

STATE OF SCIENCE
SITUATION

Inferring signs from purposeful samples: The role of context in competency assessment

Marise Ph. Born^{1,2} | Karen M. Stegers-Jager³  | Chantal E. E. van Andel³ 

¹Department of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands

²Optentia and Faculty of Economic and Management Sciences, North-West University, Vanderbijlpark, South Africa

³Institute of Medical Education Research Rotterdam, Erasmus MC, Rotterdam, The Netherlands

Correspondence

Karen M. Stegers-Jager, PhD, Erasmus MC, IMERR, Room AE-241, PO Box 2040, 3000 CA Rotterdam, The Netherlands.
Email: k.stegers-jager@erasmusmc.nl

Abstract

Context: Medical students' clinical competencies are customarily assessed using convenience samples of performance from real practice. The question is how these convenience samples can be turned into purposeful samples to extrapolate students' overall competency profile at the end of medical school, particularly given the context specificity of clinical performance. In this paper, we will address this issue of inferring signs from samples using insights from the discipline of psychology.

Theoretical perspective: We adapted Smith's theory of predictor validity of universals, occupationals and relationals to the context of clinical competency assessment. *Universals* are characteristics required by all working individuals and therefore not context dependent. *Occupationals* refer to characteristics required by certain jobs but not others and therefore are dependent on task-related features of an occupation. *Relationals* are required in a specific organisational context with habitual ways of working together.

Application: Through seven propositions, we assert that generalising from samples of assessed clinical competencies during clerkships to generic competencies (i.e., signs) is dependent on whether characteristics are universals, occupationals and relationals, with universals most and relationals least generalisable.

Conclusion: When determining what types of ratings to use to evaluate medical student competence, medical education has shown too little nuance in considering the degree to which particular characteristics are likely to be generalisable, approaching the issue in an all-or-none manner. Smith's distinction between universals, occupationals and relationals offers a promising way forward that has implications for assessment, student selection and career choice.

1 | INTRODUCTION

During clinical training, medical students' fitness for practice is customarily assessed using convenience samples of performance in

realistic learning and working environments. In these so-called workplace-based assessments, raters use direct observation to appraise students' knowledge, skills and attitudes in a particular situation, for example, while interacting with a patient. However, medical schools are not so much interested in how a student performs in a specific observed situation but rather in the extent to which they can

Marise Ph. Born and Karen M. Stegers-Jager contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Medical Education* published by Association for the Study of Medical Education and John Wiley & Sons Ltd.

use the observation to generalise about the students' ability to perform in other situations.¹ Put differently, they are interested in providing an estimation of students' general standing on the competencies important for entering the labour market as medical doctor such as those defined in the CanMEDs framework.² Among educational researchers, the issue therefore has been raised of whether the current convenience samples of workplace-based assessments are appropriate for extrapolating a student's overall medical competency profile at the end of medical school, given that clinical assessments are often context bound and specific. One may, for instance, wonder whether it is possible to draw general conclusions about students' interpersonal skills from the observation of one or two interpersonal situations. Next to context specificity, there is the issue of multidimensionality: most student-patient interactions require students to integrate and perform several competencies.¹ Although workplace-based assessments are attractive because they provide samples that enable assessing multiple competencies in an integrated manner,³ these make it harder to distinguish mastery of individual competencies. This paper's goal is to examine how we might begin to consider students' general competency profiles in more nuanced ways by using a taxonomy of characteristics that can lead to purposeful rather than convenience sampling of workplace-based assessments. By implication, we also explore the issue of which elements of physician performance are premature to assess in medical students and at what moment in a doctor's career it will become appropriate to measure them. Borrowing from the discipline of Psychology, we frame our goal as the need to *infer signs (that one is ready for practice) from (purposeful) samples*.

The use of samples of performance from real practice to assess professional competence can be regarded as an example of the *samples approach*, which refers to a situation where a sample of representative performance or behaviour is used as a predictor of future performance or behaviour.⁴ The samples approach is usually contrasted with the signs approach. In the signs approach, as commonly applied in (personnel) selection contexts, distinguishable constructs, traits, or skills, such as cognitive abilities and personality traits (= signs) are used to predict performance. Although sound assessments do not advocate for the use of personality traits to predict performance, even grades can be considered multifaceted compound measures influenced by cognitive ability, personality traits and study skills.^{5,6} Similarly, the general conditions identified for entrusting individuals with a professional activity are ability, integrity, reliability and humility,⁷ further suggesting that measures taken during medical school are likely to, indirectly, include personality traits.

An important difference between both approaches is that, in the samples approach, constructs are not measured separately, but in an integrated way by means of demonstrating a skills repertoire to perform a task. The strength of this approach lies in the notion of behavioural consistency: the predictive validity of the behavioural measures will be higher the more this behaviour and the to be predicted future behaviour are alike.⁵ However, this is exactly what complicates matters when the aim is to predict medical students' general competency level. Samples predict performance in specific situations, so how do we then generalise to performance in other situations that were not

measured? Another disadvantage is that due to the multifaceted nature of the measures, it is often not clear why a student underperforms in a particular situation. This reduces the value of samples as diagnostic instruments. So, the question is how can we ensure that our samples enable us to infer signs?

We will argue that it is important to start with defining the KSAOs (knowledge, skills, abilities and other characteristics including values⁸) that are required for the job of MD,² resulting from work analysis. The good news is that—to some extent—such work analyses have already been conducted for the medical field and have led to the definition of competency frameworks underlying the current medical curricula, including CanMEDS and ACGME competencies.^{2,9} From here, the question becomes how do we ensure that the samples we take during clerkship provide sufficient information to infer signs that will be predictive across various contexts. To address this issue, we use Smith's framework¹⁰ to classify individual characteristics into the so-called categories of universals, occupationals and relationals to propose that the effects of contextual factors when trying to extrapolate from samples to signs are dependent upon the category of individual characteristics.

In doing so, this paper offers a critical review of the current domain of workplace-based assessment in medical school and a framework for purposeful sampling for assessment by exploring how different aspects of competence might usefully be thought of as having different generalisability. Smith's framework was developed in the context of assessment and selection of personnel and offers an overall view to explain relationships between the content of assessment measures and their relationships with future workplace performance. It is essentially a validity model, focusing on whether assessments can predict useful outcomes (cf. Borman et al.¹¹). Our choice for using Smith's work is its relevance for the type of inferences that can be made from assessments of medical students. Other validity models, such as Messick's,¹² were viewed as too abstract to be useful for our purposes.

2 | FROM WORK ANALYSIS TO ASSESSING CLINICAL COMPETENCIES

To identify clinical competencies, one could use an *inductive* method in which generic rules are derived from a set of specific cases. To define which set of cases is relevant, critical incidents—*anecdotal incidents of exceptionally good and exceptionally poor behaviour*—are often used to extract a set of critical requirements for a profession.¹³ This is the oldest work analysis technique known.¹⁴ This method can be seen in recent initiatives to formulate a set of professional activities that recently graduated medical doctors need to be able to execute.^{8,15} These initiatives regarding undergraduate curricula followed a shift to focusing on entrustable professional activities (EPAs) in most postgraduate medical training programmes.¹⁶ Examples include *clinical consultation* and *communicating and collaborating with colleagues*. Generating a proposed set of professional activities then allows one to decide which samples are required during clinical training to enable statements about students' ability to execute these activities. A risk inherent in this method is that constructs or competencies are not

adequately covered by the samples. Whereas competencies describe people's capacities (e.g., medical knowledge, communication skills and professional attitude), EPAs are units of professional practice (e.g., the task of conducting a laparoscopic cholecystectomy).¹⁶ Adequate sampling, therefore, implies that samples are not deficient and do not contain irrelevant aspects.¹⁴

An alternative is the *deductive* method, which implies applying a general rule to a specific situation. For clinical training, this would mean that the set of required samples is generated deductively on the basis of an attribute-oriented job analysis (competencies, traits, aptitudes; e.g., Cook¹⁴), a literature review or an existing theory.¹⁷ The intended learning outcomes for undergraduate medical training based on the CanMEDs competencies described in the recently updated Raamplan for Dutch medical schools offers an example of such a job analysis.⁸ The main challenge, however, is to ensure that all these learning outcomes are intentionally assessed during clinical training. This requires deliberate sampling of various situations in which students' performance is going to be assessed.

Given the described advantages and disadvantages of both the deductive and inductive approaches in clerkships, we propose the following and use Smith's model to explore how a hybrid method might be created:

Proposition 1. To ensure that all learning outcomes are intentionally assessed during clerkships, a hybrid method is needed, that is, a combination of the deductive and inductive approach to sample (workplace-based) assessments in clinical training.

3 | UNIVERSALS, OCCUPATIONALS AND RELATIONALS

The competencies required to work as a doctor belong to different categories of job characteristics in Smith's model.¹⁰ He focuses on the degree to which competencies are context dependent, differentiating between the following three domains of job characteristics: *universals*, which are characteristics required by all working individuals and thus not context-dependent; *occupational*s, which refer to characteristics required by certain jobs but not others that, therefore, are dependent on task-related features of an occupation; and *relationals*, which are required in a specific organisational context with habitual ways of working together. We will now discuss these three categories more in depth by offering intuitive examples regarding what competencies fit into each category and deducing, thereby, whether extracting signs (claims about general competencies) from samples of behaviour is likely to be easy or problematic.

4 | UNIVERSALS

Universal competencies are relevant across jobs and contexts. Next to Smith's universals of cognitive capacities, vitality (physical and mental

energy¹⁸) and work ethos (including conscientiousness), we also regard agility (aka adaptability¹⁹) as a present-day universal. Changes in work through new technology and changing patient populations require employees to be much more adaptable. Smith regards universals as constructs that would hold a nonzero predictive validity for all jobs.

The question then becomes whether sampling 'universal' behaviour in a specific medical work situation can predict this behaviour in any other medical work and setting. Our premise is that if such sampling is done reliably and in a construct-valid way,¹⁴ this indeed is possible. In support of this notion, a meta-analysis by Gonzalez-Mulé et al.²⁰ shows how measures of cognitive ability among working adults across many job types and settings strongly predict general job performance, while Koczwara et al.²¹ in a longitudinal study, reported that cognitive ability among graduates as measured in diverse ways (a GMA test and more contextualised measures) predicted simulation exercises related to medical work, such as a simulated patient consultation and a work-related group discussion used for selection into UK-training for GP. Moreover, a meta-analysis by He et al.²² shows measures of conscientiousness to be predictive across occupations and types of performance ratings, whereas Hojat et al.²³ report this trait to be the strongest and most consistent personality predictor of performance in medical school and in the medical profession. These findings suggest that the predictive validity of these constructs holds not only across jobs and work settings but also across the way these universals are measured. Although we are not knowledgeable of meta-analyses related to vitality, agility and work ethos, we would predict similar findings for these concepts. This, therefore, leads us to the following:

Proposition 2. Reliable and construct-valid sampling of universal competencies, which are cognitive capacities, vitality (physical and mental energy), work ethos (including conscientiousness) and agility, in a specific medical work setting is generalisable to any other medical work setting or medical occupation.

This proposition implies that the way universals are sampled is not particularly important, provided it is done reliably and in a construct-valid way. In other words, multiple approaches to gathering evidence (i.e., different assessment methods) can be used, but all need to be supported by validity evidence. Therefore, next to traditional methods such as ratings by supervisors, other methods like situational judgement testing,²⁴ 360° ratings by relevant others such as colleagues, nurses, patients and supervisors,²⁵ and 'walk-throughs' during which students explain and describe how they execute certain tasks instead of actually executing them¹⁴ could be used. Factors influencing the choice of assessment method then could be very practical, driven by things such as time and costs. While the stability of personality traits across situations, contexts and time is a complex issue that remains a matter of discussion among personality psychologists,^{26,27} conscientiousness is one personality construct that tends to show a relatively high degree of stability, suggesting it is

assessable in many situations, contexts and times without harming generalisation of the assessment.

5 | OCCUPATIONALS

According to Smith,¹⁰ occupationals are characteristics that are relevant to particular jobs or occupations. Examples of occupationals are specific cognitive abilities (i.e., numerical or verbal skills), specific knowledge and certain personality traits that enable effective performance in a particular job. Smith reports that these lower-level abilities show considerable overlap with general cognitive ability (a universal) and are linearly related to performance.

Job clustering often starts with the notion that jobs, or groups of jobs, require specific abilities. An example of a specific cognitive ability for doctors is clinical reasoning, which could be considered as the cognitive processes of health professionals through which they interpret patient information to come to a diagnosis and/or treatment plan. Clinical reasoning is arguably the defining characteristic of the medical profession²⁸ and therefore can be considered as a 'generic' occupational for medical doctors. On the other hand, there are also occupationals that differ in their relevance between specific medical domains and therefore may be considered 'specific' occupationals. Think for example of eye-hand coordination, which is particularly relevant to the surgical disciplines, but less so for general medicine.

Proposition 3. For the medical profession, occupationals can be divided into generic occupationals, which are relevant to all medical doctors' occupations, and into specific occupationals which are relevant for medical doctors within specific disciplines.

The category of generic occupationals includes, but may not be limited to, clinical reasoning, integrity, communication skills (communication with colleagues and patients), concern for others and for society, stress tolerance, self- and other-focused learning orientation.

The category of specific occupationals consists of competencies that are relevant depending on the specific discipline. These comprise competencies such as subfacets of generic occupationals (e.g., communication with children as subfacet of communication), eye-hand coordination/manual dexterity (surgery), spatial awareness (radiology) and vigilance (anaesthesia).

Given the distinction between generic and specific occupationals, the question then is to what extent and how can medical schools sample for both types. As medical schools seek to prepare students for a wide range of future work contexts, it seems logical to focus on generic occupationals, such as clinical reasoning, during medical school. But, while it is agreed that all medical doctors must be competent in clinical reasoning, challenges exist in defining, teaching and assessing it.²⁸ The main issue here, as before, is context specificity²⁹;

that is, the context of the clinical task will have an influence on the student's performance and the type of clinical reasoning required. This implies that we need to estimate competence in the generic occupational *clinical reasoning* by means of (observed) performance in specific situations. But how do we decide on the specific situations that are required for us to be able to prove at the end of medical school that a student is generally competent in a generic occupational such as 'clinical reasoning'? We believe that here the use of a hybrid approach to sampling is likely to be particularly useful to help improve statements about medical students' competences across contexts and tasks.

As alluded to above, the hybrid approach could start inductively by identifying a set of 'critical' professional activities, which could take the form of EPAs. In order to be useful for sampling, however, these activities should be concrete and contextualised. For example, rather than using the generic activity *clinical consultation*, one could think of the different types of clinical consultations that medical students must master during clinical training, such as the neurological or the psychiatric consultation. Recently, as many as 16 subactivities, so-called nested EPAs, have been suggested for the core EPA *clinical consultation*.¹⁵ We propose, therefore, that sampling subactivities is particularly relevant for assessing specific occupationals. The question then becomes how we subsequently use activities or situations within which specific occupationals are localised, to infer something about generic occupationals, which after all is our goal?

Here, the addition of a deductive approach to sampling assessment situations becomes useful. To explain, let us reflect on another skill that is considered important to all medical doctors: communication. Given that communication skills are highly context specific, Van der Vleuten et al.³⁰ propose the use of a programmatic assessment approach to communication assessment, which implies using different methods to collect multiple samples of students' communication skills over a longer period of time. Although we agree with the value of this inductive approach of 'generalising' competence from broad sampling, we think it is important that a systematic, deductive approach based on competency frameworks and theoretical models be overlaid on this longitudinal approach to ensure adequate sampling of activities that capture the spirit of what is meant by communication competence. Another example, consider the model of teamwork knowledge, skills and abilities defined by Stevens and Campion.³¹ This model, consisting of, among others, interpersonal competences like conflict management competences, shared problem solving competences and communication competences, was used to deductively design a situational judgement test for selection procedures based on teamwork.³¹ To sample activities in an ideal manner, the patient mix (i.e., the quantity and the diversity of the patients encountered by students) must be taken into account. As suggested by De Jong et al., we might want to consider tailoring the patient mix to specific learning goals and needs of individual students.³² In sum, in contrast to starting from incidents or situations only, the deductive approach may lead to the conclusion that some aspects of communication would better be measured in other ways than by the typical way of direct observation in real clinical practice.

Proposition 4. A deductive approach supports reliable and construct-valid sampling of generic occupational, whereas an inductive approach is particularly useful for supporting reliable and construct-valid sampling of specific occupational.

According to Smith, certain noncognitive competencies may enhance performance in particular occupations. Generally, it is unlikely that attributes that are important for performing as an accountant are the same as those important for performing as a firefighter. But how is that for medical doctors? Are there noncognitive competencies that are important for all medical doctors, despite their specialty, and are there such characteristics that are important for some, but not all doctors, depending on their specialty? Using job analyses, Patterson et al.³³ identified a wide range of attributes beyond clinical knowledge and academic achievement that may help ensure that doctors are matched with a specialty for which they have a particular aptitude. They concluded that although there were more similarities than dissimilarities between the studied specialties, differences in perceived importance by professionals from the different specialties indicate context specificity of competency domains. As an example, within paediatrics, communication and empathy were rated as most important, whereas in anaesthesia, integrity and vigilance got the highest ratings. In both paediatrics and obstetrics and gynaecology, team involvement was considered more important than in anaesthesia. Next to differences in relevance, differences in context between specialties also lead to qualitative differences in required competences.³³ This implies that a generic occupational like communication could be broken-down into specific occupational. For example, being able to switch in conversation style from adult to child interactions is different from being able to discuss psycho-sexual problems with patients. By providing specialty-specific, contextualised information on relevant competency domains, job analyses for various specialties could not only inform assessment sampling but also self-selection and career choice of medical students. Nevertheless, as medical schools must prepare their graduates for a wide variety of postgraduate training programmes, medical students need to demonstrate a minimum standard across competencies that are considered important for all medical doctors, meaning that the generic occupational are particularly important in the early stages of one's career. Later, domains that are identified as priorities in a particular specialty (i.e., the specific occupational) could be used for designing selection and assessment procedures for that specialty.

Proposition 5. (a) Generic occupational should be given more weight than specific occupational in designing assessment procedures for medical students. (b) Specific occupational should be given more weight than generic occupational in selection and assessment procedures for medical residents; they could, however, also be used for medical students' self-selection into medical specialties.

6 | RELATIONALS

Relationals, as Smith¹⁰ suggested, are characteristics that are relevant to specific work contexts. In one hospital department, the relationship between a medical doctor and his or her team can be harmonious but in another department relationships in a team may be very litigious. Such differences could have a marked effect on performance as the skills and knowledge required to navigate those situations differ markedly. Consequently, relationals concern the so-called person-organisation fit and demand an approach that investigates whether a specific person and a specific organisational context are a match. A vast amount of research into person-organisation fit has been conducted by work and organisational psychologists, mostly focusing on values, namely, whether a person's values are commensurate with the values of a specific organisation. Meta-analyses by Kristof-Brown et al.³⁴ and Arthur et al.³⁵ report that a values-fit is related to a person's job satisfaction, to turnover (negatively) and to commitment to the organisation.

Examples of values are striving for prestige, having a strong achievement-orientation, and a focus on commitment to the welfare of others.³⁶ A certain academic hospital may emphasise possibilities for advancement and prestige, which may attract people who find advancement, recognition, and social status important. Another may be strongly patient oriented, valuing service orientation more than prestige. In yet another hospital, the focus could possibly be on independence, allowing its employees to work on their own and make decisions, which may form a match for those who prefer strong individual responsibility and autonomy. Based on the general values model of Schwarz,³⁷ De Clercq et al.³⁶ developed a values model for organisations, grounded on the underlying themes of self-enhancement (e.g., achievement, prestige and power), openness to change (e.g., need for stimulation and self-direction), self-transcendence (e.g., social commitment) and conservation (e.g., tradition, conformity). A recent qualitative study among medical students by Gennissen et al.³⁸ found three career-related values among these students, namely, a career orientation concerned with achievement and recognition of one's work, an orientation towards lifelong self-development and an orientation valuing work-life balance. In these three values, De Clercq et al.'s themes of self-enhancement and openness to change can easily be recognised. Future research could investigate to what extent both other themes in De Clercq et al.'s value model, that is, self-transcendence and conservation, are relevant for medical students.

Next to differences in values, work settings may differ in structural aspects, such as a high- versus low-resource environment, the absence or presence of electronic performance monitoring and needed speediness.^{20,39} Such factors may influence the interpretation of the tasks and skills of the job by the organisation and by the individual, and thus the degree to which this interpretation is shared. According to Smith,¹⁰ values and the shared interpretation of the job are among the most relevant subdomains in the category of relationals.

We believe that congruence between a specific medical setting and the medical student in terms of relationals—values, ideals and principles and job/task interpretation—is important for the student's future work-related well-being and the student's commitment to one's medical job in a specific context. Therefore, clerkships should give students the opportunity to explore whether they experience a match in terms of their value profile. In future selection procedures in a specific hospital department, this match could play a role. To illustrate, it may well be that the work–life balance values of a medical student are more fitting for one hospital department than for another. Focusing on relationals, we therefore propose that:

Proposition 6. (a) Clerkships across several settings (different hospitals, locations with low- and high resources, etc.) should provide medical students the opportunity to experience the degree of fit between their values in terms of De Clercq et al.'s work-related values model and their job/task interpretation on the one hand, and the values and job/task interpretations that characterise different settings on the other hand. (b) The degree of value-fit and shared job-interpretation fit might inform (self-)selection for specific hospital settings.

To this end, we think that medical students should develop their own self-assessed values and job interpretation profile³⁹ while they conduct their clinical training. Furthermore, while they follow a particular clerkship, we advise that they reflect upon the degree of fit between their values and job interpretation, and the values and job interpretations that are characteristic for that specific context, to be able to reflect whether they would be able to feel at home in such an environment.

An important issue connected to relationals is whether a measure of a medical student's 'fit' includes a match between a student and the *cultural* context of a specific hospital unit. This fit may refer to the student's chemistry with a specific supervisor, the team of colleagues³⁴ or with a certain demographic homogeneity or heterogeneity in terms of gender, lifestyle or language.³⁴ We believe that generalisation from samples of relationals, including ratings of the student's fit with the team and the hospital climate, to other medical work (contexts, specialisms, hospitals) to be particularly perilous, an oversimplification that should *not* be attempted. The reason for this is that such ratings will be too contextualised and unique for the work situation in which they were given. A classic example of such a rating component is *liking*. Interpersonal perception research from within psychology⁴⁰ clearly shows how liking is a fundamental judgement made about others which is strongly influenced by the so-called *relationship effect*. This effect is unique to the combination of a specific rater and a specific ratee and therefore, per definition, is not relevant outside that particular dyad of individuals. Another potentially complicating factor in generalising ratings from a specific hospital unit is the relational of *Trust*. From studies by Jones and Shah⁴¹ and Campagna⁴² (as cited in Kenny⁴⁰), it appears that for people who have worked

together for longer periods of time, a large part of the variance in ratings of ratee's trustworthiness could be attributed to relationship variance. In other words, trust is unique to the dyad of the rater and the ratee. Therefore, only when it is time to accept a job offer at a particular hospital unit for a considerable period of time will relationals, which are unique to that unit, become relevant. Following this line of thinking, we propose the following:

Proposition 7. (a) During clerkships, raters should refrain from giving ratings which include elements associated with the relationship effect (e.g., liking, trust and chemistry) and from value-fit and job interpretation fit evaluations. (b) Such relationship effects, however, become highly relevant at the point in time when the medical graduate applies for a job.

When providing ratings during clerkships, therefore raters need to be trained to become aware of nonrelevant factors that may influence their ratings of professional competences (e.g., liking and in-group favouritism⁴³).

7 | INTEGRATING SMITH'S THREE CATEGORIES

This paper started off by arguing that three different categories of characteristics required to work as a doctor can be distinguished. Universals hold for all jobs and therefore also for medical doctors; occupationals refer to characteristics required for all medical, but not other, jobs and include a subset of characteristics that are relevant only to subpopulations of medical doctors; relationals, in contrast, refer to characteristics that enable effective performance in particular settings, and which therefore are effective or not dependent on the specific organisation.¹⁰ Moving from universals to occupationals to relationals, the characteristics become more context dependent. This means that using samples of ratings during clerkships to generalise to claims about competence will be easiest for universals, but will require argued induction—or refraining from generalisation—for relationals.

One topic, affect,⁴⁴ remains hard to categorise using this model. At first sight, affect-relevant situations may appear to be relationals. For instance, to what extent is an environment characterised by work overload versus an optimal work volume, by goal disruptive events versus uninterrupted goal attainment, and by an authoritarian versus an open climate? For students who are highly agentic (achievement oriented), it is important to be knowledgeable of whether a specific hospital climate is characterised by high ambiguity, loss of control and obstacles in completing their work tasks and whether they are able to cope with such issues. Similarly, students who are highly communion oriented need to know how a particular hospital culture rates on issues of conflict in communication and problems in interactions with patients, and whether they can cope with such issues. However, to our view, it is not necessarily clear at this point whether affective reactions to such situations⁴⁴ are likely to be characteristics displayed

across *all* medical occupations, differ between *certain medical specialties*, or are specific to the climate within a *certain hospital department*. For now, we suggest treating these work characteristics as relationals to be assessed during job applications rather than drawing universal claims about trainees' affect.

For a different context, namely, personnel selection, Smith¹⁰ proposed that the predictive validity of a set of selection measures is a function of the domain of characteristics covered by the measures and the accuracy with which the domain is measured. To this end, he developed an algebraic formula. A modified version of the formula can be used for our purposes, namely, to guide thinking about performance appraisal of students during clinical training. In line with this notion, we derived a modification of this formula, as follows.

Algebraically, the quality of the assessment of medical students' universals (*U*), occupationals (*O*) and relationals (*R*) can be described as follows: the observed quality of the assessment of these competencies (Q_{AC}) is a function of the extent to which it measures each of the three domains (*U*, *O* and *R*) multiplied by their respective weight (*W*; importance), the sampling quality (*S*) and the intersubjectivity of the assessment (*I*). In this formula, *W* refers to the relative importance of the domain in relation to the decision to be made, whereas *S* and *I* are features of the accuracy of measurement (i.e., quality of the data). That is, they reflect whether the sample (*S*) of collected data points (e.g., observations and ratings) is large and representative enough to cover the competencies within the domain and whether or not that has been done as effectively as possible (*I*) (e.g., using informed and experienced observers vs. intuitive and naïve judgement). We used intersubjectivity to replace Smith's "objectivity (*O*) of assessment" given the fact that intersubjectivity is more realistic for the medical competencies of interest. In the formula, we distinguish between generic occupationals (*Og*) and specific occupationals (*Os*).

$$Q_{AC} = \sum (U \times W_U \times S_U \times I_U + O_g \times W_{Og} \times S_{Og} \times I_{Og} + O_s \times W_{Os} \times S_{Os} \times I_{Os} + R \times W_R \times S_R \times I_R)$$

This formula can be used to justify the quality of the assessment of students for which the educational programme management ultimately is responsible. This is done by explicitly explaining the importance given to the *U*, *O* and *R* competencies, and by defending the quality of sampling and intersubjectivity of the assessments on a pre-determined scale (for instance on a scale from -5 [*very bad sampling/intersubjectivity*] to +5 [*excellent sampling/intersubjectivity*]) (Table 1).

The below integrative table (Table 2) offers an overview of how, to our view, universals, occupationals and relationals can be assessed. It also illustrates how relevant generalisability issues can be dealt with for each competency domain, and several assessment recommendations. We maintain that the category of universals—if reliably measured and when having a relatively strong predictive validity coefficient—needs to have a comparatively heavy weight in selection and assessment for the medical profession. Focusing on universals is important when 'decisions about placement are likely to be delayed' (Smith¹⁰, p. 18), as is the case for the population of medical students in clinical training we are focusing on. Universals also are highly appropriate in "turbulent" ...

TABLE 1 A hypothetical worked example of some assessment methods graded on components of the formula, intended to infer a part of the generic competency profile of medical students entering the labour market

Sampling	For which Intended competencies	Coverage of domains			Accuracy of measurement			Sampling quality for current purpose
		Universals	Occupationals-generic	Occupationals-specific	Relationals	Number of data points	Representativeness of data points	
Master Knowledge test (cf Van Andel et al. ⁴⁵)	Medical expert (cognitive ability)	Moderate	High	Low	Nil	High	High	High
Single direct observation by supervisor, e.g., neurological consultation	Medical expert Health advocate Communicating	Low	Moderate	High	Nil	Low	High	Low-moderate
Ratings by colleagues of daily functioning at department	Collaborating Organising	Low	Low	Low	High	Low	Low	Low

Note: The weights of the domains and therefore the suitability of a particular assessment method depends on the specific purpose—In the present case for the general competency profile of medical school graduates the most important are the universals and the occupationals-generic.

TABLE 2 An integrative overview of assessment and generalisability issues for universals, occupationals and relationals

Competency domain	Competencies belonging to the domain	Sampling for assessment-purposes	Generalisability issues	Recommendations for assessment
Universals	Cognitive capacities Vitality (physical and mental energy) Work ethos (including Conscientiousness), Agility	Can be of any kind, e.g., observations, self-report, walk-through, etc. In principle interchangeable.	No major issues, relatively context-free	Let practical arguments prevail for the choice of assessment
Occupationals	Generic Competencies (may not be limited to): clinical reasoning, integrity, communication skills (communication with colleagues and patients), concern for others and for society, stress tolerance, self- and other-focused learning orientation. Specific Competencies such as: communication with children and empathy (paediatrics), team involvement (emergency medicine), eye-hand coordination/manual dexterity (surgery), spatial awareness (radiology), vigilance (anaesthesia), the need for recognition (cardiology/neurosurgery)	Broad, but purposeful sampling of mainly direct observation on the workplace, supplemented with 360° feedback, SJTs	Individual observations are not really generalisable, therefore broad sampling required	Combine inductive and deductive sampling to cover all required competences. Focus on assessing generic occupational skills for medical students. Use specific occupational skills mainly for (self-)selection for postgraduate training.
Relationals	Values of self-enhancement (achievement, recognition), openness to change (self-direction, self-development, a quality-balance between work and life), self-transcendence (social commitment), and conservation (e.g., tradition, conformity) <i>Job interpretation:</i> for example, whether one assumes that the job concerns making important decisions or not, one is consulted before objectives are set (or not), includes support and help from colleagues and the boss (or not), working at a high speed and with high-level of resources (or not) Chemistry ('liking', 'click') with colleagues, team, a departmental climate, supervisors	Compare self-assessment by medical students of their values on a values-questionnaire and of their job interpretation on a questionnaire from the work design domain, ³⁹ with the (perception of or an estimated) value profile/job interpretation profile of a specific hospital context Do not sample these relationals, unless during application for a (tenured) medical job	Relationals are not generalisable beyond a specific medical/hospital context. Can form irrelevant influence on assessment of other competencies, thus diminishing their generalisability	Need of preparatory work: develop, e.g., Q-sort technology (Gemmissen et al. ³⁸), values-questionnaire (based on De Clercq et al.'s values ³⁶) and have an estimation or impression of a values-profile of medical context available. Similarly, such profiles are needed related to <i>job interpretation</i> , e.g., based on work in the area of work design. ³⁹ An estimation of the degree of values-fit and shared job interpretation is important for self-selection into a medical job in a specific hospital. Make assessor aware of this element of 'chemistry' to be able to restrain from it when needed. For example, by holding them accountable for their ratings, and by warning them to avoid the influence of performance-irrelevant antecedents of this 'chemistry'. ⁴⁶ Diminishing freedom for subjective rating, for instance is needed if demographics (e.g., gender and ethnicity) influence 'liking', such as through in-group favouritism. ⁴³

conditions where rapid change is anticipated',¹⁰ which is characteristic for the medical profession given effects of technological and scientific developments but also given the potential of governmental interference in the medical systems within a country. In some places, a substantial minority of medical graduates find jobs other than medical doctor (e.g., in pharmacy and policy adviserships⁴⁵), training for which by definition highlights the relative importance of universals. Naturally, generic occupationals to our view are essential as well, given that the vast majority of all graduates find a job as medical doctor,⁴⁵ whereas as argued before, relationals, in contrast, should receive a low weight during clinical training.

Taking a broader view on the issue of assessment and (self-) selection of medical students and predicting their future job success, many other factors will come into play. Among these are whether the assessments are seen as (legally) fair and appropriate by all stakeholders,^{14,21} not too costly but practical, and what portion of students can be expected to graduate.⁴⁷ These factors are only briefly mentioned here as they are beyond the scope of the present paper. We believe Smith's validity model, which emanated from the domain of psychology, forms a novel way to frame questions of the relevance and generalisability of professional skills assessment of students in medicine. The propositions put forward will need to be put to the test in empirical studies, but we expect them to provide a useful starting point to enable medical schools to move from convenience samples towards more purposeful samples guided by the purpose for which the assessments are being made. They should further help make decisions about how, how often and under which circumstances each of the qualities in these three categories has to be measured. In other words, what are the most appropriate sources of data to inform mastery of each of the desired qualities for medical graduates?

8 | CONCLUSION

Drawing from Smith's framework of universals, occupationals and relationals, and using the distinction between samples and signs, we outlined the implications of a division of competency domains for (a) issues of sampling for assessment purposes during clerkships, and (b) generalisability to signs. The focus of sampling in medical schools should be on assessing generic occupationals. However, we suspect that specific occupationals can help to measure generic occupationals and to inform medical students on future work contexts for which they have a particular aptitude. Students are advised to use relationals to self-assess their fit with specific medical settings. An adaptation of Smith's formula for universals, occupationals and relationals is provided to assist educators to conceptualise the quality of assessment of medical students.

ACKNOWLEDGMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

AUTHOR CONTRIBUTIONS

MPhB and KS-J contributed equally to the article, developing the idea and passing drafts back and forth through several iterations for editing and comment. CvA revised it critically for important intellectual content. All authors approved the final version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to its accuracy or integrity are appropriately investigated and resolved.

ETHICS STATEMENT

No ethical approval was sought for this project.

ORCID

Karen M. Stegers-Jager  <https://orcid.org/0000-0003-2947-6099>

Chantal E. E. van Andel  <https://orcid.org/0000-0001-5651-3129>

REFERENCES

- Williams R, Klamen D, McGaghie W. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med.* 2003;15(4):270-292.
- Frank J, Snell L, Sherbino J. *CanMEDS 2015 Physician Competency Framework.* Ottawa: Royal College of Physicians and Surgeons of Canada; 2015.
- Norcini J. ABC of teaching and learning in medicine work based assessment. *Bmj.* 2003;326(7392):753-755.
- Wernimont PF, Campbell JP. Signs, samples, and criteria. *J Appl Psychol.* 1968;52(5):372-376.
- Niessen S, Meijer R. On the use of broadened admission criteria for selection into higher education. *Perspect Psychol Sci.* 2017;12(3):436-448.
- Boyd RL, Pasca P, Lanning K. The personality panorama: conceptualizing personality through big behavioural data. *Eur J Personality.* 2020;34(5):599-612.
- Ten Cate O. Entrustment as assessment: recognizing the ability, the right, and the duty to act. *J Grad Med Educ.* 2016;8(2):261-262.
- Van der Pol M, Hacfoort M, Laan R. *The 2020 Framework for Undergraduate Medical Education in the Netherlands.* NFU: Utrecht; 2020.
- Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach.* 2007;29(7):648-654.
- Smith M. A theory of the validity of predictors in selection. *J Occup Organ Psych.* 1994;67(1):13-31.
- Borman W, Hanson M, Hedge J. Personnel selection. *Annu Rev Psychol.* 1997;48(1):299-337.
- Messick S. Validity of psychological assessment. *Am Psychol.* 1995;50(9):741-749.
- Flanagan J. The critical incident technique. *Psychol Bull.* 1954;51(4):327-358.
- Cook M. *Personnel Selection: Adding Value Through People—A Changing Picture.* sixth ed. Wiley Blackwell: Chichester; 2016.
- Ten Cate O, Graafmans L, Posthumus I, Welink L, Dijk M. The EPA-based Utrecht undergraduate clinical curriculum: development and implementation. *Med Teach.* 2018;4(5):506-513.
- Ten Cate O, Chen H, Hoff R, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Med Teach.* 2015;37(11):983-1002.

17. Weekley J, Ployhart R, Holtz B. On the development of situational judgement tests: issues in item development, scaling and scoring. In: Weekley J, Ployhart R, eds. *Situational Judgment Tests: Theory, Measurement and Application*. New York, NY: Lawrence Erlbaum Associates; 2013:157-182.
18. Op den Kamp EM, Tims M, Bakker AB, Evangelia Demerouti E. Proactive vitality management in the work context: development and validation of a new instrument. *Eur J Work Organ Psy*. 2018;27(4):493-505.
19. Doeze Jager-van Vliet SB, Born MP, van der Molen HT. Using a portfolio-based process to develop agility among employees. *Hum Resour Dev Q*. 2019;30(1):39-60.
20. Gonzalez-Mulé E, Mount MK, Oh IS. A meta-analysis of the relationship between general mental ability and nontask performance. *J Appl Psychol*. 2014;99(6):1222-1243.
21. Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Med Educ*. 2012;46(4):399-408.
22. He Y, Donnellan MB, Mendoza AM. Five-factor personality domains and job performance: a second order meta-analysis. *J Res Pers*. 2019;82:103848.
23. Hojat M, Erdmann JB, Gonnella JS. Personality assessments and outcomes in medical education and the practice of medicine: AMEE Guide No. 79. *Med Teach*. 2013;35(7):e1267-e1301.
24. Christian MS, Edwards BD, Bradley JC. Situational judgment tests: constructs assessed and a meta-analysis of their criterion-related validities. *Pers Psychol*. 2010;63(1):83-117.
25. Toegel G, Conger JA. 360-degree assessment: time for reinvention. *Acad Manag Learn Edu*. 2003;2(3):297-311.
26. Huang JL, Bramble RJ. Trait, state, and task-contingent conscientiousness: influence on learning and transfer. *Pers Individ Differ*. 2016;92:180-185.
27. Josef AK, Richter D, Samanez-Larkin GR, Wagner GG, Hertwig R, Mata R. Stability and change in risk-taking propensity across the adult life span. *J Pers Soc Psychol*. 2016;111(3):430-450.
28. Gruppen L. Clinical reasoning: defining it, teaching it, assessing it, studying it. *West J Emerg Med*. 2017;18(1):4-7.
29. Rencic J, Schuwirth L, Gruppen L, Durning S. A situated cognition model for clinical reasoning performance assessment: a narrative review. *Diagnosi*. 2020;7(3):227-240.
30. Van der Vleuten C, Van den Eertwegh V, Giroldi E. Assessment of communication skills. *Patient Educ Counsel*. 2019;102(11):2110-2113.
31. Stevens M, Campion M. The knowledge, skills and ability requirements for teamwork—implications for human-resource management. *J Manage*. 1994;20:503-530.
32. Jong JD, Visser M, Van Dijk N, Van der Vleuten C, Wieringa-de Waard M. A systematic review of the relationship between patient mix and learning in work-based clinical settings. A BEME systematic review: BEME Guide No. 24. *Med Teach*. 2013;35(6):1181-1196.
33. Patterson F, Ferguson E, Thomas S. Using job analysis to identify core and specific competencies: implications for selection and recruitment. *Med Educ*. 2008;42(12):1195-1204.
34. Kristof-Brown AL, Zimmerman RD, Johnson EC. Consequences of individuals' fit at work: a meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Pers Psychol*. 2005;58(2):281-342.
35. Arthur W Jr, Bell ST, Villado AJ, Doverspike D. The use of person-organization fit in employment decision making: an assessment of its criterion-related validity. *J Appl Psychol*. 2006;91(4):786-901.
36. De Clercq S, Fontaine JR, Anseel F. In search of a comprehensive value model for assessing supplementary person-organization fit. *J Psychol*. 2008;142(3):277-302.
37. Schwartz SH. Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In: Zanna M, ed. *Advances in Experimental Social Psychology*, Vol. 25. Orlando, FL: Academic Press; 1992:1-65.
38. Gennissen L, Stegers-Jager KM, Van Exel N, Fluit C, De Graaf J, De Hoog M. Career orientations of medical students: a Q-methodology study. *Plos One*. 2021;16(5):e0249092.
39. Parker SK, Morgeson FP, Johns G. One hundred years of work design research: looking back and looking forward. *J Appl Psychol*. 2017;102(3):403-420.
40. Kenny DA. *Interpersonal Perception: The Foundation of Social Relationships*. 2nd ed. New York: Guilford Publications; 2020.
41. Jones SL, Shah PP. Diagnosing the locus of trust: a temporal perspective for trustor, trustee, and dyadic influences on perceived trustworthiness. *J Appl Psychol*. 2016;101(3):392-414.
42. Campagna R. The Accuracy of Initial Trust Judgments. All Theses and Dissertations (ETDs). 2011;53.
43. Van Anandel CE, Born MP, Themmen AP, Stegers-Jager KM. Broadly sampled assessment reduces ethnicity-related differences in clinical grades. *Med Educ*. 2019;53(3):264-275.
44. Ohly S, Schmitt A. What makes us enthusiastic, angry, feeling at rest or worried? Development and validation of an affective work events taxonomy using concept mapping methodology. *J Bus Psychol*. 2015;30(1):15-35.
45. van der Velde F, van de Leemkolk B, Lodder A. *Loopbanen en Loopbaanwensen van Basisartsen Meting 2019 [In Dutch]*. Utrecht: Prismant; 2019.
46. Sutton AW, Baldwin SP, Wood L, Hoffman BJ. A meta-analysis of the relationship between rater liking and performance ratings. *Hum Perform*. 2013;26(5):409-429.
47. Cascio W, Boudreau J. *Investing in People: Financial Impact of Human Resource Initiatives*. Upper Saddle River, US: Ft Press; 2010.

How to cite this article: Born MP, Stegers-Jager KM, van Anandel CEE. Inferring signs from purposeful samples: The role of context in competency assessment. *Med Educ*. 2022;56(1):117-126. doi:10.1111/medu.14669