

A PROPOSED MODEL OF COGNITIVE APPROACH TO ASSESS CLINICAL PERFORMANCE OF DENTAL STUDENTS

Ayman M. Khalifah

Department of Dental Education, College of Dentistry, Taibah University, Saudi Arabia

ABSTRACT

In dental education, assessment of students is of critical value to improve their performance in clinical settings. The assessment of students' clinical performance includes various stages, where in certain cases, assessors encounter challenges in providing final grades. This study sheds a light on assessment practices in clinical settings and focuses on assessors' modulation of the whole cognitive process. The argument involves discussing critical thinking of assessors before, during, and after the event of assessment. Then, it analyzes a cognitive approach of assessment implied by assessors during students' performance. Further, it proposes a model with step-by-step approach in decision-making along with different factors, which may strongly influence final grades. Four main stages were identified for the purpose of analysis, such as pre-decision, driver, primary decision, and communication stages. Each stage was supported by literary data, along with evidences worth consideration. Possible factors related to the assessment and assessors' cognition that derived from literature were discussed in terms of the influence on the final decision towards more stringent or lenient decisions, following a temporal sequence for the proposed model events. Finally, both primary and secondary factors involved in each stage were presented.

KEY WORDS: decision-making, cognitive approach, curriculum, clinical performance, clinical assessment.

J Stoma 2021; 74, 1: 57-64

DOI: <https://doi.org/10.5114/jos.2021.104700>

INTRODUCTION

There has been a significant change in the meaning of process of assessment in medical education. In traditional terms, the assessment process refers to an instrument, which helps educators to understand students' learning, and to prepare and provide them with the required knowledge for satisfactory decision-making process in their practical life [1]. During education, students usually focus on assessment-oriented education based on completing necessary requirements, which serve as a tool to pass a certain course. In non-traditional terms, the assessment process deals with optimizing learning potential of students, and ensures development of re-

quired abilities, including self-directed learning, critical thinking, life-long learning, creativity, and innovation in education [2].

The process of assessment has many objectives, which include students' assessment on how they achieve the learning outcomes. Moreover, decision-making is another important part of the assessment process. However, the process of decision-making is mostly based on the main cognitive approach and factors that contribute to the development of meaningful assessment grades, such as 'pass' or 'fail'. The cognitive approach affecting the assessment of students has been discussed in the literature, which covers the development of major skills, including cognition, interpretation, application, synthesis of decision-making process, and the ability to

**JOURNAL OF
STOMATOLOGY**
CZASOPISMO STOMATOLOGICZNE

OFFICIAL JOURNAL OF THE POLISH DENTAL ASSOCIATION | ORGAN POLSKIEGO TOWARZYSTWA STOMATOLOGICZNEGO



ADDRESS FOR CORRESPONDENCE: Dr. Ayman M. Khalifah, Department of Dental Education, College of Dentistry, Taibah University, Prince Abdulmajeed Ibn Abdulaziz, Bani Muawiyah, Al-Madinah, Al-Munawwarah 42313, KSA, e-mail: akhalifah@taibahu.edu.sa

RECEIVED: 07.10.2020 • ACCEPTED: 27.12.2020 • PUBLISHED: 30.03.2021

judge students accordingly [3]. The cognitive structure of decisions has been recognized as a new direction, which has recently been embraced in medical education to study the cognition in practices of assessments [4, 5]. This study summarizes and recommends new methods of cognitive thinking to understand the relationship between cognition and the process of decision-making during clinical assessments. Furthermore, it can assist in discovering potential influential factors that affect cognition during the process of assessment in a more systematic way, which can decrease the complications of cognitive structure. Factors contributing to the proposed model interact with a decision at particular stages of the process. Understanding their effect and impact on decision-making can direct a research in medical education to improve accuracy, reliability, and utility of clinical assessments [6]. The aim of the study was to propose a new model following the review of available literature to clarify the cognition of decisions in assessing clinical students and trainees.

LITERATURE SEARCH AND AVAILABLE MODELS

A literature search was conducted using MEDLINE, Embase, and Google Scholar databases. Different search terms were applied, including an assessment of clinical performance, cognitive approach, and decision-making in medical education. Based on the literature review, one model has been reported to discuss the cognitive approach of clinical assessment [7]. Briefly, the model uses internal and external information as primary factors to affect cognitive reactions in the process of decision-making. The focus was to discover an expert assessors' cognition based on the theory of expertise, which emphasized the role of assessment experience in shaping the cognition of future tasks [8]. The study sought to discover how the experts balance between external and internal factors. The external factors have been defined as non-personal sources of information, including assessment rubrics, structure, program out-

comes, and institutional and national expectations. Internal sources according to the model were those related to assessor's experience, knowledge, instinct feeling, and expectations. They also referred to assessors' own clinical practices and reasoning as factors, which facilitate the process of internalization of external criteria to be compared with a trainee performance. The comparison also included assessors' knowledge based on their level of training. Without considering the time frame of events, the previous model described the cognitive process in a condensed short manner. For example, internal and external factors were considered as a bulk of cognition, despite the presence of other factors. Expertise was the main focus of the previous model to shape the cognitive structure. In a new model, these factors are mainly related to the period of experience gained during work-life of assessors. Moreover, a temporal sequence considers expertise as a primary, but not the only factor to affect the cognition during any specified task assessment. The factors in a new model are classified into primary and secondary factors to facilitate tracing and application. It was arguable, whether assessors in a new model should be clinically qualified practitioners who usually are not academics since the previous model did not mention this difference of experience and its influence on the cognition of assessment.

THE NEW PROPOSED MODEL

The proposed model aims to clarify the relation between all contributing factors that relate to guiding assessors in decision-making during the process of assessment through different stages of the cognitive process. For a better understanding of assessors' cognition and to confirm the overall assessment and grading of students' performance, the model was divided into four successive stages, such as pre-decision, driver, primary decision, and moderation stage (Figure 1). More explicit and specific definitions in the current model to each nominated stage of cognition with the predicted influences

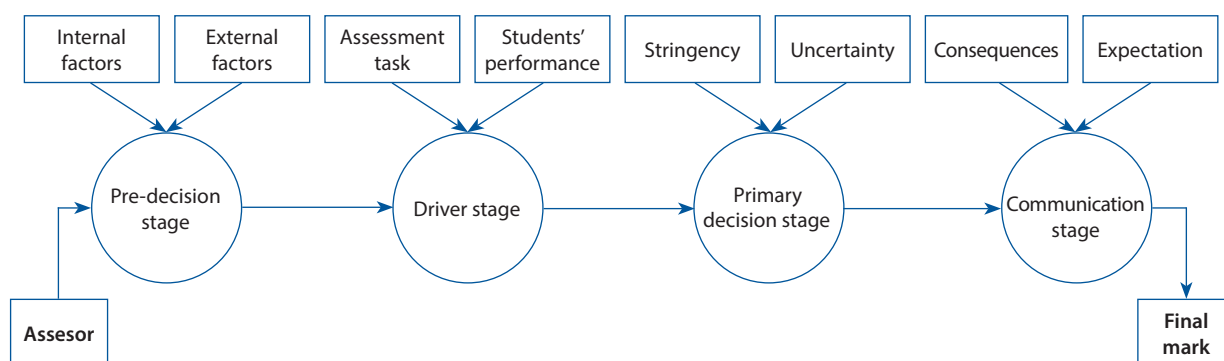


FIGURE 1. Decision making stages in students' assessment. Stages and related factors that have been proposed by the author based on the available literature

TABLE 1. Primary and secondary factors involved in decision-making process

Stage/factor	Primary	Secondary
Pre-decision	Internal (expertise, contextual knowledge)	Internal (personality of the assessor)
	External (continuous professional development, curriculum context)	External (political and administrative situations, expectation, increased number of candidates during the exam)
Driver	Student performance (quality or competence)	Student performance (impression formation)
	Task (assessment criteria)	Task (employing the frame of reference)
Primary decision	Similar to preliminary stage, assessor's self-confidence	Similar to preliminary stage; cultural attributes
Communication	Legal consequences, community, and patient's safety	Cultural consequences, institutional expectations

were used. In fact, the focus was to discover the impact of influencing factors on each different stage, aiming at producing decisions that could be as accurate as possible, reflecting the real clinical assessment. These influencing factors were classified into primary and secondary factors. Primary factors consisted of those, which influence the decision and update with a new decision in a reciprocal relationship. Secondary factors may influence the decision, but could not be affected by the decision in return (Table 1).

Briefly, the four stages explain the flow of information during the cognitive process with conjoint factors. For example, internal and external sources are the main factors to affect a pre-decision (non-task-specific) cognitive stage in clinical settings. The driver stage starts when the assessor is present to judge the performance of specific clinical tasks of assessment. The third stage is the primary decision stage, which begins when the assessor finds or sees (interpret) students' performance according to the defined frame of reference. The resulted primary decision predicts a range of options between 'being sure' and 'uncertainty'. Then, the refinement process of decision grade, the fourth stage, is the moderation of the decision, which is affected by another set of factors, such as legal consequences, community, and patient safety to direct the decision towards grading. In the following sections, the most influencing factors, which have been discussed in earlier studies, are incorporated within each stage explaining the full proposed model.

PRE-DECISION STAGE

The pre-decision stage involves the pre-assessment attitude of the assessors that expresses their characteristics with certain internal and external factors. The pre-assessment attitude plays a fundamental role in shaping the judgement during assessment. Therefore, these factors have indirect relationship to the existing assessment task or students' performance. In other words, these internal and external factors contribute to building up the cognitive decision about students' performance within the assessor's mind.

INTERNAL FACTORS

Internal factors are based on personal and professional characteristics that match and mix, to help predicting the assessors' decisions. For example, gender, expertise, and content knowledge are part of internal information at this stage [4]. However, some studies have indicated that students' assessment preferences are influenced and depending on several factors [9, 10]. Not surprisingly, researchers paid a little attention to the impact of assessor's gender on the cognition of clinical assessment [6]. With reported poor inter-rater decision agreements, female assessors tend to be less rigorous in their judgments [6,9].

Expertise is believed to be a primary internal factor, which influences clinical reasoning to update assessors' judgment capacity. Clinical reasoning consists of two types: content-dependent and context-dependent. However, in therapeutic and diagnostic reasoning, expertise generally differs among assessors [11, 12]. For example, in psychology, deliberate practice acts as a key to expert performance. It may further benefit clinical reasoning as that experts possess more stringent decisions than early career assessors [11, 13]. In the light of behavioral learning perspective theory, expertise is guided by disciplines' specific knowledge and skills [14], which in turn affect internal assessors' attitudes, emotions, intentions, and personalities [4, 6]. These also reflect 'gut feeling' of expert assessors to defend uncertainty against students' performance and reaching consistency of decisions [7]. The 'gut feeling' evokes the sense of alarm. Therefore, expertise may result in an increased assessors' stringency towards trainee's performance.

Content knowledge is another internal factor, especially the concerns on the accuracy of rating and comparison among students [15]. For example, an increased level of content knowledge of assessors affects their rating. Moreover, those assessors who have a direct interaction with students tend to give higher marks, which could be intentional, when compared to those assessors who present only during the assessment task without having previous interaction with students [16]. It has been reported and identified that content knowledge is one of the major factors behind a good assessment.

Many respondents in that study have indicated that an unfair assessment arises due to the lack of content-related knowledge [15]. However, it is not easy to determine, which internal factor overrides the other, but assessor's familiarity with students' overall progress make their judgments more consistent.

EXTERNAL FACTORS

External resources represent the other aspect of the pre-decision stage. They are flexible and changeable according to learning environment. In other words, they underline the indirect influence of political and administrative situations, which consequently reflect accountability and trust towards the host (institutions) environment [17]. External factors include, but are not limited to, curriculum context, direct institutional expectations, and continuous professional development [6, 18]. In fact, limited information exists on the relative impact of curriculum context and institutional/international expectations on decision-making, especially for evaluating clinical competences.

In clinical learning and assessment, curriculum context is usually dynamic and demands flexibility of assessors. More understanding of the learning context results in clearer and more defensible decisions by assessors [19]. As noted above, teaching experience within the clinical context increase the consistency of rating and may increase the passing rates [1]. However, this is very much related to the impact of the content to be assessed. For instance, if the assessment is for interpersonal skills, such as professionalism, assessors tend to give higher marks [9]. In contrast, if the assessment is to evaluate clinical skills, such as history taking or physical examination, assessors tend to be more stringent in their decisions [9]. A recent study has analyzed the experiences of tertiary and industry-based experiences of grading nursing students in clinical courses, specifically in situations when the students' performance was neither a clear pass or fail. Findings of the study indicated that most assessors took advantage when students' performance remained doubtful. They further reported that most of assessors preferred failing students, based on their academic performances [20].

Institutional expectations, on the other hand, present certified criteria and standards, which reflect the complexity and values of the organization to provide guidance and roles of expectations to the raters. St-Onge *et al.* [7] supported this argument and stated that following the external criteria of assessment, which include assessment grids, accredited institution's expectations, licensing contributing to framing their observations, and assessments related to students' clinical performance. Assessors need to consider these roles during evaluation to avoid unnecessary conflicts, especially when academic freedom is of concern [21]. Therefore, assessors' deci-

sions, when considering institutional expectations, tend to be less stringent.

Training and faculty development programs play an important role as they enhance conceptions of shared responsibilities, expectations, and accountability between educators and hosting cultures. Consequently, this improves the consistency of judgments [4, 22]. In certain cases, assessors tend to have different expectations in relation to the assessment of standards for students [23]. These expectations are related to clinical knowledge, attitude, and technical ability of students. Since assessors generally have their own set of values, their decisions are generally influenced by standards. Assessors can predict boundaries of performance by developing an agreed-upon reference framework, improved understanding of assessment criteria, and aligning them with individual beliefs of standards [3]. Nevertheless, assessors who lack training, especially if they do not have teaching roles, would rely on their own clinical experiences to judge trainee's performance. If, for example, the performer used disorganized history taking or abrupt physical examination, no doubt the assessor would give lower marks. Therefore, training of assessors on assessment strategies improve decision-making certainty and decrease stringency.

Interestingly, one external factor that is found to increase the stringency of assessors' decision is the increased number of candidates during time of assessment [9]. This could be related to the feeling of fatigue or other factors, which may be further investigated in future studies. To conclude, it is now clear that the above-mentioned internal and external factors can influence the cognitive process during the pre-decision stage to initially predict assessors' appraisals of learners' performances. A combination of assessment task and students' performance represent the next stage of this model, as they drive the assessor cognition to prepare for comparisons and interpretations of what was observed during the driver stage of cognition.

DRIVER STAGE

This stage is characterized by the presence of a specific assessment task followed by student performance that drives the cognition of assessor towards reaching a preliminary decision. In fact, when considering the above internal and external factors, it is vital to discuss the influence of "impression formation", a subconscious stereotyping [24]. Once the trainee presents to the assessment task, the categorization begins, then the decision may tend to be less stringent or affected. However, this is a secondary factor.

Another factor that needs to be considered during assessment is how critical the assessment task is. In other words, the assessment is affected by the reaction of assessor towards critical performance during formal

assessments [25]. They pay more attention to students' precision of interactions when making inferences and when compared with informal assessments. This enables assessors to make more stringent decisions since the primary purposes of summative assessment is grading, certification, and accountability [26]. However, once the task is specified at the time of performance, assessor's mind evokes the frame of references to be used when observing the performance [6]. These frames of references include self-expertise, expertise of other doctors, students' performances, patients' outcomes, or assessment criteria in the form of rubrics.

Occasionally, assessors present what is called "bias due to a recent experience" when judging a particular student's performance. Therefore, their rating is influenced by a recent rating experience of previous students [27]. This might occur due to underestimation or overestimation of how well some students can perform, but unexpectedly they show professional competence. Thus, this results in unintentional bias and, therefore, being more stringent in their decisions. Likewise, if the assessor considers a qualified performer like him/ herself or a colleague as a reference for rating, then decisions would be based on high stringency. The reason might be attributed to their expectations that are higher than the actual performance of students.

Furthermore, if patients' outcomes dictate assessor considerations during the assessment, then stringency also masters the decision. Also, in the presence of assessment criteria, the consistency of grading increases, which indicates an increased certainty of decisions [28]. While with rubrics, precision of performance increases and can be traced by evaluator, because of the implementation of reasoning that promotes fairness and accuracy. In fact, choosing the reference frame may be dependent on criticality or difficulty of the task as seen by the evaluator.

However, in highly complex tasks, assessors' ability to identify the quality of students' performance in such complex tasks may be decreased, leading to underestimation of the overall grade [7-29]. Following this, the expected decisions are found to be less stringent. Subsequently, when actual performance begins, at the same time, making inferences starts with encoding process [30]. The encoding process has been defined to occur between the initial observation and the storage of that specific performance inside the memory of the assessor. Denisi and Peters [31] stated that a rater's ability to accurately recall information is dependent on how the information was organized in rater's memory during the encoding process. On the other hand, it has been argued that the context of assessment affects the degree of stringency during the encoding process [32].

In addition, during the process of making inferences, a task-specific reasoning occurs. There are three types of reasoning known for educational assessments, including deductive, inductive, and abductive reasoning [33].

In brief, Mislevy [34] defines deductive approach as what accompanies up-down reasoning, (i.e., assessors who start reasoning from disease down to symptoms, and prefer using standards or criteria-based assessments). The same previous study has shown that with using rubrics, the average scores of students' increase [34], which indicate a decrease in stringency of decisions. While the bottom-up approach mainly uses inductive and sometimes abductive approaches, they are those who advocate for non-standard assessment systems (i.e., their reasoning starts from symptoms up to disease). Their rating is more dependent on expertise, which is known as source for increased stringency [9].

Assessment type is another considerable factor, which affects the direction of decisions. For performance assessment, the usual method is using a checklist or rating scale forms [35]. Using rating scales requires more descriptive judgment from the assessors, which involves providing students with assessment's feedback for their earned grades [36]. For example, when using mini clinical evaluation exercise format (mini-CEX) for oral examination, assessors give high marks for humanism when compared with other competencies [37]. However, this may be attributed to the learning context rather than to the assessment type. A recent study has indicated that the complex data, like competency-based portfolios, must be assessed following a different approach for critical judgements and interpretation. Findings suggested that such a complex data must be assessed utilizing different and multiple approaches for more critical assessments [38].

Considering how the task is chosen in terms of the assessment context, the type of assessment is important for the assessor to make a meaningful interpretation of the performance. Inability to make stringent and more certain decisions occur if the above factors in the assessment task construction are not compatible with assessor's preferences. This may explain the resistance of some instructors to use rubrics in their work [28].

PRIMARY DECISION STAGE

After completion of the assessment task, depending on the previous cognitive stages, assessors reach a decision that ranges from 'being certain' to 'uncertainty'. In fact, uncertainty accompanies every decision-making process, unless some details are visible to direct the decision towards assertion [39]. These details represent areas of performance and how they match with the criteria of competence within the assessor's mind. On the other hand, assessors' characteristics play a central role in directing their decisions during this stage. Expert assessors, for instance, have the capacity to use alternative reasoning in different situations. This allows them to systematically reach better and more consistent judgments, which are more stringent than provided by in-

experienced assessors [40]. Experts pay more attention to situation-specific cues, particularly in complex tasks, such as in the clinical performance assessment. During observation, they request additional information, extra performance, like repeating certain tasks or verbal explanation, from students to confirm their ratings.

Regardless of expertise, if the performance of task was persuasive and matches the assessors' beliefs of standards, then uncertainty of decision decreases. However, some assessors have variability of their ratings despite the clarity of performance, indicating an increase in uncertainty of decisions that requires further investigation. Uncertainty increases due to the influence of different variables. For instance, lack of content knowledge and familiarity with presenting student [41]. Uncertainty, therefore, could lead assessors to be less stringent and to provide overestimated grades.

Another factor that may affect certainty of decision-making in educational settings is cultural attributes [42]. Assessors who have individualistic backgrounds usually tend to focus on the task, while those from collectivist cultures focus more on contexts [43]. These features are more prominent in communication styles of either assessors or trainees. In other words, the way the assessor acquires or interacts with information is dependent on cultural features. For example, assessors from individualistic background expect verbal and non-verbal communication to be direct, explicit, and acquire open manners. Those from collectivist cultures tend to acquire indirect, implicit, and more contextual manner of communication styles [42, 43]. Therefore, the decision process is longer for people from collectivist culture, as they need to consider not only the task, but also the consequences of decision on surrounding people. Collectivist individuals also value social obligations and harmony of relationships. These characteristics require the person to consider more careful approach during decision making [44]. As a result, assessors (collectivist) tend to be less stringent and possibly more uncertain of decisions. Studies on the influence of cultural features on the cognitive process are rare. It may be essential to find out how strong cultures are in causing variation of decisions or uncertainty during this stage.

In conclusion, uncertainty can relate to different factors of assessor's personality, while other are associated with the environment that has external or contextual nature. These factors reflect the degree of assessors' self-confidence in their own skills, their confidence in the assessment instrument used, the degree of risk or difficulty of the task, and assessors' knowledge [44].

COMMUNICATION STAGE

It is a refinement stage, which represents self-communication period of decision that needs confirmation. If the proposed grades are far from 'pass' or 'fail' border-

lines, then considerations are not required, and the final grade is asserted. If not, then assessors revisit their cognitive processing to find reasons to fail or pass the student. During this stage, two main factors are to be considered that may affect the communication of decision: the consequences of decision and the expectations associated with final grades. Both reflect the criticality of uncertain decisions reached during the previous stage. The possible consequences that the assessor usually considers at this stage include the impact of assessment decision on student future learning, where cultural, legal, and human safety consequences are considered [45, 46].

Cultural consequences have been discussed as a factor to influence stringency of decisions despite individual differences of students' clinical abilities [47]. However, considering the effectiveness of social cognition, it can be assumed that the interaction between people and social environment unavoidably impacts individual's decision-making [41]. Moreover, cultural bias is observed as one major reason associated with unexpected differences in students' assessment results [47]. This reflects the degree of matching of cultural attributes of both assessor's and students, which results in a decreased stringency of decisions. Another factor is legal consequences and appeals [45]. In grading performances and particularly borderline cases, assessors consider this factor if uncertainty dominates their judgments. Assessors, therefore, tend to avoid these consequences by giving higher marks.

The third consequence is related to community and patients' safety [46]. If the student's performance implies any potential harm (based on what the assessor perceives) to the patient, the assessor becomes more certain and stringent in decisions. Finally, grading policies play an important role to affect expectations of educational institutions in general, beside expectations of faculty members and students [48]. Institutional expectations sometimes require assessors to pass some students who fulfill certain requirements, although their performances were assumed to be under the assessor's expectations, resulting in decreasing the level of assessor's stringency in decision [21].

CONCLUSIONS

This model for cognitive process explains different levels and stages of cognition, which may occur in the assessment of students. There are many factors that contribute to the final and overall grading. Those factors have been incorporated in different stages of the proposed model, which could help to improve the accuracy of the assessment system. In the light of the above arguments, this paper provides a valuable contribution in the existing literature as it includes sufficient information, which would assist assessors in critical analyzing and distinguishing considerable factors. The study is further effective in improving the assessment processes implemented in different clinical institutions.

ACKNOWLEDGMENT

We would like to express our appreciation to Assistant Professor, Dr. Nahar Ghouth for his comments and linguistic edits on an earlier versions of the manuscript. Moreover, any errors are the author's own and should not destroy the reputations of such esteemed professional.

CONFLICT OF INTEREST

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Carneiro V, Pequeno A, Machado M, Aguiar D, Carneiro C, Carneiro R. Assessment process in dental schools: perspectives and teaching challenges. *RGO – Rev Gaucha Odontol* 2020; 68. DOI: 10.1590/1981-863720200002220180004.
- El-Kishawi M, Khalaf K, Al-Najjar D, Seraj Z, Al Kawas S. Rethinking assessment concepts in dental education. *Int J Dent* 2020; 2020: 8672303-8672303. DOI: 10.1155/2020/8672303.
- Khan N, Saeed M, Bari A, Butt A. Dental students perceptions about assessment methods. *J Pak Dent Assoc* 2018; 27: 202-206.
- Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. Seeing the 'black box' differently: assessor cognition from three research perspectives. *Med Educ* 2014; 48: 1055-1068.
- Gingerich A, Ramlo S, van der Vleuten C, Eva K, Regehr G. Inter-rater variability as mutual disagreement: identifying raters' divergent points of view. *Adv Health Sci Educ* 2017; 22: 819-838.
- Kogan J, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: a conceptual model. *Med Educ* 2011; 45: 1048-1060.
- St-Onge C, Chamberland M, Lévesque A, Varpio L. Expectations, observations, and the cognitive processes that bind them: expert assessment of examinee performance. *Adv Health Sci Educ* 2016; 21: 627-642.
- Ross K, Shafer J, Klein G. Professional judgments and "naturalistic decision making". In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR (eds.). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge: Cambridge University Press; 2006, pp. 403-419.
- McManus I, Thompson M, Mollon J. Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP (UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Med Educ* 2006; 6: 1-22.
- Alenezi H. Evaluating dental students' preferences of the current assessment methods used in dental education and their impact on learning approaches. ProQuest Dissertations Publishing; 2018.
- Ten Cate O, Durning S. Understanding clinical reasoning from multiple perspectives: a conceptual and theoretical overview. *Principles and Practice of Case-based Clinical Reasoning Education* 2018. pp. 35-46.
- Asch D, Nicholson S, Srinivas S, Herrin J, Epstein A. How do you deliver a good obstetrician? Outcome-based evaluation of medical education. *Acad Med* 2014; 89: 24-26.
- Mamede S, Van Gog T, Sampaio A, De Faria R, Maria J, Schmidt H. How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Acad Med* 2014; 89: 121-127.
- Wraga W. Understanding the Tyler rationale: basic principles of curriculum and instruction in historical context. *Espacio, Tiempo y Educación* 2017; 4: 227-252.
- Berendonk C, Stalmeijer R, Schuwirth L. Expertise in performance assessment: assessors' perspectives. *Adv Health Sci Educ* 2013; 18: 559-571.
- East L, Peters K, Halcomb E, Raymond D, Salamonson Y. Evaluating objective structured clinical assessment (OSCA) in undergraduate nursing. *Nurse Educ Pract* 2014; 14: 461-467.
- Govaerts M, Van der Vleuten C, Schuwirth L, Muijtjens A. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Educ* 2007; 12: 239-260.
- Burrack F, Urban C. Strengthening foundations for assessment initiatives through professional development. *Assessment Update* 2014; 26: 5-12.
- Williams R, Klamen D, McGaghie W. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med* 2003; 15: 270-292.
- Hughes L, Mitchell M, Johnston A. Just how bad does it have to be? Industry and academic assessors' experiences of failing to fail – a descriptive study. *Nurse Educ Today* 2019; 76: 206-215.
- Sadler D. Academic freedom, achievement standards and professional identity. *Qual High Educ* 2011; 17: 85-100.
- Eva K, Bordage G, Campbell C, et al. Towards a program of assessment for health professionals: from training into practice. *Adv Health Sci Educ* 2016; 21: 897-913.
- Poole C, Boland J. What influences assessors' internalised standards? *Radiography* 2016; 22: e99-105.
- Gingerich A, Regehr G, Eva K. Rater-based assessments as social judgments: rethinking the etiology of rater errors. *Acad Med* 2011; 86: S1-7.
- Cornell D, Krosnick J, Chang L. Student reactions to being wrongly informed of failing a high-stakes test: the case of the Minnesota Basic Standards Test. *Educ Policy* 2006; 20: 718-751.
- Newton P. Clarifying the purposes of educational assessment. *Assess Educ* 2007; 14: 149-170.
- Yeates P, O'Neill P, Mann K, Eva KW. 'You're certainly relatively competent': assessor bias due to recent experiences. *Med Educ* 2013; 47: 910-922.
- Reddy Y, Andrade H. A review of rubric use in higher education. *Assess Eval High Educ* 2010; 35: 435-448.
- Tavares W, Eva K. Impact of rating demands on rater-based assessments of clinical competence. *Educ Prim Care* 2014; 25: 308-318.
- Goldberg E. Effects of prior expectations on performance appraisal: a social-cognitive approach. Doctoral dissertation, University at Albany, State University of New York, Department of Psychology; 1993.
- DeNisi A, Peters L. Organization of information in memory and the performance appraisal process: evidence from the field. *J Appl Psychol* 1996; 81: 717.
- Hauenstein N. An information-processing approach to leniency in performance judgments. *J Appl Psychol* 1992; 77: 485.
- Mislevy R. Evidence and inference in educational assessment. *Psychometrika* 1994; 59: 439-483.
- Reddy Y, Andrade H. A review of rubric use in higher education. *Assess Eval High Edu* 2010; 35: 435-448.
- Swanson D, van der Vleuten C. Assessment of clinical skills with standardized patients: state of the art revisited. *Teach Learn Med* 2013; 25 (Supp 1): S17-25.
- Berk R, Theall M. Thirteen strategies to measure college teaching: a consumer's guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians. Stylus Publishing; 2011, pp. 47-63.
- Norcini J, Blank L, Arnold G, Kimball H. Examiner differences in the mini-CEX. *Adv Health Sci Educ Theory Pract* 1997; 2: 27-33.
- Pool A, Govaerts M, Jaarsma D, Driessen E. From aggregation to interpretation: how assessors judge complex data in a competency-based portfolio. *Adv Health Sci Educ Theory Pract* 2018; 23: 275-287.
- Fischhoff B, Davis A. Communicating scientific uncertainty. *Proc Natl Acad Sci* 2014; 111 (Suppl 4): 13664-13671.

40. Hyde C, Yardley S, Lefroy J, Gay S, McKinley R. Clinical assessors' working conceptualisations of undergraduate consultation skills: a framework analysis of how assessors make expert judgements in practice. *Adv Health Sci Educ Theory Pract* 2020; 25: 845-875.
41. Berendonk C, Stalmeijer R, Schuwirth L. Expertise in performance assessment: assessors' perspectives. *Adv Health Sci Educ Theory Pract* 2013; 18: 559-571.
42. Wilby K, Govaerts M, Austin Z, Dolmans D. Exploring the influence of cultural orientations on assessment of communication behaviours during patient-practitioner interactions. *BMC Med Educ* 2017; 17: 61.
43. Oguri M, Gudykunst W. The influence of self construals and communication styles on sojourners' psychological and sociocultural adjustment. *Int J Intercult Relat* 2002; 26: 577-593.
44. Brew F, Hesketh B, Taylor A. Individualist-collectivist differences in adolescent decision making and decision styles with Chinese and Anglos. *Int J Intercult Relat* 2001; 25: 1-19.
45. Gynnild V. Student appeals of grades: a comparative study of university policies and practices. *Assess Educ Princ Pol Pract* 2011; 18: 41-57.
46. Dudek N, Marks M, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med* 2005; 80: S84-87.
47. Kruse A. Cultural bias in testing: a review of literature and implications for music education. *Update Appl Res Music Educ* 2016; 35: 23-31.
48. Voge D, Higbee J. A "grade A" controversy: a dialogue on grading policies and related issues in higher education. *Research and Teaching in Developmental Education* 2004; 21: 63-77.