

Supervisor competence as an assessor of medical trainees

Citation for published version (APA):

McGill, D. A. (2019). *Supervisor competence as an assessor of medical trainees: evaluating the validity and quality of supervisor assessments*. [Doctoral Thesis, Maastricht University]. ProefschriftMaken Maastricht. <https://doi.org/10.26481/dis.20190529dm>

Document status and date:

Published: 01/01/2019

DOI:

[10.26481/dis.20190529dm](https://doi.org/10.26481/dis.20190529dm)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Valorization

What is the social and educational relevance of the research results?

The educational and social relevance of the research observations from this thesis relate to informing educational and regulation bodies of the recurring inadequacies of judgement-based assessment results; and therefore their intended use of measuring the performance and educational and training achievements of postgraduate medical trainees. The research observations and methods used not only inform the jurisdictional institutions, but also should motivate and drive change towards educational and policy reform that promote the quality of assessment practice, fit-for-purpose development and implementation, quality assessment programmes, and the development of assessment literacy for assessors. In addition the development of accreditation practices for assessment programmes to make them more relevant to trainees, supervisors, and society is particularly important. Why society? Society is reliant on health regulatory bodies to ensure the populace is protected from incompetence and poor medical practice as much as possible within reasonable intellectual, social and economic bounds.

Those with knowledge and experience, and those with regulatory powers have the responsibility to set those bounds, but buttressed within an educational scientific framework. The scientific framework is necessary to ensure that information used to formulate healthcare policies is as valid as can be achieved in order to free society from the tyranny of biased and ill-informed decisions based on possibilities and untested “good” ideas. The substance of such concepts is equally crucial for medical education practice and regulation. The disturbing thing though is that the medical profession has not attended to this latter problem first. That is, has not addressed the quality problem before the regulators need to step in to enforce the development and implementation of evidence-based methods to improve supervisor performance and assessment literacy.

The personal relevance to trainees is clear. Feedback for learning and clinical and professional formation, with improvement based on validly predicated instructional guidance is integral to quality training. Being given unreliable instructional information, and being assessed on phenomena about which the assessor has no, or poor understanding is an anathema to quality training. Similarly, the assessor may be in a position where they require training and experience in those same phenomena, but are not provided with the feedback about this need. Being asked to assess something about which one has little

understanding, or an incorrect understanding, borders on being a corruptive influence propagating less than mediocrity.

The dissertation provides both educational bodies and regulatory bodies with knowledge suitable for use, as pragmatically shown in this dissertation, for quality improvement processes to be implemented; with knowledge and methodology suitable for immediate implementation; and the translation of the information and methods into services, processes and new educational activities.

Who are the target groups for the research results?

For medical education in general, the groups for whom the research results would be of interest can be considered generic because they are identified simply by the use of judgement-based assessment methods in any context. The results can raise concern that similar issues may exist for any group involved in training using assessment and feedback by judgement-based observation. In addition learners, trainers and supervisors, institutions and governing bodies of the assessments and work-type, and the vocational professions are all stakeholders¹⁵.

Even though a specific group was studied, namely supervisors of postgraduate medical trainees, the problems demonstrated are ubiquitous. They extend well beyond the specific context of this thesis, are generic to judgement-based assessment, and are not specific to supervisors of medical trainees or medical education. The methods employed are also generalisable, especially those used to identify and measure the constructs assessed. Similar controversies identified in this dissertation have caused ongoing angst for results involving judgement-based assessment in most jurisdictions where they are used. Reframing the approach highlighted in this thesis offers remedies for some issues, especially clarifying competency-constructs used in judgement-based assessments.

Specific groups for which the research results are particularly pertinent can be readily identified. For example, in post-graduate medical education the target groups are the training programme organisers at a certifying level. At an implementation level, the corresponding example is a hospital-based training programme. Within programme implementation, an example is supervisors but also other assessors that may be fit-for-purpose for assessing specific performance, skills and/or knowledge. However the groups identified can be changed to any similar identity in other professions and work places, the problem is so wide spread. The group for which particular concern should exist are medical trainees whose education and training is partly, possibly largely dependent on judgement-based assessment and feedback, and especially the formative instruction based on that assessment. Unless they can be reassured that the information and feedback they

¹⁵ Those involved in any assessments that are used for feedback and formative instruction derived from judgement-based assessments used for the training of individuals to an acceptable standard of a required competency.

are being provided is valid and reliable, then misgivings can arise about their quality of training¹⁶.

Into what concrete products, services, processes or educational activities will the results be translated and shaped?

Certification and regulating bodies can use methods exemplified in this dissertation to assist the evaluation of the training programmes they accredit using objective data. Similarly the training programmes can immediately evaluate and obtain at least some measure of the quality of their supervisors' assessment literacy, and then develop quality improvement processes and measure their effectiveness. The supervisors as a group will also be able to apply the information from the perspective of professional self-regulation and development. Additionally, in an equitable and transparent education system, the adult learners will be able to obtain at least some quantitative measure of the quality of their assessments, and if not satisfactory be able to petition vigorously for time and cost resources for implementing improvement processes. Standard setting bodies are in a position to ensure adequate and timely progress faster than achievable by traditional peer reviewed publications and educational meetings. They have the ability to create change quickly and in a methodical transparent way by mandating active quality improvement and reporting for training programmes as part of accreditation requirements¹⁷.

A specific concrete product relates to the part universities can play in the implementation. Universities have the capability to attract the required individuals with appropriate training and experience, to create departments that provide assessment evaluation services to many separate training and stand-setting institutions thus *assuring equal standards*. Funding could be by above-cost recovery fee-for-service for the evaluative service. Also, traditional grant funding for innovative research aimed at improving assessments and human competence in this research domain¹⁸.

¹⁶ Fortunately for healthcare, medical schools and postgraduate training programmes are populated with highly intelligent and capable individuals who mostly are able to work around the problems. However, a better use of that intelligence is to provide valid, reliable and effective feedback and formative instruction to efficiently and effectively accelerate learning for most if not all of the time. This is particularly important given current worktime rules for medical training compared to when the current educational system was developed.

¹⁷ The caveat to the above potential for implementation of the services, processes and educational activities is that many of the concepts are relatively novel. A great deal of translational activity will be needed given the perceived complexity to many of the methods used. Translational activities will involve the use of traditional approaches, for example peer-review publications of research, learning and training workshops for the methods, and audit and feedback for all involved. Further explorative research is essential to effectively identify and therefore specifically focus on assessors who need improvement and/or calibration. In this dissertation only leniency/stringency bias was evaluated.

¹⁸ A university-based institution that provides independent and impartial evaluation and quality measurement services, and assessment-dedicated research, would be in a position to be

To what degree can the results be called innovative in respect to the existing range of products, services, processes and educational activities?

Providing detailed information about the implementation and methodology used to collect and then act on validity evidence is not usually a routine part of the evaluation of many, possibly most training programmes in post-graduate medical education and training. Yet this is recommended by professional medical educators. One of the many arguments made in the dissertation is that there should be such routine evaluation. Therefore it would be innovative for these and other strong methods to be employed routinely to ensure a minimal quality for assessments using feedback and formative instruction as well as for assessments used for making high-stakes decisions.

Another partial innovation is the use of continuous quality control methods and processes for routine evaluation of the quality of judgement-based assessments. These processes are able to identify assessors who have an understanding of assessee's competencies that is not appropriately calibrated. As a new conceptual approach they can be implemented for regular quality improvement using data that is already routinely collected for assessment purposes.

Another innovation is that assessment programmes can use the methods to ensure that the assessment for feedback and formative instruction is minimally adequate. It is the feedback and formative instruction within the workplace that is the important part of training and assessment for the trainee. Feedback and formative instruction are a major source of learning, and the content of that learning has to be valid and reliable. If the feedback and formative instruction is not valid and reliable, as measured by the supervisors' ability to reliably and validly identify and grade competence, then the supervisors need to be replaced or adequately trained to improve their assessment literacy. The background research review for this dissertation has highlighted a lack of appropriate and consistent methodological approaches to provide validity evidence for judgement-based assessments in a pragmatic and operational way¹⁹.

Rather than creating variable processes and structures for conducting evaluation, validation and innovative research within each training and standard-setting institution, or by using processes from scattered individuals at the work-face with little coordination, Universities are in an ideal position to establish large-scale centres dedicated to high-quality validation processes for service and research.

completely autonomous from the training programmes and standard setting and certifying intuitions involved. Such an approach also will allow unprejudiced comparison of quality between training programmes, which could then be used to drive improvement as required by the standard-setting bodies.

¹⁹ The concept of ensuring assessor quality for judgement-based assessments are fit for purpose and supervisors have assessment literacy appears to have been assigned to the too-hard-basket for decades.

Schedule and Implementation

How will these plans for valorization be shaped?

As stated, the observations and methods emerging from this dissertation could be of particular interest to certifying and standard setting bodies, the governing training institutions and the implementing training programmes. The plans include further explorative research into expanding the application of the concepts, advocating for replication by other researchers and institutions, and providing detailed information to the standard-setting bodies in the form of the published dissertation. The potential number of implementable and new research-based innovations is substantial and each would require appropriate planning for implementation with the development of a realistic schedule. The example chosen is the planning of newly defined University Departments to provide independent evaluation and validity evidence of the quality of judgement-based assessments and assessors, and in addition being a potential assessment-developer (see schematic below).

What is the schedule?

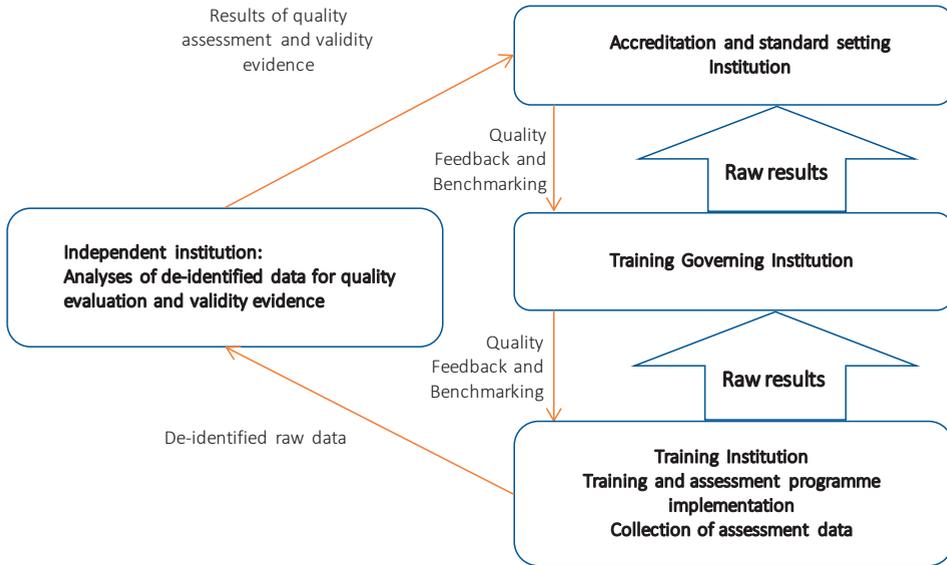
First, obtain cooperation from a credentialing body and a training programme with a well-established assessment package is the priority. Simultaneously a pilot study is being planned in association with the accrediting body and a related training institution to obtain real-time de-identified assessment data in order to produce the information and then determine the level of interest and identify pragmatic issues related to implementation²⁰.

Clarifying the necessary type and number of personnel needed for optimal effectiveness is also central to planning. In addition, identifying other needed resources and infrastructure for the establishment and then maintenance of such processes will be a major part of the pilot. Clearly ascertaining and categorising the potential barriers to implementation will also be an essential component of the pilot activity.

An iterative process will be created so that the schedule includes the development of consensus on the initial methodology and standards required. Not only is the establishment of agreed and definable consensus of standards an essential component to be reviewed after the pilot implementation, such an iterative process will need to be embedded as normal procedure for ongoing developmental improvement.

²⁰ Negotiations and planning for these are in progress.

Schematic representation of assessment data use:



Since this is a nascent concept, all the information required for the development of an academic institutional business case will need to be gathered so a subsequent substantive case can be developed. Providing the concepts are acceptable and the implementation sufficiently pragmatic, a number of follow-on activities can be instigated. Firstly, an agreement and elaboration on the sources of funding need planning. In addition, establishing the methods of dissemination of the concepts, and follow-on promotion, to all stakeholders will be an important process. Furthermore the creation of agreed expected outcomes and establishment of the efficiency capabilities that can be measured are an essential minimum requirement for future review and evaluation procedures²¹.

What are the risks involved?

An immediate risk is a lack of initial uptake. If early interest is however developed the follow-on risk is diminishing long term enthusiasm and participation. As for most academic-based initiatives, another risk is insufficient understanding and/or agreement among the stakeholders. This is a particular risk since personal ability and bias in perception of assessment constructs is part of the underlying problem, and why the current standard-setting practises are proposed. Adequate comprehension and assessment literacy of stakeholders is a potential major barrier for implementation.

²¹ Each of the over-riding components of the schedule will need micro-schedules. For example, obtaining cooperation from a credentialing body and a training programme with a well-established assessment programme needs specifying.

Although the market potential and stakeholders have been identified, the “market” may see itself as functioning well with the *status quo*. Given the length of time inaction has been prevalent for implementing quality processes, a general belief may be that tradition has it correct and no innovation is needed. Standard setting bodies may resist innovation and the market and stakeholders may need education through the traditional but slower means of research publication and ubiquity of evidence.

At present the extent of the likely competition is an unknown. Potential competitive organisations will exist, and the stakeholders may even take up the ideas and implement them in isolation²². So the likely market share needs to be identified, and a sensitivity analyses is necessary to start estimating likely uptake. The initial ideas and estimates about the understanding of the stakeholders are not known and their willingness for change may be wrongly estimated. A structured business plan needs to be developed and when done may indicate that the propositions are not viable.

What market opportunities are there?

Although at present the likely competition is not exactly known, there is a dearth of activity in the academic medical literature for assessment construct identification and measurement of assessment literacy. Quality control processes around assessment literacy are not a routine requirement for assessment programmes in post-graduate medical education. In addition there is no standardisation of those programmes with respect to the form quality control should be implemented. So this area is a potential open market place yet to be filled.

What are the costs involved?

A structured business plan needs to be developed to identify all costs. Initially these may be high given a need for well-trained personnel. Once the required people and systems have been identified more accurately the time/cost to profit ratio will improve. The professional infrastructure exists in Universities so the professionals’ available time and any additional training will be a cost.

Appropriate software and computer systems with suitable digital security will be necessary. This however does not need to be housed in a specific physical space. Given the importance of personal communication and collegial interaction for initial success and long term sustainability, some form of physical meeting space would be desirable to engage in the interactive scientific development. Information transfer and ongoing publicising costs will depend on initial uptake and demand.

²² Educationalists with their educational and training organisations can potentially use this dissertation to inform themselves on how to undertake the evaluation of the reliability and construct validity of their judgement-based assessments. The methods are generic and because of the detail in the explanation and also referencing, the methods sections allow easy replication. This is normal academic transparency.