

The medical licensing examination debate

ABSTRACT

National licensing examination are typically large scale examinations taken in early career or near the point of graduation, and importantly where success is required to subsequently be able to practice. They are becoming increasingly popular as a way of quality assuring the medical workforce but debate about their contribution to patient safety and improving healthcare outcomes continues.

A systematic review of the national licensing examination literature demonstrates that there is disagreement between assessment experts about the strengths and challenges. This is characterized by a trans-Atlantic divide between the dominance of psychometric reliability assurance in North America and the wider interpretations of validity, to include consequences, in Europe. We conclude that the debate might benefit from refocusing to what national licensing examination should assess; trying to achieve a balance between assessing a breadth of skills and the capacity for such skills in practice and worrying less about reproducibility.

5 keywords: national licensing examination, validity, international medical graduates, testing, assessment

INTRODUCTION

Healthcare regulators around the world are charged with protecting the public and supporting standards within the healthcare professions. One approach to this is the adoption of a national licensing examination (NLE). Typically this is a large scale examination taken in early career or near the point of graduation, and importantly where success is required to subsequently be able to practice.

The adoption and delivery of these examinations varies around the world. Some countries have established NLEs exclusively for their home graduates as a way of quality assuring both their graduates and their medical schools. These include Switzerland, Poland, Bahrain, Qatar, Croatia and Germany (Seyfarth et al., 2010). Others ask all prospective doctors who hope to work in their jurisdiction to undergo a NLE process such as the United States, Canada, Hong Kong, Japan, Korea, Chile, and the United Arab Emirates, with North America dominating the literature and the methodologies (Sutherland, 2006).

A third distinct group of countries exclusively involve international medical graduates (IMGs), who undergo an additional examination alongside an accreditation process, where they are required to provide evidence of language competence and validated documentation of their primary qualifications, in order to practice in the new country. These include Australasia, parts of Europe (Sonderren et al., 2009), including Sweden (Musoke, 2012), and the United Kingdom (UK) (Kovacs et al., 2014).

However such systems are not without implications. There is a worldwide shortage of physicians (Leitch and Dovey, 2010) and so additional examination hurdles have workforce planning implications. NLEs may also have implications for candidates if they achieve scores

in the lowest quartile resulting in poorer career prospects for both home graduates and IMGs. In contrast, those who get the highest scores are likely to secure the best posts (Kenny et al., 2013, Green et al., 2009, Noble, 2008). NLEs may also have an impact on the health of graduates, many of whom complain of 'stress and burnout' as a result of the process (McMahon and Tallia, 2010).

In 2015 the General Medical Council (GMC), which regulates doctors in the UK, announced that they were planning to develop and ultimately implement a new Medical Licensing Assessment by 2021 (General Medical Council, 2015a). This will transform the UK from only currently assessing IMGs if they wish to enter the UK to practice (General Medical Council, 2015b), to assessing all doctors who wish to practice – although as European Law stands, EEA and Swiss graduates, and EC rights holders will remain exempt under freedom of movement legislation (Gullard A, 2015); although the GMC has stated that "it is our strong aspiration also to cover European nationals and we will continue to explore options' (Hulf J and Hart M, 2015).

As part of the process of developing this new NLE, the GMC commissioned researchers at the XXX to undertake a systematic review of the international literature (XXX). As part of this review we were interested in exploring the debate surrounding NLEs as it is an area not without policy controversy. We present this debate and argue that the adoption and focus of NLEs is as much a socio-political venture as one about patient safety.

METHODS

We conducted an extensive appraisal of the literature following accepted guidance for narrative synthesis within systematic reviews (Popay J et al., 2006).

We carried out a structured search in the following databases: Embase (Ovid Medline); Medline (EBSCO); PubMed; Wiley Online; ScienceDirect; and PsychINFO. We used a combination of keywords which we believed would yield the best results following an initial scoping exercise. These included: national licensing examination; doctor; dentist, nurse; midwife; healthcare professional; international medical graduate; accreditation; credentialing; registration; and certification.

One author (NL) undertook a first screening of potentially relevant papers based on titles and their abstracts, and then four authors (NL, JA, MR and LC) independently performed the final selection of included articles based on reading their full text. Consensus between the reviewers was supported by all the authors. A review of websites of medical regulators or those bodies responsible for licensing healthcare professionals including doctors in 49 'very high human development' countries was undertaken (UNDP, 2014). The United Nations Development Programme (UNDP) calculates 'Human Development' by evaluating: life expectancy at birth, mean years of schooling, expected years of schooling, gross national income per capita. Countries are ranked as having 'very high', to 'low' human development and the UK is in the 'very high' category (UNDP, 2014).

We included all papers in our review that discussed national or large regional (State level) examinations for medical and healthcare professionals published between January 2005 and April 2015 within the 'very high' category of nations (UNDP, 2014). Papers were eligible if they were linked to examinations in early career or near the point of graduation from school, and importantly that a successful outcome was required to subsequently be able to practice. We placed no restriction on language but specialist examinations were not included.

Using a standardized data extraction form, two reviewers (NL with one of JA, MR and LC) independently extracted study characteristics from the included papers. We placed papers into one of two categories; either that they were important opinion based pieces that helped to shape the debate around NLEs or that they demonstrated validity evidence, as defined by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) framework as reported by Downing (2003).

RESULTS

Our initial search produced 202 documents. We reviewed the abstracts for relevance and the total number of reports was reduced to 103 papers. One researcher (NL) then reviewed their full text with a second review performed by at least one other researcher (JA, MR, LC). Thirty papers were excluded as they did not meet the inclusion and exclusion criteria at this stage. There were 73 papers in the final review prior to a framework mapping exercise.

Of these 73 papers, when mapped to the validity framework, 50 papers were found to be important in terms of shaping the debate through informed opinion and editorials. The other 23 articles discussed the evidence for NLEs and are reported elsewhere. The overall review process is summarized in figure 1.

We found that the debate around national and other large scale licensing examinations is polarized (Ferris, 2006, Neilson, 2008) with the arguments for and against limited to a relatively small number of core themes based on a limited number of papers (Harden, 2009, Melnick, 2009, Ricketts and Archer, 2008).

National licensing examinations are not new and indeed some suggest that their longevity is in itself now an endorsement of their value (Melnick, 2009, Neumann and Macneil, 2007). Proponents argue that testing everyone to a minimum required standard (Cosby Jr, 2006), which all graduates must achieve to enter their chosen profession must be 'fair' (Melnick, 2009). After all, reliability is improved by standardization and structure (Van der Vleuten C, 1996). Without all candidates undertaking the same examination one cannot fairly discriminate between them (Lee, 2008); as, for example, it is known that medical students perform differently in large scale testing depending on which medical school they attended (Boursicot KAM et al., 2007), and that this difference cannot be entirely explained by underlying student ability (McManus et al., 2008, McManus and Wakeford, 2014). In other words differences in medical school training matter. So NLEs might help to support a common medical curriculum and provide an opportunity for this to be better informed by stakeholders including patients and employers (Wass, 2005).

However, there is a growing body of evidence that home graduates, familiar with the language and systems, do consistently better than graduates trained elsewhere (Tiffin et al., 2014, Holtzman et al., 2014). The Anglo-French model employed in Canada provides evidence that language should not be a barrier, and if a practitioner cannot cope with the language and cultural differences of the examination process they may not be well placed to cope with the local culture of healthcare delivery. If fairly administered, NLEs should promote the movement of doctors (Archer, 2009), though some believe that such examinations do little to assist in increasing the mobility of health professionals (Philipsen and Haynes, 2007, Cooper, 2005). There is also an argument that additional layers of intra- and inter-state regulation involving certification, credentialing, and accreditation

interwoven with regulatory politics makes it very complicated for practitioners and public alike (Rehm and DeMers, 2006).

Advocates of NLEs infer that the standards associated with the testing process are improved due to the logistics of staging large scale examinations which necessarily require resources and expertise to be pooled (Neumann and Macneil, 2007, Lehman and Guercio, 2013).

Furthermore, it may also be advantageous financially by economies of scale (Lehman and Guercio, 2013).

There is little doubt that while educational assessment and theory evolves at a rapid pace, most specialists regard the United States Medical Licensing Examination (USMLE) and Medical Council of Canada Qualifying Examination (MCCQE) as world-leading examples in licensing examinations (Norcini et al., 2014, Bajammal et al., 2008, Lillis, 2012). This claim is supported by good empirical evidence (Lillis, 2012, USMLE CEUP, 2008, Hecker and Violato, 2008, Margolis et al., 2010, Stewart et al., 2005, Guttormsen et al., 2013). Yet difficulties arise from the additional claims that national licensing examination assures patient safety and that this form of assessment directly leads to better patient care (McMahon and Tallia, 2010, Tamblyn et al., 2007, Wenghofer et al., 2009, Stewart et al., 2005). While Norcini *et al.* (2014) has established a correlation between USMLE scores and some specific patient outcomes and Tamblyn *et al.* (2007) a correlation with later complaints, there is no evidence as yet to establish causality (Sutherland, 2006, Boulet and van Zanten, 2014).

The arguments that licensing examinations make patients safer, some medical educationalists suggest, rests on assumptions rather than on evidence. In fact, it could even be damaging by standardizing medical education and therefore reducing innovation and advancements in curricula (van der Vleuten, 2009, Harden, 2009, Gorsira, 2009).

Certainly a number of assumptions have been made about NLEs. First, Melnick (2009) argues that the public are safer because assessment specialists provide a credible external audit of the quality of NLEs and therefore healthcare workers that are allowed to practice. Second, the assessment specialists are able to accurately establish and assess minimum standards of knowledge and competence (Melnick, 2009). Lastly, some reason that the statistical correlation between NLE scores, patient care outcomes, and the incidence of disciplinary action in later professional life is evidence that NLEs directly lead to better care (Tamblyn et al., 2007, Wenghofer et al., 2009).

Opponents counter that NLEs are actually outdated and that the social, educational, and professional contexts are different to when, for example, the USMLE was first established. “We see a system designed over 100 years ago to solve a problem that no longer exists – proprietary diploma mills that had no educational standards, or accreditation” (Ferris, 2006 p.129). Ferris (2006) argues that we should be testing competence in the workplace over time and not at a point-in-time. Others agree stating that “the best preparation for the practice of dentistry is the practice of dentistry” (Calnon, 2006). Opponents of licensing examinations also suggest that those learning outcomes that are easily testable become the focus of such examinations and that those competences needed more for actual practice are often missing from the blueprint (Harden, 2009, Noble, 2008, Neilson, 2008). Their answer is for more job assessment, appraisal, and professional development to provide more accurate and up-to-date and importantly ongoing evidence of practitioner competence (Kovacs et al., 2014, Waldman and Truhlar, 2013, Calnon, 2006).

Neilson (2008) develops this argument. “The standardisation of final, licensing, and fitness to practice examinations may make educationalists weep with joy, but there is no clear

evidence that it makes for better doctors. My colleagues and I deal with the immediate postgraduate training of juniors and know that, regardless of where the doctors have qualified, their practical education starts when they start working with patients for real.”

Ultimately some believe NLEs are not real barriers to practice anyway and therefore do not protect the public as, for example, nearly everyone who takes USMLE passes it in the end (Margolis et al., 2010). They also point out that while there is evidence that those who do well in early career examinations go on to get the best jobs (Green et al., 2009, Kenny et al., 2013), there are studies that argue that trying to predict which doctors may appear in disciplinary hearings or administer poor patient care is subject to a myriad of other variables and consequences (Harden, 2009, Norman, 2015a). Finally many proponents themselves point out that those doctors who are subsequently identified and disciplined still initially passed the NLE (Tamblyn et al., 2007, Wenghofer et al., 2009, Norcini et al., 2014), and that they are expensive to run (Brown et al.).

Transatlantic divide

We discovered in the literature a distinct North American / European divide in the arguments. North America has the well-established USMLE and MCCQE (Norcini et al., 2014, Bajammal et al., 2008, Lillis, 2012), with the literature focusing on the psychometric / measurement aspects of the examinations (Lillis, 2012, USMLE CEUP, 2008, Hecker and Violato, 2008, Margolis et al., 2010, Stewart et al., 2005, Guttormsen et al., 2013). While in Europe it is more commonly argued that the challenges and practicalities of introducing an equivalence within or across European boundaries would be considerable (Harden, 2009), despite imperatives that ensure citizens are able to freely move and work (de Vries, 2009). Securing a consensus whereby *all* doctors in Europe might sit a pan-European licensing

examination would be very challenging (Gorsira, 2009). The arguments against a European equivalent though are not just logistical as we found that European experts worry about the wider picture. Harden (2009) argues that large scale examinations tend to target learning outcomes that can be easily assessed and that these are rarely related to the overall competence required of a doctor, and that NLEs depress change and innovation while failing to recognize diversity. Although van de Vleuten (2009) concludes that a NLE might work in Europe he explains that: "...we need to start thinking very carefully about how qualifying systems could be set up to achieve the desired effects without doing too much harm to learning and to innovation power..." (van der Vleuten, 2009 p191).

DISCUSSIONS AND CONCLUSIONS

The ultimate aim for educators with researchers is to be able to select, train and continue to support the best professionals across the continuum of healthcare careers. This results in common checks of performance at key stages in career paths. There is much to be said for national licensing examinations. In assessment terms, the pooling of expertise to produce the highest quality assessment process that can drive up standards and improve patient safety makes sense. An examination that is delivered to all possible candidates also has many psychometric strengths and should produce reliable results, but it does not appear to be as simple as that. Candidates are not constant in their traits or abilities over time and good people can perform badly in difficult settings (Weiner B, 2013). In statistical terms, subsequent poor performance is rare making reliable predictions almost impossible (Norman, 2015a), and while recognized experts argue over the various merits of national licensing examination, it is perhaps the discourse of a transatlantic divide that is of the most interest.

North American has a long history of large scale testing in healthcare; the National Board of Medical Examiners was founded in 1913. The testing culture is grounded and further reinforced by the perception of an ever growing litigious society. In Europe, at least historically, more assessment decisions have been locally developed and delivered and quality assurance has been based on a peer review model of inspections.

This difference is important as it speaks to the very heart of assessment methodologies and the psychosocial constructs of validity. A useful framework, developed and revised by US based educational organizations, divides validity into 5 categories: content; response process; internal structure; relationship to other variables; and consequences (Downing, 2003). Yet the emphasis for collecting and reporting validity evidence from across this framework appears to vary. North American psychometricians have traditionally focused on internal structure which deals with the assessments' statistical performance including reliability (Brannick et al., 2011). The reproducibility of a test is a fundamental part of validity evidence supported by a concrete statistical group of methods that can withstand legal challenge (Cavanaugh SH, 1991).

Other areas of validity start to require mixed methods and draw on theoretical constructs of what is validity, what are we trying to assess and what might the impacts of the intervention itself be? Such developments have not been universally welcomed by some North American experts (Norman, 2015b).

Indeed there is a balance if not a conflict between these. Clinical examinations, often a core component of NLEs and commonly in the form of Objective Structured Clinical Examinations (OSCEs), are sometimes criticized for sacrificing content for reliability, taking candidates

away from the realities of healthcare into such controlled environments that candidates fail to demonstrate the very skills we are trying to assess (McEvoy et al., 2014).

We would argue that the NLE debate is part of a wider ideological battle between standardization and contextualized reality construction, the two 'sides' of validity. Within the validity literature, while reliability and other validity constructs inhabit the same framework, there is a conflict that is fundamentally shaped by socio-political dimensions. With a background of a broader move, under neo-liberalism as the state seeks to promote its control on professionals from a distance (Rose N and Miller P, 1992), we increasingly witness a desire to centralize and standardize assessment methodologies so they run autonomously with self-regulation. This is part of an approach to assure an increasingly risk-averse public that professionals are safe and fit-for-purpose. This drive puts pressure on the need to secure reliable results that can be 'trusted' but at the risk of an unintended detrimental impact on other aspects of validity. The result is an assessment that is robust but potentially irrelevant as it struggles to assess constructs that are of real interest, such as professionalism and capability.

With examinations tending to focus on applied knowledge, it is perhaps not surprising that those who end up in front of disciplinary boards are mainly there with concerns about their professional behavior and not knowledge (Tamblyn et al., 2007).

The debate might benefit from refocusing from whether or not we should have a NLE to what they should assess. Understanding better the assessment of professionals in the workplace, as opposed to simulated environments such as OSCEs, might take us in the right direction. We ultimately would benefit from trying to achieve a balance between assessing a

breadth of skills and the capacity for such skills in practice and worrying less about psychometric reproducibility.

Ultimately research needs to focus on if NLEs add *uniquely* to assuring patient safety. Jolly (2016) in his recent opinion piece argues that in order to be convinced of the positive impact of NLEs we should seek to ascertain their “added value” on existing assessment techniques and explore potentially more cost-effective alternatives to the NLE such as a more “beefed up” accreditation of the existing assessment process” (Jolly, 2016).

We know that performance in NLEs can be predicted by prior examination success and they predict future success (Hecker and Violato, 2008, Ranney, 2006, Stewart et al., 2005, Tiffin et al., 2014). We know that those who do better in them deliver better care, at least in some specific domains (Norcini et al., 2014), so testing is undoubtedly important. However testing specifically using NLEs remains up for debate as there is a lack of unequivocal evidence to either support or refute their use (Boulet and van Zanten, 2014).

Ethical approval: was not required

REFERENCE LIST

XXX

ARCHER, J. C. 2009. European licensing examinations – The only way forward. *Medical Teacher*, 31, 215-216.

BAJAMMAL, S., ZAINI, R., ABUZNADAH, W., AL-RUKBAN, M., ALY, S., BOKER, A., AL-ZALABANI, A., AL-OMRAN, M., AL-HABIB, A., AL-SHEIKH, M., AL-SULTAN, M., FIDA, N., ALZHRANI, K., HAMAD, B., AL SHEHRI, M., ABDULRAHMAN, K., AL-DAMEGH, S., AL-NOZHA, M. & DONNON, T. 2008. The need for national medical licensing examination in Saudi Arabia. *BMC Medical Education*, 8, 53.

BOULET, J. & VAN ZANTEN, M. 2014. Ensuring high-quality patient care: the role of accreditation, licensure, specialty certification and revalidation in medicine. *Medical Education*, 48, 75-86.

BOURSICOT KAM, ROBERTS T & PELL G 2007. Using borderline methods to compare passing standards for OSCEs at graduation across three medical schools. *Medical Education*, 41, 1024.

BRANNICK, M. T., EROL-KORKMAZ, H. T. & PREWETT, M. 2011. A systematic review of the reliability of objective structured clinical examination scores. *Medical Education*, 45, 1181-1189.

BROWN, C., ROSS, S., CLELAND, J. & WALSH, K. Money makes the (medical assessment) world go round: The cost of components of a summative final year Objective Structured Clinical Examination (OSCE). *Medical Teacher*, 0, 1-7.

CALNON, W. R. 2006. The Residency Pathway to Dental Licensure: The Paradigm Shift from Inception to Policy. *Journal of Evidence Based Dental Practice*, 6, 138-142.

CAVANAUGH SH 1991. Response to a legal challenge. Five steps to defensible credentialing examinations. *Eval Health Prof*, 14, 13-40.

COOPER, S. L. 2005. The licensure mobility experience within the United States. *Optometry - Journal of the American Optometric Association*, 76, 347-352.

COSBY JR, J. C. 2006. The American Board of Dental Examiners Clinical Dental Licensure Examination: A Strategy for Evidence-Based Testing. *Journal of Evidence Based Dental Practice*, 6, 130-137.

DE VRIES, H., SANDERSON, P., JANTA, B., RABINOVICH, L., ARCHONTAKIS, F., ISMAIL, S., KLAUTZER, L., MARJANOVIC, S., PATRUNI, B., PURI, S., TIESSEN, J. 2009. International Comparison of Ten Medical Regulatory Systems.

DOWNING, S. M. 2003. Validity: on the meaningful interpretation of assessment data. *Medical Education*, 37, 830-837.

- FERRIS, R. T. 2006. A Sea-Change in American Dentistry: A National Clinical Licensure Examination, or a Residency-Based Pathway? *Journal of Evidence Based Dental Practice*, 6, 129.
- GENERAL MEDICAL COUNCIL. 2015a. *GMC Council approves development of UK medical licensing assessment* [Online]. <http://www.gmc-uk.org/news/26549.asp>: General Medical Council. [Accessed 27 July 2015].
- GENERAL MEDICAL COUNCIL. 2015b. *Professional and Linguistic Assessments Board (PLAB)* [Online]. <http://www.gmc-uk.org/doctors/plab.asp>: General Medical Council. [Accessed 27 July 2015].
- GORIRA, M. 2009. The utility of (European) licensing examinations. AMEE Symposium, Prague 2008. *Medical Teacher*, 31, 221-222.
- GREEN, M., JONES, P. & THOMAS, J. X. J. 2009. Selection Criteria for Residency: Results of a National Program Directors Survey. *Academic Medicine*, 84, 362-367.
- GULLARD A. 2015. GMC proposes single test for all doctors wishing to work in UK. *BMJ Careers* [Online], http://careers.bmj.com/careers/advice/GMC_proposes_single_test_for_all_doctors_wishing_to_work_in_UK.
- GUTTORMSEN, S., BEYELER, C., BONVIN, R., FELLER, S., SCHIRLO, C., SCHNABEL, K., SCHURTER, T. & BERENDONK, C. 2013. The new licencing examination for human medicine: from concept to implementation. *Swiss Med Wkly*, 143, 1-10.
- HARDEN, R. M. 2009. Five myths and the case against a European or national licensing examination. *Medical Teacher*, 31, 217-220.
- HECKER, K. & VIOLATO, C. 2008. How Much Do Differences in Medical Schools Influence Student Performance? A Longitudinal Study Employing Hierarchical Linear Modeling. *Teaching and Learning in Medicine*, 20, 104-113.
- HOLTZMAN, K. Z., SWANSON, D. B., OUYANG, W., DILLON, G. F. & BOULET, J. R. 2014. International Variation in Performance by Clinical Discipline and Task on the United States Medical Licensing Examination Step 2 Clinical Knowledge Component. *Academic Medicine*, Publish Ahead of Print, 10.1097/ACM.0000000000000488.
- HULF J & HART M 2015. Taking forward work on a UK licensing assessment. http://www.gmc-uk.org/10___Taking_forward_work_on_a_UK_licensing_assessment.pdf_61114454. pdf: General Medical Council,.
- JOLLY, B. 2016. National licensing exam or no national licensing exam? That is the question. *Medical Education*, 50, 12-14.
- KENNY, S., MCINNES, M. & SINGH, V. 2013. Associations between residency selection strategies and doctor performance: a meta-analysis. *Medical Education*, 47, 790-800.

- KOVACS, E., SCHMIDT, A. E., SZOCSKA, G., BUSSE, R., MCKEE, M. & LEGIDO-QUIGLEY, H. 2014. Licensing procedures and registration of medical doctors in the European Union. *Clinical Medicine, Journal of the Royal College of Physicians of London*, 14, 229-238.
- LEE, Y. S. 2008. OSCE for the Medical Licensing Examination in Korea. *Kaohsiung Journal of Medical Sciences*, 24, 646-650.
- LEHMAN, E. P. & GUERCIO, J. R. 2013. The Step 2 Clinical Skills Exam — A Poor Value Proposition. *New England Journal of Medicine*, 368, 889-891.
- LEITCH, S. & DOVEY, S. M. 2010. Review of registration requirements for new part-time doctors in New Zealand, Australia, the United Kingdom, Ireland and Canada. *Journal of primary health care*, 2, 273-280.
- LILLIS, S., STUART, M., SIDONIE, TAKAI, N. 2012. New Zealand Registration Examination (NZREX Clinical): 6 years of experience as an Objective Structured Clinical Examination (OSCE). *The New Zealand Medical Journal*, 125, 74 - 80.
- MARGOLIS, M. J., CLAUSER, B. E., WINWARD, M. & DILLON, G. F. 2010. *Validity Evidence for USMLE Examination Cut Scores: Results of a Large-Scale Survey. [Miscellaneous Article]*, *Academic Medicine* October 2010;85(10) Supplement, RIME: Proceedings of the Forty-Ninth Annual Conference November 7-November 10, 2010:S93-S97.
- MCEVOY, M. D., DEWAAY, D. J., VANDERBILT, A., ALEXANDER, L. A., STILLEY, M. C., HEGE, M. C. & KERN, D. H. 2014. Are Fourth-Year Medical Students as Prepared to Manage Unstable Patients as they are to Manage Stable Patients? *Academic medicine : journal of the Association of American Medical Colleges*, 89, 618-624.
- MCMAHON, G. T. & TALLIA, A. F. 2010. Perspective: Anticipating the challenges of reforming the United States medical licensing examination. *Academic Medicine*, 85, 453-6.
- MCMANUS, I. C., ELDER, A. T., DE CHAMPLAIN, A., DACRE, J. E., MOLLON, J. & CHIS, L. 2008. Graduates of different UK medical schools show substantial differences in performance on MRCP(UK) Part 1, Part 2 and PACES examinations. *BMC Med*, 6, 5.
- MCMANUS, I. C. & WAKEFORD, R. 2014. PLAB and UK graduates' performance on MRCP(UK) and MRCGP examinations: data linkage study. *BMJ*, 348, g2621.
- MELNICK, D. E. 2009. Licensing examinations in North America: Is external audit valuable? *Medical Teacher*, 31, 212-214.
- MUSOKE, S. B. 2012. Foreign Doctors and the Road to a Swedish Medical License Experienced barriers of doctors from non-EU countries.
- NEILSON, R. 2008. Authors have missed gap between theory and reality. *British Medical Journal*, 337.

- NEUMANN, L. M. & MACNEIL, R. L. 2007. Revisiting the National Board Dental Examination. *Journal of dental education*, 71, 1281-1292.
- NOBLE, I. S. G. 2008. Are national qualifying examinations a fair way to rank medical students? No. *British Medical Journal*, 337.
- NORCINI, J. J., BOULET, J. R., OPALEK, A. & DAUPHINEE, W. D. 2014. The Relationship Between Licensing Examination Performance and the Outcomes of Care by International Medical School Graduates. *Academic Medicine*, 89, 1157-1162
10.1097/ACM.0000000000000310.
- NORMAN, G. 2015a. Identifying the bad apples. *Advances in Health Sciences Education*, 20, 299-303.
- NORMAN, G. 2015b. The negative consequences of consequential validity. *Advances in Health Sciences Education*, 20, 575-579.
- PHILIPSEN, N. C. & HAYNES, D. 2007. The Multi-State Nursing Licensure Compact: Making Nurses Mobile. *The Journal for Nurse Practitioners*, 3, 36-40.
- POPAY J, ROBERTS H, SOWDEN A, PETTICREW M, ARAI L, RODGERS M, BRITTEN N, ROEN K & DUFFY S 2006. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews. *ESRC Methods Programme*.
http://www.researchgate.net/profile/Mark_Rodgers4/publication/233866356_Guidance_on_the_conduct_of_narrative_synthesis_in_systematic_reviews_A_product_from_the_ESRC_Methods_Programme/links/02e7e5231e8f3a6183000000.pdf.
- RANNEY, R. R. 2006. What the Available Evidence on Clinical Licensure Exams Shows. *Journal of Evidence Based Dental Practice*, 6, 148-154.
- REHM, L. P. & DEMERS, S. T. 2006. Licensure. *Clinical Psychology: Science and Practice*, 13, 249-253.
- RICKETTS, C. & ARCHER, J. 2008. Are national qualifying examinations a fair way to rank medical students? Yes. *BMJ*, 337.
- ROSE N & MILLER P 1992. Political Power Beyond the State: Problematics of Government. *British Journal of Sociology*, 43, 173-205.
- SEYFARTH, M., REINCKE, M., SEYFARTH, J., RING, J. & FISCHER, M. R. 2010. Grades on the Second Medical Licensing Examination in Germany Before and After the Licensing Reform of 2002: A study in Two Medical Schools in Bavaria. *Dtsch Arztebl International*, 107, 500-4.
- SONDEREN, M. J., DENESSEN, E., CATE, O. T. J. T., SPLINTER, T. A. W. & POSTMA, C. T. 2009. The clinical skills assessment for international medical graduates in the Netherlands. *Medical Teacher*, 31, e533-e538.

- STEWART, C. M., BATES, R. E. & SMITH, G. E. 2005. Relationship Between Performance in Dental School and Performance on a Dental Licensure Examination: An Eight-Year Study. *Journal of Dental Education*, 69, 864-869.
- SUTHERLAND, K. L., S. 2006. Regulation and Quality Improvement a review of the evidence. *Health Foundation*.
- TAMBLYN, R., ABRAHAMOWICZ, M., DAUPHINEE, D. & ET AL. 2007. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA*, 298, 993-1001.
- TIFFIN, P. A., ILLING, J., KASIM, A. S. & MCLACHLAN, J. C. 2014. Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. *BMJ*, 348, g2622.
- UNDP 2014. Human Development Report 2014 Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience
- USMLE CEUP 2008. Comprehensive Review of USMLE Summary of Final Report and Recommendations.
- VAN DER VLEUTEN C 1996. The assessment of professional competence. *Adv Health Sci Educ*, 1, 41-67.
- VAN DER VLEUTEN, C. P. M. 2009. National, European licensing examinations or none at all? *Medical Teacher*, 31, 189-191.
- WALDMAN, H. B. & TRUHLAR, M. R. 2013. Impact of residency requirement for dental licensure: an update. *The New York state dental journal*, 79, 30-32.
- WASS, V. 2005. Ensuring medical students are "fit for purpose". *BMJ*, 331, 791-792.
- WEINER B 2013. Drive Theory. In: WEINER B (ed.) *Human Motivation*. New Jersey: Lawrence Erlbaum Associates, Inc. Publishers.
- WENGHOFER, E., KLASS, D., ABRAHAMOWICZ, M., DAUPHINEE, D., JACQUES, A., SMEE, S., BLACKMORE, D., WINSLADE, N., REIDEL, K., BARTMAN, I. & TAMBLYN, R. 2009. Doctor scores on national qualifying examinations predict quality of care in future practice. *Medical Education*, 43, 1166-1173.