

Greatest Hits of 2023! Notable Non-JGME Medical Education Articles

Gail M. Sullivan, MD, MPH

Deborah Simpson, PhD

Anthony R. Artino Jr, PhD

Nicole M. Deiorio, MD

Lalena M. Yarris, MD, MCR

The editors of the *Journal of Graduate Medical Education* have again dusted off articles, published in other journals, that we found intriguing and/or useful from the past year. We compile this summary before the end of the calendar year; thus, we miss key late-breaking articles relevant to graduate medical education (GME). However, we've given December 2022 articles a careful look, so we are not unfairly neglecting the final, festive month of the calendar year. As always, our diverse backgrounds produce an eclectic mix of articles to share with you (BOX 1, 2). We regularly peruse a wide variety of medical education journals and venues, although not all, and we use no scientific approach for our choices; we just like them. Take our recommendations with a pinch of salt (or a glass of wine) and return the favor with recommendations of your own. Salut!

Tony Artino's Pick

This past year Sawatsky and an impressive team of collaborators published a compelling empirical paper, "Professional Identity Struggle and Ideology: A Qualitative Study of Residents' Experiences."¹ The authors used qualitative methods to explore the complex dynamics of medical ideology and its influence on the formation of professional identity among GME trainees. Medical ideology, defined by the authors as "the system of ideas, often not explicitly stated, behind the structures and processes of medicine and medical education that drive the practices and discourses of medicine and medical education," served as the lens through which the researchers examined the intricacies of medical socialization.¹ In short, this study explored how individual beliefs and professional expectations intersect and shape the trajectory of GME residents.

Central to this investigation is a recognition that medical residents face identity struggles. The research

illuminates that becoming a medical professional extends beyond acquiring knowledge and skills; it involves a complex integration of personal and professional identities. The study spanned 3 medical specialties—internal medicine, emergency medicine, and family medicine—at 3 US academic institutions. Sawatsky et al employed a creative data collection approach, which combined picture-drawing exercises and in-depth interviews, to gather nuanced qualitative data from 12 participants. The authors used *rich pictures* (ie, asking participants to draw a picture of a challenging residency experience that included an identity struggle) as an elicitation technique, to help participants deeply reflect on experiences of identity struggle. The methodology was aimed at capturing the multifaceted nature of residents' professional identity struggles. The rich pictures not only enriched data collection, but also provided a unique window into the complex and often abstract concepts of professional identity and ideology. Thematic analysis of the narratives elicited from these rich pictures delved into the nuances of residents' experiences to reveal layers of meaning potentially overlooked by conventional text-only methods.

The study's findings coalesce around several key themes central to the residents' experiences of identity struggle. First, the demanding nature of GME and a pervasive culture of perfectionism were major contributors to personal distress and career doubts among residents. Second, the research highlighted a tension between residents' preexisting personal identities and their developing professional identities, which underscores a friction between their authentic selves and expected professional personas. Lastly, the study unearthed a gap between the idealized concept of medicine and the harsh realities of clinical practice. For instance, residents noted that systemic constraints in health care often impeded their ability to provide optimal patient care, which led to struggles in enacting their desired professional identities and disillusionment, further complicating their identity formation.

DOI: <http://dx.doi.org/10.4300/JGME-D-24-00009.1>

Box 1 Recommended Non-JGME Articles From the Past Year

Bearman M, Ajjawi R, Castanelli D, et al. **Meaning making about performance: a comparison of two specialty feedback cultures.** *Med Educ.* 2023;57(11):1010-1019. doi:10.1111/medu.15118

Kendrick DE, Thelen AE, Chen X, et al. **Association of surgical resident competency ratings with patient outcomes.** *Acad Med.* 2023;98(7):813-820. doi:10.1097/ACM.0000000000005157

Lees AF, Beni C, Lee A, et al. **Uses of electronic health record data to measure the clinical learning environment of graduate medical education trainees: a systematic review.** *Acad Med.* 2023;98(11):1326-1336. doi:10.1097/ACM.0000000000005288

Mann A, Shah AN, Thibodeau PS, et al. **Online well-being group coaching program for women physician trainees. A randomized clinical trial.** *JAMA Netw Open.* 2023;6(10):e2335541. doi:10.1001/jamanetworkopen.2023.35541

Sawatsky AP, Matchett CL, Hafferty FW, et al. **Professional identity struggle and ideology: a qualitative study of residents' experiences.** *Med Educ.* 2023;57(11):1092-1101. doi:10.1111/medu.15142

My take home: This research underscores the critical need for medical educators and institutions to acknowledge and address the ideological influences impacting the development of a medical professional's identity. Professional identity development is not the smooth, somewhat linear process that is often depicted. Instead, there is struggle, much of it ideologically driven. GME leaders need to be more aware of the nature of this struggle and how best to support trainees. By bringing these often-overlooked factors to light, the study not only highlights the internal conflicts faced by medical residents but also prompts a reevaluation of the current GME system. Ultimately, the authors advocate for a more comprehensive approach to GME training, one that fosters both professional competence and personal growth.

Box 2 Honorable Mention, Non-JGME Articles From the Past Year

Anderson LM, Rowland K, Edberg D, Wright KM, Park YS, Tekian A. **An analysis of written and numeric scores in end-of-rotation forms from three residency programs.** *Perspect Med Educ.* 2023;12(1):497-506. doi:10.5334/pme.41

Chen A, Chen DO. **Accuracy of chatbots in citing journal articles.** *JAMA Netw Open.* 2023;6(8):e2327647. doi:10.1001/jamanetworkopen.2023.27647

Farrell L, Cuncic C, Hartford W, Hatala R, Ajjawi R. **Goal co-construction and dialogue in an internal medicine longitudinal coaching programme.** *Med Educ.* 2023;57(3):265-271. doi:10.1111/medu.14942

Khan R, Hodges BD, Martimianakis MA. **When I say... burnout.** *Med Educ.* 2023;57(8):704-705. doi:10.1111/medu.15088

Nicole Deiorio's Pick

Kendrick et al's "Association of Surgical Resident Competency Ratings With Patient Outcomes" uses the Medicare database to examine patient postoperative outcomes in early-career surgeons, based on their final-year Milestone ratings.² Interestingly, although Milestones are designed to track readiness toward unsupervised practice, no association was found between complication rates of surgeons and proficient vs not-yet-proficient Milestone rating levels.

In the accompanying commentary, Montgomery et al rightly point out the limitations to this paper in understanding the complex relationships among an early-career surgeon's patients, patient complications, and Milestone data.³ However, this analysis is a solid step forward in looking at high-level, patient-care outcomes, which is usually difficult to achieve in medical education.

My take home: The original research paper and accompanying commentary are required reading for those interested in staying up to date on the conversations around competency-based medical education (CBME). While this particular data analysis is limited to surgical residents and patients with Medicare, the discussions of the predictive value of assessment tools, such as Milestones and entrustable professional activities, touch on issues and nuances germane to any specialty and medical education as whole.

Honorable Mention

I encourage readers to also look at Farrell et al's paper, "Goal Co-Construction and Dialogue in an Internal Medicine Longitudinal Coaching Programme," which improves our understanding of what is happening within coaching conversations related to goal co-construction.⁴ Coaching in medical education continues to be a hot topic in the literature. Coaching requires investments in time and money, so we hope for meaningful returns on these investments. Thankfully, as coaching spreads, we now see studies reporting tangible outcomes from these programs.

Deb Simpson's Pick

Skeptics of big education data for trainee assessment in GME beware—it's becoming a reality! Per a systematic review by Lees et al of objective assessments to measure resident and fellow competencies, the use of electronic health record (EHR) data is rapidly evolving.⁵ In this first systematic review of EHR data for GME performance assessment, the authors searched MEDLINE from its inception through 2021, which yielded 3558 articles. Following PRISMA guidelines,⁶ the authors derived 86 articles for final review based

on exclusion (eg, not original research, non-English language) and inclusion (lacking the use of routinely collected EHR health data or GME trainee as the unit of observation or analysis) criteria. Data was coded for multiple topics, including article theme, trainee specialty, Accreditation Council for Graduate Medical Education (ACGME) Core Competencies, and attribution method. The gnarly issue of how to attribute a patient's care to a specific resident varied by article, and none provided quantitative validity evidence.

Articles clustered around 16 themes,⁷ including training experience, work patterns, and continuity of care. However, framing the results using ACGME Core Competencies highlighted marked between-trainee variation. The most frequently studied competency was Patient Care and Procedural Skills, found in 33% (n=28) of articles. All 9 articles that compared trainee data with national standards revealed gaps between trainee and national standards. The review also found that manual procedure logs were typically incomplete and/or inaccurate when compared with EHR-derived logs.

Twenty percent (n=17) of the articles focused on problem-based learning and improvement. Just over half of these articles reported providing trainees with their individual practice data through reports and dashboards. Systems-Based Practice competencies were found in 15% of the articles (n=13); most focused on continuity of care and panel size, with considerable variation between trainee experiences. Professionalism and Medical Knowledge was explored in a few articles (n=5). The authors created a sixth category to account for 29% articles (n=25), labelled the clinical learning environment, which encompassed EHR data associated with workload, work patterns, and work hours.

The authors concluded with a proposal for a "digital learning cycle framework," aligned with the Kolb learning cycle, to support sequential uses of trainee EHR data over time. They also recommend 3 technical components essential to optimizing the use of EHR data for GME assessment purposes.

Methodologically this article exhibits the hallmarks of a quality systematic review: searching multiple databases, focused questions, and describing and appraising the quality of included studies.⁸ Articles varied widely in their questions and assumptions, particularly around attribution methods, which prevented meta-analysis. Other problems included variability among EHRs, even with the same vendor, how to count work hours with inaccurate sign outs, inability to see the original note for attribution, and faculty attesting for procedures for billing purposes. All of these issues required authors to make coding decisions that are

well documented in the supplemental information, providing transparency for future investigators.

My take home: If you have not yet started considering how to use EHR data for resident assessments, this review will provide a strong introduction to the field and its challenges. The future is coming faster than we anticipated, as EHR data are becoming drivers for learning and assessment, through linking individual residents to patient and outcome data with attribution algorithms.⁹

Gail Sullivan's Pick

The COVID-19 pandemic's effects on GME were not all deleterious: we developed creative ways to connect. In programs and institutions where some trainees are a small minority, whether related to gender, racial, sexual orientation, or other backgrounds, remote connection to outside coach or mentor programs, to enhance professional success and personal satisfaction, could be a game changer. But humans usually develop trusting connections through real-world contact: do remote interactions work, and are they worth our time and effort?

Dr. Mann and colleagues begin to answer this question in their randomized controlled trial of a remote well-being group coaching program for women physician trainees.¹⁰ The authors focused on women, as studies report higher burnout in women. Most studied interventions to date have been short-duration, single-program, or single-site interventions. The authors used the Better Together Physician Coaching (BT) program, developed and piloted by the University of Colorado, from September to December 2022 at 26 diverse GME institutions. All GME trainees self-identifying as women were invited, with voluntary participation. Participants were randomized, stratified by site to BT vs control (delayed access to BT, after study conclusion). Both groups received emailed access to online wellness resources. Outcomes were pre- and post-intervention surveys of well-being and distress (Maslach Burnout Inventory [MBI] subscales, Young Imposter Syndrome scale, Moral Injury Symptom Scale-Health Professionals, Neff Self-Compassion Scale-Short Form, and Secure Flourish Index).

BT is a 4-month, online, group coaching program led by physicians certified by The Life Coach School. It consists of weekly content (self-monitoring, behavior change, etc). Participants have access to 3 to 4 weekly, live video coaching calls (also recorded for asynchronous watching), individual written coaching, and weekly self-study modules. Topics are modified by the pre-survey and discussions during the program.

The authors used their pilot data to calculate participant number for sufficient power to detect changes and analyzed their data with linear and logistic regression models. They generated odds ratios adjusted for baseline values, with intention-to-treat and sensitivity analyses: strong methods.

Of the 1017 participants in the study, 53% self-identified as White; 21% were postgraduate year (PGY)-1, 20% PGY-2, 60% PGY-3+, 19% in a surgical specialty, and no differences found at baseline between intervention and control groups. At the start, scores suggested a high prevalence of burnout, imposter syndrome, moral injury, and low flourishing,

Only 40% of participants did the post-survey. With this limitation, BT participants experienced notable improvements in MBI subscales. For overall burnout, participants had an 18% (95% CI 5% to 30%) reduction in burnout, with 53% lower odds of experiencing burnout at follow-up. The number needed to treat (NNT), to go from positive to negative burnout score, was 11 (95% CI 7.1 to 22.4). There were similarly positive findings in the other scales, such as NNT of 9 (95% CI 6.8 to 17.8) to go from positive to negative on the imposter scale. The sizes of improvement reported here were generally higher than those found in prior single site studies.

For feasibility, the authors report approximately 5 hours total, weekly, for coaches. Time to become a certified Life Coach was not provided; a Google search offered many different courses, costs, and times, from 90 days to 6 months. The authors describe scalability by comparing the group program to their single-site pilot: using the online group format yielded a 10-fold increase in participants at 26 sites and required 30 additional (total) coach hours.

My take home: With voluntary participants, low post-survey completion, and 4 months duration, I'm not ready to conclude that this program should be offered at every GME institution. But this approach appears more feasible than single-program or site interventions and should spur additional projects, especially for smaller or more remote institutions and specialties with shortages of coaches. I hope that this well-designed protocol will not be degraded by commercial entities focused on profit. I herald the University of Colorado for financially supporting this work.

Honorable Mention

Drs. Arjun Chen and Drake Chen examined the value of a generative pretrained chatbot (ChatGPT) for creating content for learning health systems.¹¹ They used question prompts for broad as well as specific topics (eg, build a stroke index) and then

asked for supporting references. The authors exhaustively checked each reference and reported error rates for GPT-3.5 and GPT-4 models. With the ChatGPT-3.5 model, an astonishing 98% of the references were fake vs with ChGPT-4, where 21% of references were fake. The authors noted that the narrower the focus of the topic, the greater the number of fake references. Also, more recent (real) references were usually missing. The authors remind us that chatbots can “answer any question” but cannot “fact check” their own responses.¹¹

Lainie Yarris' Pick

Despite a large body of health professions education literature about the importance of feedback in improving performance, educators and learners still struggle to engage in this complex practice in a way that feels helpful to them, in real-time clinical environments. Our understanding of feedback has evolved over the past several decades, with a shift in emphasis from feedback delivery to feedback receptivity and incorporation. Furthermore, recent work has emphasized the importance of learner factors, emotions, relationships, and culture in feedback effectiveness. In “Meaning Making About Performance: A Comparison of Two Specialty Cultures,” Bearman and colleagues explore the role of culture in feedback processes by interviewing trainees in surgery and intensive care medicine about their experiences with feedback.¹² In this constructivist grounded theory qualitative study, the authors ask how trainees come to understand their performance, across 2 specialty cultures, and explore what role feedback conversations play in this process.

The authors find commonalities across the 2 specialties. First, trainees actively seek information about their performance. Second, they understand their progress through both explicit and tacit cues, ranging from direct guidance to analyzing their internal emotional response. Finally, they understand overall progress mostly from patching together these cues, rather than from feedback conversations. However, despite both specialties taking place in acute care environments and employing high-stakes procedural skills, there were notable divergences between the 2 fields in how trainees made meaning about the quality of their performance and the role of feedback conversations. Surgical trainees were more likely to use patient outcomes and procedural skill feedback as cues for their performance quality, receive feedback based on direct observation, and describe their performance with more certainty than intensive care trainees. Intensive care trainees found value in supervisor emotional validation and support and experienced more ambiguity about their performance than surgical trainees.

In highlighting the differences between specialty cultures and the variable role that feedback conversations play in the meaning trainees make about their performance, the authors do not imply that educators should aim to change specialty culture. Rather, they suggest that programs consider where, when, and how feedback interactions in their specific culture result in trainee meaning-making about their performance.

My take home: In this rigorous qualitative study, Bearman et al contribute to the conversations about the role specialty culture plays in feedback and how trainees make meaning of feedback to understand their performance. This work suggests that there are large gaps in the ways trainees understand their performance in actual feedback situations versus the ideal of CBME. In CBME, trainees understand their performance based on myriad workplace-based assessments as well as longitudinal conversations with coaches and advisors. This paper reminds us that there is much we still don't understand about how to implement feedback that results in actionable trainee meaning-making about performance. Program directors can directly apply these results by considering the reality of their specialty and culture: what works currently to guide trainees in meaning-making, and what opportunities are still untapped in their specific programs?

Honorable Mention

In “When I Say... Burnout,” Khan, Hodges, and Martimianakis discuss modern definitions and usages of the term “burnout” and pose that the current widespread use of the term offers little insight into individual experiences of burnout.¹³ The paper includes a succinct review of 3 evolutionary constructs of burnout: (1) as an individual illness; (2) as a syndrome resulting from workplace or system stresses; and (3) as an existential crisis. As Khan et al describe how burnout constructs “exist both in the semantic space of linguistics (as a word in the vernacular) and in the dynamic space of the imagination (based on the meaning that is ascribed to it)” they illustrate that what is missing in our current lexicon is illumination of the underlying concerns or contributing factors to the experience of burnout.¹³ This thought-provoking read is of interest to educators, burnout researchers, and those touched by burnout—in other words, all of us.

References

1. Sawatsky AP, Matchett CL, Hafferty FW, et al. Professional identity struggle and ideology: a qualitative study of residents' experiences. *Med Educ.* 2023;57(11):1092-1101. doi:10.1111/medu.15142
2. Kendrick DE, Thelen AE, Chen X, et al. Association of surgical resident competency ratings with patient outcomes. *Acad Med.* 2023;98(7):813-820. doi:10.1097/ACM.0000000000005157
3. Montgomery B, Kelsey B, Lindeman B. Using graduating surgical resident milestone ratings to predict patient outcomes: a blunt instrument for a complex problem. *Acad Med.* 2023;98(7):765-768. doi:10.1097/ACM.0000000000005165
4. Farrell L, Cuncic C, Hartford W, Hatala R, Ajjawi R. Goal co-construction and dialogue in an internal medicine longitudinal coaching programme. *Med Educ.* 2023;57(3):265-271. doi:10.1111/medu.14942
5. Lees AF, Beni C, Lee A, et al. Uses of electronic health record data to measure the clinical learning environment of graduate medical education trainees: a systematic review. *Acad Med.* 2023;98(11):1326-1336. doi:10.1097/ACM.0000000000005288
6. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. doi:10.1136/bmj.n71
7. Xiao Y, Watson M. Guidance on conducting a systematic literature review. *J Plan Educ Res.* 2019;39(1):93-112. doi:10.1177/0739456X17723971
8. Maggio LA, Samuel A, Stellerrecht E. Systematic reviews in medical education. *J Grad Med Educ.* 2022;14(2):171-175. doi:10.4300/JGME-D-22-00113.1
9. Simpson D, Sullivan GM, Artino AR Jr, Deiorio NM, Yarris LM. Envisioning graduate medical education in 2030. *J Grad Med Educ.* 2020;12(3):235-240. doi:10.4300/JGME-D-20-00292.1
10. Mann A, Shah AN, Thibodeau PS, et al. Online well-being group coaching program for women physician trainees. A randomized clinical trial. *JAMA Netw Open.* 2023;6(10):e2335541. doi:10.1001/jamanetworkopen.2023.35541
11. Chen A, Chen DO. Accuracy of chatbots in citing journal articles. *JAMA Netw Open.* 2023;6(8):e2327647. doi:10.1001/jamanetworkopen.2023.27647
12. Bearman M, Ajjawi R, Castanelli D, et al. Meaning making about performance: a comparison of two specialty feedback cultures. *Med Educ.* 2023;57(11):1010-1019. doi:10.1111/medu.15118
13. Khan R, Hodges BD, Martimianakis MA. When I say ... burnout. *Med Educ.* 2023;57(8):704-705. doi:10.1111/medu.15088



Gail M. Sullivan, MD, MPH, is Editor-in-Chief, *Journal of Graduate Medical Education* (JGME), and Associate Director for Education, Center on Aging, and Professor of Medicine, University of Connecticut Health Center, Farmington, Connecticut, USA; **Deborah Simpson, PhD**, is Deputy Editor, JGME, and Director of

EDITORIAL

Education, Academic Affairs at Advocate Aurora Health, and Clinical Adjunct Professor of Family & Community Medicine, Medical College of Wisconsin, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA; **Anthony R. Artino Jr, PhD**, is Deputy Editor, JGME, and Professor and Associate Dean for Evaluation and Educational Research, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA; **Nicole M. Deiorio, MD**, is Executive Editor, JGME, Professor, Department of Emergency

Medicine, and Associate Dean, Student Affairs, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA; and **Lalena M. Yarris, MD, MCR**, is Deputy Editor, JGME, and Professor of Emergency Medicine, Oregon Health & Science University, Portland, Oregon, USA.

Corresponding author: Gail M. Sullivan, MD, MPH, University of Connecticut Health Center, Farmington, Connecticut, USA, gsullivan@uchc.edu