Realist evaluation of faculty development: What works for who, in what context and why

Olanrewaju Sorinola*, University of Warwick, Warwick Medical School, Coventry, UK
Jill Thistlethwaite, University of Technology, School of Medicine, Sydney, Australia
David Davies, University of Warwick, Warwick Medical School, Coventry, UK
Ed Peile, University of Warwick, Warwick Medical School, Coventry, UK

Introduction: The central mission of every medical school is education, however most teachers have not received formal teacher training. Hence faculty development (FD) with planned programmes to prepare faculty members for their role in teaching is integral to medical school life. The widespread investment in FD is predicated on the belief that it enhances the effectiveness of teaching but the evidence is limited. The aim of this study was to carry out an in-depth exploration of FD for educators in UK medical schools and evaluate its effectiveness. The realist framework with its principle of explanatory causation was chosen to find out what works for whom in FD in what context, and why.

Methods: A detailed literature review was carried out, informing the development of eight realist hypotheses in the form of contexts, mechanisms and outcomes (CMO) based on the researchers’ experience of and insight into FD. Contexts are the individual, interpersonal, institutional and infrastructural factors that influence programme efficiency. Mechanisms describe what it is about interventions that bring about change while outcomes are the consequences of the intervention. Data to support, modify, or challenge the hypotheses were collected in three phases. Phase I was a review and scoring of data of UK medical school FD webpages using a scoring index. Phase II was observation and informal interviews of two cohorts of educators attending a FD course followed by detailed interviews of 12 educators six months later. Phase III was interviews of FD coordinators and educators at eight medical schools. Quantitative and qualitative data were collected. Quantitative data were analysed by descriptive methods, slope diagrams, bi-axial constructs and Fisher’s exact test. Qualitative data were analysed by categorising and connecting strategies followed by summarising the relevant data under each hypothesis to check if the hypothesis was supported, modified or refuted.

Results: The literature review revealed a paucity of publications on FD in the UK, however 30 medical schools’ webpages showed data on FD activities. Data from all three phases supported four CMOs hypotheses which explain how the contexts of participatory approach/reflective practice, needs identification, supportive setting and standardization of medical teaching facilitated FD mechanisms of engagement, motivation, positive perception and professionalization respectively. These led to the outcomes of improved confidence, competence, credibility and career progression.

Theory 1 Engagement: Participatory approach / reflective practice in FD + Engagement = Increased confidence in teaching
Theory 2 Motivation: Needs assessment of educators / FD relevance + Motivation to learn = Improved competence
Theory 3 Positive perception: FD in a supportive setting / access to programmes + Positive perception of the value of FD = Credibility as a teacher
Theory 4 Professionalization: Demand to standardise teaching / accreditation + professionalization of teaching = Career progression / improved recognition

Discussion and Conclusions: Four realist theories of FD were identified (engagement, motivation, positive perception and professionalization) illuminating the effectiveness of FD both in the short and long term. We make recommendations to all FD stakeholders (educators, FD developers, universities and policy makers) and for future medical education researchers who choose the realist evaluation model.


Educating physicians in evidence based medicine: Current practices and curricular strategies

Lauren Maggio*, Stanford University, Stanford, USA
Olle ten Cate, University of Utrecht, Stanford, Netherlands
David Irby, University of California, San Francisco, USA
Bridget O’Brien, University of California, San Francisco, USA

Introduction: Evidence based medicine (EBM) is “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients.”(1) The practice of EBM is an expectation of professional healthcare and requisite component in many medical school curricula. Yet, despite training, physicians sub-optimally practice EBM due to multiple factors (2). Education may mitigate these factors and improve EBM practice. However, there is limited knowledge on how is EBM practiced and taught in medical schools.

Methods: To understand how EBM is practiced, specifically the first two steps of EBM (ask and acquire), we 1) interviewed American and Dutch physicians to identify clinical information needs. We next 2) developed a web log methodology that automates the mass retrieval of article-level metadata of articles accessed via PubMed and UpToDate and analyzed a year of Stanford University Hospitals’ web logs to understand information access in clinical care. To characterize how EBM is taught we 3) systematically reviewed the literature to identify EBM educational interventions. We also 4) interviewed faculty at North American medical schools, identified as graduating students confident in their EBM abilities,
to identify challenges students face when learning EBM and how institutions approached these challenges.

**Results:** Based on 1) 22 interviews, we identified six information needs, including refreshing, confirming, logistics, teaching, information generating and personal learning, which align with physicians’ varied roles. Our 2) web log study revealed twice as much use of UpToDate (110,336 visits), a topical review resource, as PubMed (47,244 visits). Of total journal articles viewed (81,851), 30% (24,529) were review articles and 12% (10,261) clinical trials. Our 3) literature review described educational settings, learner levels, teaching methods and EBM skills covered. No intervention addressed all EBM steps and due to lack of details reported we were unable to draw conclusions about the efficacy of interventions. Based on 4) 31 interviews with educators at 17 medical schools, we identified six student learning challenges, including sub-optimal role models, and five learning approaches to overcome these challenges, such as incorporation of clinical content and faculty development.

**Discussion and Conclusions:** Study 1) presents information needs, which can help structure educational approaches to support physicians’ clinical inquiry in their varied roles, including teacher and lifelong learner. Study 2) demonstrates heavy use of review resources for EBM. This suggests the traditional approach to EBM training, which generally focuses on critically analyzing individual studies, should expand to include appraisal of review resources to better synchronize training and practice. Both studies provide a glimpse into practices potentially observed by students, therefore contributing to their EBM learning experience. Study 3) demonstrates a need for training to comprehensively cover all EBM skills and more detailed reporting of EBM interventions to enable evaluation and replication. Student challenges and educational approaches identified in Study 4) provide a window into students’ experience of EBM training, which medical educators can use to design learner-informed EBM curricula. This PhD report provides knowledge of how EBM is practiced and taught and equips medical educators with suggestions, drawn from practice, for examining and modifying their EBM curriculum.

**References:**

**#713 (23739)**
Understanding assessors’ behaviours in the context of performance-based assessments

**Stefanie S Sebok**, Queen’s University, Education, Kingston, Ontario, Canada
Don A Klinger, Queen’s University, Education, Kingston, Ontario, Canada

**Introduction:** Performance-based assessments are often used in medical education to assess the competency of medical candidates. However, one of the challenges with this particular form of assessment is the variability among assessors (i.e., individuals responsible for providing an evaluative judgment). Previous studies have shown that assessors are variable (see Gingerich, Regehr, & Eva, 2011; Gingerich, Kogan, Yeates, Govaerts, Holmboe, 2014), which can make it difficult to obtain consistent assessment information from these assessors. Looking at the problem of assessor variability differently, this work proposes studying assessor cognition in order to better understand what assessors are thinking as well as the approach and information they use when evaluating medical candidates.

**Methods:** This work used artifacts (i.e., ratings, checklists, and comments) within a multi-study quantitative approach. The first study examined what assessors’ ratings can tell us about individual assessors. Using Rasch analyses and hierarchical clustering techniques, this study identified instances where assessor variability was high and explored how raters differentially constructed individualized notions of candidate competence based on various constructs (e.g., communication and professionalism). The second study explored the use of checklists and how these could provide information about individual assessors. When using only two checklist options (i.e., done and not done), assessors tend to default to “done” (even in situations where a task was not completed satisfactorily). Assessors were provided with an alternative checklist option to determine whether the precision of assessors’ judgments could be enhanced. The final study examined assessors’ comments and what these could tell us about individual assessors. This study investigated the different types of comments found in workplace-based assessments, more specifically in terms of how these comments corresponded to assessors’ use of anchored benchmarks (e.g., outstanding, satisfactory, unsatisfactory).

**Results:** The results of the first study demonstrate that, at times, assessors were unable to distinguish between the various constructs used to assess competence. Additionally, differences in terms of how individual assessors saw the various constructs were also observed. Through the use of an addition checklist option (attempted), the second study highlighted instances when raters required a third option to more accurately articulate the observed competency of a medical candidate. This study also found that assessor variability was a result of failing to involve assessors in all stages of the assessment process. The final study highlighted patterns in the valence and quality of feedback assessors provide. Furthermore, this study also demonstrated that sometimes assessors use blank spaces or simply fail to provide comments as a way of communicating negative or redundant information about a candidate’s performance.
Introduction: 
Persistent concerns about rater variability in clinical performance ratings has provoked a new field of study in medical education that focuses on better understanding how physicians think while performing assessment tasks. In this doctoral program of research, the assumption that “error” variance can largely be attributed to raters making idiosyncratic judgments is challenged by drawing on research methods from the social cognition literature investigating variability in judgments when people are forming impressions of others.

Methods: A review of current perspectives on assessor cognition in medical education research plus a review of social categorization in impression formation research informed the design of three mixed-methods studies. These studies examined whether meaningful clusters (or meta-impressions) can be extracted across raters’ idiosyncratic impressions of a given clinical performance. The capacity of these different meta-impressions (for the same performance) to explain variance in Mini-CEX ratings was used as an indirect measure of their influence on assessors’ ratings.

Results: When multiple physicians observed and assessed the same resident, a limited number of meta-impressions were formed. The meta-impression with which each rater’s “idiosyncratic” impression was most closely associated explained substantial variance in the ratings. Different meta-impressions arose as different subgroups of physicians 1) emphasized different aspects of the performance as most salient, 2) disagreed on the interpretation of the same aspect of the performance and 3) drew different inferences about the reasons the resident performed as he/she did.

Discussion and Conclusions: Consistent with impression formation research, there appears to be more consensus than implied by the “idiosyncratic error” model of rater judgments. There were always multiple clusters of consensus for each performance. These clusters of consensus, or meta-impressions, often conflicted with each other in ways that would not be easily reconciled into a single point of view or summarized by a single numerical rating. Thus, rater variability may be less a matter of “idiosyncratic error” masking a single true signal, and more a matter of multiple legitimate (and sometimes conflicting) perspectives on the same performance. As such, solutions aimed at reducing idiosyncrasy may not decrease error variance in ratings and we may need to grapple with the challenge of incorporating meaningful diversity of opinion when trying to summarize individuals’ abilities. In conclusion, the pervasive variability seen in clinical performance ratings is not well explained by thinking of it being due to raters making “error-prone” assessment judgments that are simply idiosyncratic to each rater. Rather, the variability may represent multiple points of view that each contain a portion of the potentially relevant information about that resident and his/her performance. Novel assessment designs would be required to appropriately collect and analyze assessment judgments of this nature.

References:

#714 (23568)
Questioning the rater idiosyncrasy explanation for error variance
Andrea Gingerich*, University of Northern British Columbia, Northern Medical Program (UBC Medicine), Prince George, Canada
Cees van der Vleuten, Maastricht University, School of Health Professions Education, Maastricht, Netherlands
Kevin Eva, University of British Columbia, Centre for Health Education Scholarship, Vancouver, Canada
Glenn Regehr, University of British Columbia, Centre for Health Education Scholarship, Vancouver, Canada

Discussion and Conclusions: Assessors’ ratings, checklist options, and narrative comments provide valuable insights regarding how assessors think and make evaluative judgments about medical candidates’ competence. Further, these studies illustrate the different methods these assessors use to provide supporting information about medical candidates’ performances. Assessing medical competence depends upon knowing what assessors are doing as well as when they exhibit certain behaviours. Through the utilization of an assessor cognition approach, these studies illustrate the importance of exploring assessor variability in order to figure out how to better extract information from assessors.


#715 (23602)
Cognition, culture, and credibility: deconstructing feedback in medical education
Christopher Watling*, Western University, Clinical Neurological Sciences, London, Canada

Discussion and Conclusions: Questioning the rater idiosyncrasy explanation for error variance
Andrea Gingerich*, University of Northern British Columbia, Northern Medical Program (UBC Medicine), Prince George, Canada
Cees van der Vleuten, Maastricht University, School of Health Professions Education, Maastricht, Netherlands
Kevin Eva, University of British Columbia, Centre for Health Education Scholarship, Vancouver, Canada
Glenn Regehr, University of British Columbia, Centre for Health Education Scholarship, Vancouver, Canada

Introduction: Persistent concerns about rater variability in clinical performance ratings has provoked a new field of study in medical education that focuses on better understanding how physicians think while performing assessment tasks. In this doctoral program of research, the assumption that “error” variance can largely be attributed to raters making idiosyncratic judgments is challenged by drawing on research methods from the social cognition literature investigating variability in judgments when people are forming impressions of others.

Methods: A review of current perspectives on assessor cognition in medical education research plus a review of social categorization in impression formation research informed the design of three mixed-methods studies. These studies examined whether meaningful clusters (or meta-impressions) can be extracted across raters’ idiosyncratic impressions of a given clinical performance. The capacity of these different meta-impressions (for the same performance) to explain variance in Mini-CEX ratings was used as an indirect measure of their influence on assessors’ ratings.

Results: When multiple physicians observed and assessed the same resident, a limited number of meta-impressions were formed. The meta-impression with which each rater’s “idiosyncratic” impression was most closely associated explained substantial variance in the ratings. Different meta-impressions arose as different subgroups of physicians 1) emphasized different aspects of the performance as most salient, 2) disagreed on the interpretation of the same aspect of the performance and 3) drew different inferences about the reasons the resident performed as he/she did.

Discussion and Conclusions: Consistent with impression formation research, there appears to be more consensus than implied by the “idiosyncratic error” model of rater judgments. There were always multiple clusters of consensus for each performance. These clusters of consensus, or meta-impressions, often conflicted with each other in ways that would not be easily reconciled into a single point of view or summarized by a single numerical rating. Thus, rater variability may be less a matter of “idiosyncratic error” masking a single true signal, and more a matter of multiple legitimate (and sometimes conflicting) perspectives on the same performance. As such, solutions aimed at reducing idiosyncrasy may not decrease error variance in ratings and we may need to grapple with the challenge of incorporating meaningful diversity of opinion when trying to summarize individuals’ abilities. In conclusion, the pervasive variability seen in clinical performance ratings is not well explained by thinking of it being due to raters making “error-prone” assessment judgments that are simply idiosyncratic to each rater. Rather, the variability may represent multiple points of view that each contain a portion of the potentially relevant information about that resident and his/her performance. Novel assessment designs would be required to appropriately collect and analyze assessment judgments of this nature.

References:

#715 (23602)
Cognition, culture, and credibility: deconstructing feedback in medical education
Christopher Watling*, Western University, Clinical Neurological Sciences, London, Canada

Introduction: Despite widespread endorsement of feedback as an essential facilitator of learning, there have been few explorations of how it becomes meaningful for learners. In reality, feedback is far from straightforward, and may fail to usefully impact learners. In this thesis, we explore feedback within and outside of medical education in order to understand why and how it succeeds or fails, with a view to informing real progress in its use.

Methods: Constructivist grounded theory methodology guided 5 linked studies. Purposive sampling drove recruitment of 99 individuals to participate in interviews or focus groups. Constant
comparative analysis to identify themes was conducted iteratively.

Study 1 explored the learning experiences physicians perceived as influential in their training, considering why those experiences had resonated. Study 2 tested the usefulness of regulatory focus theory in explaining learners’ responses to feedback. Study 3 compared the learning cultures of medicine and music to unveil medicine’s tacit assumptions about learning. Study 4 examined feedback in 3 learning cultures – medicine, music, teacher training – exploring how culture influences the handling of feedback. Study 5 studied individuals traveling between learning cultures, doctors trained in music and sports – to explore how individual and sociocultural influences on feedback interact.

Results: As they participate in clinical work, medical learners are shaped by a variety of learning cues. Feedback competes with other learning cues for learners’ attention, becoming influential only if it survives a judgement of its credibility. Regulatory focus theory provides insight into learners’ feedback responses, but there are challenges in applying regulatory focus theory to real feedback scenarios: regulatory focus may be mixed and may shift over time, and other factors (eg. feedback’s credibility) may trump regulatory focus in driving learner responses. Medicine’s pedagogical approaches reflect cultural assumptions; medicine privileges learning by doing, guides its learners toward competence more than toward excellence, and values the clinical skill more than the instructional skill of its teachers. Credibility and constructiveness transcend culture as essential components of meaningful feedback, but each culture creates distinct definitions of credibility and constructiveness. Learning culture modulates feedback’s impact, creating the conditions for good feedback to occur and for learners to respond.

Discussion and Conclusions: Together, these studies offer a model for understanding feedback that incorporates both the individual learner and the learning culture as essential and inseparable elements. Individual variability in feedback response is inevitable, but learners can only hope to benefit if the learning culture makes meaningful feedback possible. Learning culture contributes more than just opportunity for meaningful feedback to occur; its inherent norms and values shape what counts as credible feedback and what demands learners’ attention. This thesis directs medical educators’ attention towards two historically marginalized influences on feedback, learner perceptions and learning culture – and offers concrete guidance for improving our status quo. Medicine’s learning culture is vulnerable in the very elements that matter most to feedback: routine learner observation, trusting longitudinal teacher-learner relationships, and authentic coaching. To become a feedback culture, medicine must remedy impediments to credible, constructive feedback, and commit to building the supports necessary for meaningful feedback to occur.