear whether the lack of detail description reflected a lack of reporting in the individual studies or in the systematic review process, however, a lack of detailed description has been found to be a common feature of intervention research reports.¹⁷ In contrast, the reviews did provide more information about the duration of the various FD activities included in their respective studies – of which there was quite a lot of heterogeneity across studies, with activities lasting from one hour to three years.

Only one of the reviews¹³ included studies of accredited FD activities which resulted in a certificate of qualification being awarded: eight studies of family medicine fellowships that lead to an advanced degree. The lack of studies of accredited certificated programmes in the reviews is notable given that it has been argued that accreditation is an indicator of the quality of a programme.¹⁸ Moreover, our personal experience has been that course certification can be a key factor in initial learner engagement. The systematic reviews

themselves did not comment on the absence of accreditation in FD programmes. It is not clear whether this absence reflects the relative paucity of accredited FD programmes or rather a lack of impact evaluations of these programmes.

Reported impact on learning outcomes

Six of the reviews^{10-12, 14-16} provided details of the impact on outcomes categorized as FD activity. One of the reviews however did not offer details of the individual outcomes measured in each study but only aggregated information linking FD to reported outcomes¹³.

We classified the reported outcomes using a modified outcomes typology¹⁰ (Table 3). The figures in each cell of the table represent the number of outcomes of that type and level reported in the review. The majority of outcomes reported are at level 1 (learner reaction), 2a (changes to learner attitudes/perceptions), 2b (changes to learner knowledge/skills) and 3 (learner behavioral change). Far fewer outcomes are reported for levels 4a (changes to organisations), 4b (changes in health care student learning resulting from a FD intervention received by their clinical teacher) or 4c (changes to patient care). Furthermore, most studies in the reviews used self-report data (e.g. surveys, interviews) to measure outcomes from a FD activity (i.e. the outcome measures were based on the learners' own reports about self-perceived change). There appears to be very little outcome data derived from rigorously developed, independent and/or blinded, data sources (e.g. systematic observations of teaching practice).

INSERT TABLE 3 ABOUT HERE

Across the reviews, FD activities described as 'workshops' and 'longitudinal programs' report the largest amount of outcomes of all types. However, as the reviews did not provide detailed information that described the nature of the included FD activities it is not clear whether (or how) the activities in these categories are similar or different to each other, or to activities in other categories. The review by Steinert et al.¹⁶ included two outcomes: 'building of a Faculty Development community' and 'building of a community of practice in the workplace' which were not featured in the other reviews. However, unlike the other outcomes reported in in Table 3 (based on results reported in the primary

studies), these additional outcomes were based on Steinert and colleagues' own judgements on whether these outcomes had occurred (or not) in the included studies.

Quality issues in the systematic reviews

Table 4 summarizes the methods used in the reviews. All seven reviews reported using standard systematic review methods for searching, selecting, abstracting, analysing and report writing. All of the reviews undertook searches that involved the use of several bibliographic databases. All but one¹³ used additional sources to find studies including searches of reference lists of included articles, journal hand searches, expert recommendations. Only one review¹² appeared to search for studies that were not published in academic journals by searching the Internet.

Given our selection criteria, we anticipated that the identified systematic reviews would have included within them many of the same studies. However, of the 317 references that we identified in the reviews as being an included study (the total number of studies reported in all the reviews themselves was 328), only 38 studies appeared in two of the reviews, and a further two studies appeared in three of the reviews. This seemed unexpectedly low given that all of the reviews were published after 2006 and had search dates that covered roughly similar periods of time. A closer scrutiny of the selection criteria used in each of the reviews suggest in part this was due to the specific way in which individual reviews had focused their respective selection criteria. For example, on a particular group of health professionals,¹¹ specific type of FD activity,¹⁵ or specific study design (only experimental or quasi experimental research designs with control groups¹²).

In some cases the authors of a systematic review were clearly aware of at least one of the other systematic reviews included in our SREA, and excluded studies from their own review that had been included elsewhere. For example, Steinart et al.¹⁶ excluded studies from the Steinart et al.¹⁴ review. In such cases the justification given was that they wanted to 'develop' or 'build on' the previous review. But this 'development' was not apparent in their review reporting, and indeed it was difficult to see how it could be undertaken in a systematic manner without the inclusion and analysis of the studies in the previous review.

These reviews all reported that a quality assessment of the included studies was completed. The amount of detail about the method of quality assessment reported varied between the reviews. For example, Sorinola & Thistlethwaite¹³ and Steinert et al.^{14,16} used two 5-point scales to assess ‘study quality’ and ‘study findings’. However, both reviews offered little information about how this scoring was operationalized to measure possible threats to validity (bias), or how the scores awarded to the included studies were calculated. In a few of the reviews, the quality scores were simply ‘aggregated’ across the included studies to provide a mean rating – an approach which does not offer a meaningful insight into the methodological quality of any individual study. By contrast, Pearce et al.¹² provided a more transparent and systematic approach to quality assessment as they employed a well-established ‘risk of bias’ approach which scored each of the 18 included studies as being at ‘high’, ‘medium’ or ‘low’ risk of bias, using a set of quality criteria based on establishing descriptive causality.

INSERT TABLE 4 ABOUT HERE

The use of a systematic approach to synthesis is a key element for any systematic review.¹⁹ In three of the reviews,^{11,14,16} the approach used for synthesizing study results was not reported. In two of the reviews, the synthesis method was described as “narrative”¹²⁻¹³ and in one as “integrated”.¹⁵ Leslie et al.¹⁰ reported using SPSS to produce a “synthesized descriptive account of the articles”. It was not transparent what these methods meant in practice in any of these reviews, other than a descriptive reporting of the results of the individual studies. Furthermore, it was not clear in any of the reviews how study quality assessment ratings were used/incorporated in the process of synthesizing results from individual studies (e.g. through ‘weighting’ on the basis of study quality), or in the reviewers’ interpretation of study findings.

Discussion

We undertook this SREA in an attempt to generate a comprehensive account of the research evidence about the impacts of FD activities to help inform the decisions about which kinds of activities may be effective for which groups of practitioners in which contexts. We were therefore looking for evidence that facilitated understanding of how

much 'benefit' was obtained from a specific outcome, from a specific FD activity for a specific group of clinical teachers. Evidence that facilitated interpretation, comparison of different FD approaches and exploration of the various moderators of effect, not simply evidence that 'everything works'.⁷

With one exception,¹⁰ the reviews reported that FD activities for clinical teachers resulted in a number of positive impacts. Specifically, they claimed that the use of workshops and longitudinal programs can lead to positive impacts for clinical teachers in relation to their improvement in their reactions, attitudes/perceptions, knowledge/skills and behaviors). A number of these reviews^{10,13,16} also reported that FD activities can generate positive impacts for clinical teaching organisations.

Though the claims made in all seven reviews seemed to be encouraging, there were some common limitations. For example, these reviews contained little information from their included studies about the characteristics of the reported FD activities or any estimate of the amount of benefit obtained from different activities. As a result, one cannot tell, for example, whether a seminar has greater benefit than a workshop or whether the duration of an activity makes any difference to outcomes or whether different FD activities have different outcomes for different clinical teachers.

The majority of studies contained in these reviews measured impact using self-report data, i.e. the learners were asked whether their participation in a FD benefited them in some way. As self-report data rely on an individual's perception, well-known problems with recall and social desirability bias raise questions about the validity of such data as a measure of impact in this context.²⁰⁻²¹ Consequently, the claims made in the reviews about the impact of FD on learner knowledge, practice and organizational quality is largely based on data with very well-known quality limitations.

Furthermore, the statements about the impacts of FD activity in the reviews imply that outcomes, whatever they may be, are directly attributable to the use of the evaluated FD activity itself. Thus there is a need for these statements to be supported by evidence of descriptive causality²². While there is some debate about the relative merits of different

research designs for this purpose, the single group post-test only design used to measure outcomes in most of the studies in the seven reviews provides a very weak warrant for claims of causality.²³

It may be argued that the relationship between specific aspects of research design and study outcome is debatable.²⁴ It is therefore it is reasonable to include studies using different research designs (notwithstanding the issue of establishing descriptive causality) to investigate the same research question within in a systematic review. However, where this is the case we would expect that the effects of heterogeneity of study design would be one of the factors explored as part of the analysis within a systematic review. Unfortunately, there was a limited quality assessment of the included studies in most reviews and no assessment of heterogeneity. Furthermore with one exception,¹² the individual study quality assessment methods used in reviews, did not appear to be linked to establishing a warrant for a causal claim i.e. it is not clear what the quality assessment judgement for any study means, e.g. 'good' but for what?

Implications

We undertook the SREA to identify evidence to help inform the decisions about which kinds of FD activities may be effective for which groups of clinical teachers in which contexts. Due to the accelerated nature of the SREA approach we may have missed other evidence and/or details about the quality of evidence that was reviewed.⁵⁻⁶ However, our analysis of these reviews identified numerous limitations in reporting and methods. These limitations mean that individually and collectively these reviews do not provide practitioners with detailed, transparent and rigorous evidence that can inform choices in practice in the way that we hoped they would, even where the review authors claim they do.

In terms of developing the evidence base for FD in future, one step to consider is whether the evidence could be improved by more detailed systematic review of individual studies of the impact of FD contained within these reviews. It does appear that in some of these reviews there were studies which investigated the impact of FD activities using research designs that are explicitly concerned with establishing a warrant for descriptive causality, for

example, the studies in the Pearce et al. review.¹² It may be useful to identify these studies as the basis for a new systematic review of the evidence.

However, even if a new review was carried out using these studies it would appear that it would only cover a limited range of FD types (e.g. 'Train the Trainer' model¹²), be limited to self-reported learning outcomes, and only include medical teachers. Therefore, while such a review may give us more confidence about the quality of the evidence, it would be focused on a limited range of FD activities in a limited range of contexts. There may be other relevant studies that were not included in these reviews. However, based on our knowledge of research in this field we suspect that this will add to the quantity of the evidence base, but without necessarily strengthening the warrant for claims about for impact. Given these limitations, it would seem imperative that commissioners and providers of FD activities work together with researchers to improve the quality of the evidence in this area. This work could focus on improving the systematic descriptive reporting of FD activities using something similar to the TIDieR Checklist,²⁵ and also considering the requirements for demonstrating descriptive causality when designing evaluations of FD activities.

The findings of our analysis also provide some food for thought for the community about the scope and methods of systematic reviews in this field. Our initial excitement at identifying seven systematic reviews of the impact of FD of recent provenance and apparently broad scope did not yield the results we anticipated. Two particular issues stand out here. Although the systematic reviews appeared to follow the methods of systematic review in their early stages, i.e. searching, selection, use of quality assurance processes, they did not use systematic transparent synthesis methods that incorporated assessment of individual study quality. Without this it is not possible to make judgements about the warrant for any claim about the impact of, for example, different types of FD activities, or to assess any potential causes of heterogeneity of outcome such as study design. Labelling whatever process of synthesis which was undertaken as 'narrative' or 'integrated' does not substitute for detailed systematic synthesis reporting. The omission of this key stage in the systematic review process undermines the potential value of such reviews considerably.

A second issue is the way in which reviews are focused. The systematic reviews included in this SREA were focused either on a particular type of FD, a particular health profession and/or a desire to expand on other reviews. While this may seem like a reasonable approach for an individual review it does not appear to be leading toward a coherent comprehensive account of the field as a whole. We are not clear how this can be addressed given the prevailing disciplinary/professional boundaries, academic incentives and limited funding. Nevertheless, there may be a role for collaborations such as BEME, The Cochrane Collaboration, The Campbell Collaboration or the Joanna Briggs Institute in seeking to manage curation of relevant studies and coordinate reviews. At the very least it would seem appropriate for reviews seeking to 'build on' or 'develop' from other systematic reviews to do so in a systematic transparent fashion. For example, a review that is intended to update or expand upon an existing review should include the studies from the original review in a revised synthesis. Given the limitations we identified in the seven included FD reviews, this is likely to mean re-analyzing the studies from the first review with any newly identified studies in a completely new systematic review.

References

1. Austerberry A & Newman M. Review of qualifications and training for clinical educators in the healthcare professions. London. London Deanery; 2013. Available at: <http://www.faculty.londondeanery.ac.uk/other-resources/review-of-qualifications-and-training-for-clinical-teachers-in-the-healthcare-professions-4>
2. Steinert Y. Faculty Development: Future Directions. In: Steinert Y. (eds) Faculty Development in the Health Professions. Innovation and Change in Professional Education, vol 11. Springer, Dordrecht; 2014
3. Stritter F. Faculty evaluation and development. In: C McGuire, R Foley, A Gorr, R Richards (eds). Handbook of Health Professions Education. San Francisco, Calif: Jossey Bass; (p294–318); 1983.
4. Davies, P. What is evidence-based education? British Journal of Educational Studies, 1999; 47(2) 108 -121
5. Thomas J, Newman M, Oliver S. Rapid evidence assessments of research to inform social policy: taking stock and moving forward. Evidence and Policy, 2013; 9: 5-27.
6. Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. Implementation Science, 2010; 5:56. DOI: <https://doi.org/10.1186/1748-5908-5-56>

7. Hattie J. Visible learning: A synthesis of over 800 meta-analyses relating to achievement. London Routledge; 2009
8. Higgins, S. Meta-synthesis and comparative meta-analysis of education research findings: some risks and benefits, *Review of Education*, 2016; 4.1: 31–53. DOI: <http://dx.doi.org/10.1002/rev3.3067>
9. Reeves S, Fletcher S, Barr H, Birch I, Boet S, Davies N, McFadyen A, Rivera J, Kitto S. A BEME systematic review of the effects of interprofessional education: BEME Guide No. 39. *Medical Teacher*, 2016; 38(7):656-68
10. Leslie, K, Baker, L, Egan-Lee, E, Esdaile, M, Reeves, S. Advancing Faculty Development in medical education: A systematic review, *Academic Medicine*, 2013; 88: 1038-1045
11. Hill, A. G, Tzu-Chieh, Y, Barrow, M, Hattie, J. A systematic review of resident-as-teacher programmes, *Medical Education*, 2009; 43: 1129-1140
12. Pearce, J, Mann, M. K, Jones, C, Buschbach, S, Olf, M, Bisson, J. The most effective way of delivering a Train-the-Trainers program: A systematic review, *Journal of Continuing Education in the Health Professions*, 2012; 32: 215-226
13. Sorinola, O, Thistlethwaite, J. A systematic review of Faculty Development activities in family medicine, *Medical Teacher*, 2013; 35: 1309-1318
14. Steinert, Y, Mann, K, Centeno, A, Dolmans, D, Spencer, J, Gelula, M, Prideaux, D. A systematic review of Faculty Development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8, *Medical Teacher*, 2006; 28: 497-526
15. Steinert, Y, Naismith, L, Mann, K. Faculty Development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME Guide No. 19, *Medical Teacher*, 2012; 34: 483-503
16. Steinert, Y, Mann, K, Anderson, B, Barnett, B, Centeno, A, Naismith, L, Prideaux, D, Spencer, J, Tullo, E, Viggiano, T, Ward, H, Dolmans, D. A systematic review of Faculty Development initiatives designed to enhance teaching effectiveness: A 10-year update: BEME Guide No. 40, *Medical Teacher*, 2016; 38: 769-786
17. Hoffmann TC, Oxman AD, Ioannidis JPA, Moher D, Lasserson TJ, Tovey DI, Stein K, Sutcliffe K, Ravaud P, Altman DG, Perera R. Enhancing the usability of systematic reviews by improving the consideration and description of interventions. *BMJ*, 2017;357:j2998
18. Harvey L. A history and critique of quality evaluation in the UK. *Quality Assurance in Education*, 2005; Vol. 13 (4) pp.263-276
19. Gough D & Thomas J. Commonality and diversity in reviews in In Gough D, Oliver S, Thomas J (Eds) *An Introduction to Systematic Reviews (2nd Edition)*, London: Sage, 43-70; 2017

20. Bowman NA & Hill PL. Measuring how college affects students: Social desirability and other potential biases in college student self-reported gains. *New Directions for Institutional Research*, 2011; 150, 73-85
21. Pike GR. Using college students' self-reported learning outcomes in scholarly research, *New Directions for Institutional Research*, 2011; 150, 41-48
22. Newman M. Fitness for purpose evaluation in Problem Based Learning should consider the requirements for establishing descriptive causation. *Advances in Health Sciences Education*, 2006; 11, 391–402
23. Torgerson, C. & Torgerson, D. The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, 2001; 49, 316–318.
24. Wells, K & Littell JH. Study quality assessment in systematic reviews of research on intervention effects. *Research on Social Work Practice*, 2009; 19(1), 52-62
25. Hoffmann TC, Glasziou PP, Boutron IP, Milne R, Perera R, Moher D , Altman DG, Barbour V, Macdonald H, Johnston M, Lamb SE, Dixon-Woods M, McCulloch P, Wyatt JC, Chan AC, Michie S, Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, 2014; 348:g1687

Implications for practice

- Faculty Development (FD) is a key activity to support the professional growth (knowledge, skills and abilities) of clinical teachers.
- While research about the impact of FD is growing, this Systematic Rapid Evidence Assessment (SREA) found that the positive claims made about the impact of FD in seven systematic reviews contained number of key limitations which generated uncertainty about their reported outcomes.
- To improve the quality of future systematic reviews of FD, reviewers need to address the methodological shortfalls identified in this SREA, paying particular attention to providing more transparency in their work.