General principles for the assessment of communication skills

Claudia Kiessling (Germany), Geurt Essers (The Netherlands), Tor Anvik (Norway), Katarzyna Jankowska (Poland), Rute Meneses (Portugal), Zoi Tsimtsiou (Greece), Marcy Rosenbaum (Iowa, U.S.A.), Jonathan Silverman (UK), for the tEACH Assessment Subgroup

Aim of the paper

This paper provides a summary of the current literature on assessment of communication skills within health care education. The aim is to help teachers who are thinking about a way to assess their students and the results of their communication skills courses to grasp the most important aspects of a fair and state-of-the-art assessment system.

Assessment drives learning

Assessment helps you as a teacher to decide whether your students are fit for later professional life and have acquired sufficient skills to be able meet the demands of clinical reality (Hulsman, 2011). For learners, assessment helps to identify their learning needs. Content and format of your assessment tools will influence students’ learning behaviour, implicitly and explicitly. Students tend to study what they expect to be tested. If we do not assess communication skills, students could assume that communication skills are not important: assessment legitimises the subject. On the other hand, the way we assess communication skills will send out a strong signal to the students about what we as professionals think good communication skills are in clinical practice. Therefore, it is of utmost importance to be aware of these considerations.

Establish a coherent entity of teaching, training, and assessment

The content and form of assessment need to be aligned with the purpose and desired outcomes of the teaching programme (Norcini et al., 2011). The core message is: assess what you teach and train. A mature communication skills curriculum is based on theoretical principles and scientific evidence of effective communication. Educational objectives1 should be derived from these principles. Teaching methods should be selected to help students to meet the educational objectives. The curriculum should encourage cumulative learning and self-reflection and prepare students for their later professional life. The system of assessment should be based on the same theoretical principles and educational objectives (Duffy et al., 2002). The methods of assessment should mirror the methods of instruction. Students should be assessed according to the educational objectives and their level of competence. This can be achieved by a careful blueprinting2 where the content of the assessment is planned to be congruent with the conceptual frameworks and educational objectives of the curriculum. Blueprinting enables sampling from across the curriculum and guarantees that assessment is seen as an integral part of the communication skills curriculum as a whole.

Assess communication skills within the clinical context

Like other clinical competencies, communication skills seem to be content and context bound (Wass et al., 2001; Schuwirth et al., 2010). Learning should take place as close as possible to the later professional reality (Schuwirth & Van der Vleuten, 2004). This means that for health care professionals, communication skills should be taught and learned within the clinical context, either in clinical practice or in clinically relevant simulations (Silverman, 2009; Kurtz et al., 2003). Even in the early years of medical training, when students do not have a lot of patient contact, it is helpful to use clinical examples and simulations to teach basic communication skills. For students, it does not make intuitive sense to isolate communication skills from later professional practice. Similarly, the assessment of communication skills should also be made within a meaningful context e.g. combined with and integrated into the assessment of other clinical skills, clinical knowledge and clinical reasoning. For instance, assessing the communication process skills of gathering

---

1 An educational objective is a statement that provides a clear description of the outcome or performance you expect from learners as a result of a lesson or course. An objective should be SMART: specific, measurable, attainable, relevant, and timely.

2 A test blueprint can be described as a template of specifications that shows the structure of the test. It can include the test content, knowledge domains, or levels of expertise. In medicine, it can also include age and gender of patients, setting of care etc.
information could be undertaken in an integrated fashion with the content of history taking and the perceptual processes of clinical reasoning so that learners are assessed approaching a clinically realistic task in which the communication skills are necessary to achieve a clinical end.

**Establish a system of assessment with formative and summative assessment**

Assessment can have a formative purpose which means detecting strengths and weaknesses of students and guiding future learning, providing reassurance and providing motivation (“forming” students’ learning). Formative assessment is often informal, integrated into a teaching course, timely, and low stakes. Summative assessment usually is intended to achieve another purpose. The focus is on making a decision about competence and fitness for clinical practice (“summing up” students’ learning). Summative assessment acts as a barrier to protect patients and public by identifying incompetent health care providers (Epstein, 2007). Both formative and summative aspects of assessment are important for learners becoming a professional and should be taken into consideration when planning a comprehensive assessment system.

**Provide feedback that enhances learning**

There is a growing awareness of the relation between assessment, feedback and continuous learning (Norcini, 2011). Assessment, especially formative assessment, provides helpful feedback to learners and teachers. Assessment needs to provide as much constructive feedback as possible. Moreover, although summative assessment concentrates on pass — fail decisions, it is still useful to provide feedback to students. Adult learners seek feedback from external sources and provide helpful feedback for other learners. If you manage to strengthen the self-regulation of your learners, feedback and self-assessment will accompany and complement the formal assessment methods set by the institution. Constructive feedback from teachers and peers will help learners understand and accept the criteria that teachers apply in the assessment of their communication skills. A broader understanding of assessment would therefore include feedback, self-assessment and peer-assessment into an assessment system (Dochy et al., 1999; Gielen, 2007).

**Establish a multi-source and longitudinal assessment programme**

All assessment methods have strengths and weaknesses (Schuwirth & Van der Vleuten, 2003). “Each single assessment is a biopsy, and a series of biopsies will provide a more complete, more accurate picture” (Van der Vleuten et al. 2010). Therefore, it is important to try to establish the use of different assessment formats, multiple observations, and independent times of measurement in different settings. Epstein provides a helpful overview about different commonly used methods of assessment (Epstein, 2007). Depending on available resources, it is helpful to try to assess students with different methods over a longer period of time with a broad sampling. Among others, Rider and colleagues (Rider et al., 2006) have published an interesting example of a model for communication skills assessment across the undergraduate curriculum in medical education.

**Use methods of assessment according to your goals and levels of competence**

There are some general criteria to consider when you implement an assessment strategy (Schuwirth & van der Vleuten, 2004):

- Reliability (the degree to which the measurement is reproducible and consistent)
- Validity (whether the assessment measures what it claims to measure)
- Educational impact (impact on future learning and practice)
- Credibility (to students and faculty)
- Feasibility and costs (to individual trainee, the institution, the society)

Assessment is always a compromise between what is desirable and what is achievable, involving a trade-off between the weightings attached to each of the five components listed above. If the assessment is summative, the tools need to meet higher requirements, especially of reliability. Reproducibility of decisions based upon the test results is essential when we “sum up” students’ learning and draw decisions about their professional advancement. The following two statements highlight the balance between assessment purpose and assessment criteria in a concise way: “What is important is the relationship between the accuracy of the scoring and the level at which decisions are made” (Schuwirth & Van der Vleuten, 2004 p.804). “The
stake of the evaluation determine the requirements for the psychometric rigor of the measurement” (Duffy et al., 2004 p.500).

- **Use an established standard setting procedure**

  In summative assessments, the cut-off point for pass and fail needs to be defined (Friedman Ben David 2000). A transparent and robust standard setting guarantees a fair judgment of an examinee’s performance. There are different methods on how to set a cut-off score, with a large body of relevant literature. The choice of method is dependent on the assessment method and tool you are going to use, the resources you have and on the consequences of misclassifying examinees as having passed or failed (Wass et al., 2001; Friedman Ben David, 2000).

**Categorising assessment tools according to the stages of clinical competence**

A selection of different assessment tools has been provided by the tEACH website. All of them are applicable to the assessment of communication skills within an educational programme. Most of the tools in the tEACH website are directed at assessing communication in an entire patient–provider consultation. Some tools have been developed for specific professions; some are aimed at assessing specific topics (e.g. patient’s perspective and health beliefs, reasoning and decision-making).

There are different ways of categorising assessment strategies and instruments. A simple way is to divide them according to the level of competence you want to assess: basic and applied knowledge, performance in standardised quasi-real situations, and performance or action in the work-place (Miller, 1990).

- **Knowledge and written assessment: to assess basic and applied knowledge**

  Multiple methods of written assessments have been developed over the last decades. Questions can be open-ended or multiple choice (which means, selecting the best possible answer out of the choices from a list); context-rich (e.g. with a clinical vignette) and context-poor; media-enhanced (e.g. video, audio); paper-and-pencil-based or computer-based. Written assessment can be administered in a relatively short time and with high standardisation. With clear constructing and grading guidelines, different types of written assessments tend to reach a high reliability per hour of testing. Written assessment allows you to test different types of cognitive abilities: knowledge and understanding of facts, processes, and concepts. It does not allow you to test skills although it may be able to predict the performance of skills (Van Dalen et al., 2002). Nevertheless, there seems to be a place for written assessment in the field of communication, especially in the beginning of training and especially following observation of prepared videos (Humphris & Kaney, 2000; Hulshman et al., 2004). Newer written assessment methods like the script concordance test (Charlin & Van der Vleuten, 2004), key feature assessment (Page et al., 1995), situational judgement test (Strahan et al., 2005), and reflective portfolio assessment (Rees & Sheard, 2004) may well bring new stimuli to the field of assessing communication skills.

- **Performance in clinical simulations: performance in standardised quasi-real situations**

  The best known assessment method using clinical simulations is the Objective Structured Clinical Examination (OSCE) (Hodges et al., 2002; Schuwirth & Van der Vleuten, 2003; Newble, 2004). Its development and implementation into medical education has become a world-wide success in the last decades. Students complete a number of stations with standardized patients who are trained to consistently repeat typical clinical situations. Students have to perform medical interviews, physical examinations, clinical procedures etc. Raters - either standardized patients or clinicians - use either a checklist or global ratings to evaluate the students’ performance. In order to achieve acceptable reliability, it is important to consider the number of cases, stations, examiners and patients. For example, with an overall testing time of 3 to 4 hours and a minimum of 10 stations, OSCEs have shown to be reliable enough for high stakes examinations, e.g. a fed-

---

3 Standard setting is a method used in assessment to define levels of achievements or proficiency and the cut-off scores corresponding to these levels. A cut-off score classifies students whose score are below into one level and students whose scores are above into the next higher level (e.g. level “fail” and level “pass”). A variety of standard setting procedures are available to set a reasonable cut-off score for assessment purposes. A comprehensive introduction about different methods is given by Friedman Ben David (2001) and PMETB (2007).
eral licensing examination (Van der Vleuten & Schuwirth 2005; Wass et al., 2001). Although OSCEs have been proven to be a worthwhile method for the assessment of clinical skills, there is still a need for further evidence. Additional research is needed around scoring and standard setting (Norcini et al., 2011). There is an on-going discussion about the use of checklists versus global ratings for the assessment of communication skills. Important aspects for consideration include among others reliability (e.g. between different raters, between different OSCE stations), validity (e.g. reducing complex competencies to easy measurable but trivial patterns of behaviours), and feasibility (e.g. amount of time for rater trainings). At the moment, there seems to be a slightly positive tendency for global ratings in concerns of validity (Regehr et al., 1998; Hodges et al., 1999; Van der Vleuten et al., 2010). However, the scoring method should match your educational objectives and teaching models for maximal educational impact (Duffy et al., 2004).

Simulation, the imitation of a real-world task or process, has become increasingly important for the assessment of such things as clinical reasoning and teamwork (Epstein, 2007). Assessment situations based on simulation can include simulated patients, computer simulations, mannequins, and high fidelity simulators. One of the most important tasks of simulation-based education is providing feedback to learners (Issenberg et al., 2005; Bradley, 2006). Therefore, these new developments also offer innovative research and assessment opportunities for training and assessing communication skills in health care settings.

• **Direct observation of performance or action in every-day clinical situations: performance or action in the work-place**

Unstructured and structured observations from supervisors, peers, co-workers, or other health professions are commonly used to evaluate learners’ performance with patients (Kogan et al., 2009). Many of these strategies are called “workplace-based assessments” (Govaerts, 2011). For example, the mini-clinical-evaluation exercise (mini-CEX) has become a helpful instrument to support structured feedback for postgraduate training in medicine (Norcini & Burch, 2007; Kogan et al., 2009). It turns the unstructured observation into a structured formative assessment, which provides feedback to learners and teachers to enhance learning and teaching. These kinds of workplace-based observations can be combined with written exercises and oral presentations, and documentation of these may well be collected in a learning portfolio (Friedman Ben David et al., 2001). The formative character of providing multisource feedback to guide future learning is certainly one of the most valuable advantages of all kinds of direct observations. Nevertheless, the need for effective supervision and follow-up of direct observation and feedback should not be underestimated.

Patients’ ratings may also be valuable for formative assessment. At the moment, more evidence is needed on how to establish a reliable and valid assessment system that includes patients as observers and raters of clinical performance (Epstein, 2007; Duffy et al., 2004). Patient-related outcomes can serve as an indicator of the quality of doctor/provider-patient communication, but to include patient ratings into an assessment programme is not an easy task. For example, patients and doctors do not necessarily agree about the quality of the interaction, and they may have contradictory goals. Patient satisfaction may not be an adequate indicator for high quality medical care. A satisfied patient can still have an incompetent doctor. “Defining functions and endpoints and connecting those to theory will improve our understanding of the effectiveness of communicative behaviour. Eventually, this will be of uttermost importance to teaching and clinical practice” (De Haes & Bensing, 2009). Nevertheless, some validated patient rating instruments are promising for medical education (Reinders, 2009). Pre-planned but unannounced standardised patients who present incognito in real clinical settings (Siminoff et al. 2011) may also be particularly valuable for postgraduate or higher level training, and can be used for the assessment of diagnostic reasoning, treatment decision, and communication skills.

• **Ratings based on videotaped encounters**

Another approach to assess the communication and interaction with patients and standardised patients is the rating of videotaped real-life or simulated encounters. It allows students to review and analyse their own behaviour, third persons (including peers) can be invited to assess the performance, and raters can explain or justify their scoring in a documented way. Similar to simulation-based assessments, the quality of real-life encounters can be measured by rating scales and checklists. In addition, measurement is also pos-
sible with interaction analysis coding systems (Boon & Stewart, 1998). This kind of performance assessment is particularly suited to postgraduate training. If an adequate case mix is made, a sample of eight consultations can suffice for a valid and reliable assessment (Ram, 1999).

- **Attitudes, values, and “multidimensional constructs”**

  The most difficult challenge in education is assessing the students’ attitudes, values, and “multidimensional constructs”. Multidimensional constructs are, for example, empathy or patient-centeredness when knowledge, skills, and attitudes all work together to form a whole. Stepień and Baernstein (Stepień & Baernstein, 2006) describe four dimensions of clinical empathy: the emotive dimension (the ability to imagine patients’ emotions and perspectives), the moral dimension (the physician’s internal motivation to empathize), the cognitive dimension (the intellectual ability to identify and understand patients’ emotions and perspectives), and the behavioural dimension (the ability to convey understanding of those emotions and perspectives back to the patient) (see also Mercer & Reynolds, 2002). In comparison to technical skills, like suturing or venepuncture, multidimensional constructs are much more difficult to measure. Hemmerdinger and colleagues (Hemmerdinger et al., 2007) recommend classifying instruments measuring such things as empathy from three different dimensions: self-ratings (first person assessment), patient-ratings (second person assessment), and observer ratings (third person assessment). First and second person instruments are often used in the field of education research. They may well have a potential to enhance self-reflection in communication skills trainings and further formative assessment. Third person instruments typically focus on the behavioural dimension for example demonstration of verbal or non-verbal clinical empathy. They are commonly used in OSCEs to assess behaviourally measurable skills rather than intention (e.g. Hodges et al., 2002).

### Search for evidence and look for advice

The last decades have brought a large array of assessment tools that can be used for assessing communication. Sometimes when you implement a new assessment system, it is difficult to anticipate the effects on students, other stakeholders, and the institution. There are always unintended effects of testing. Increasingly, “there is little agreement on ideal assessment tools” (Schirmer et al., 2005 p.185). And perhaps the ideal assessment tool does not exist. They all have their advantages and disadvantages and they need to match the educational context. Therefore, if you are considering communication skills assessment, apart from reading the literature, it is worthwhile talking to people who are experienced and experts in the field. Implementing assessment in the curriculum will need careful change management to guarantee sustainability and continuous improvement. One of the main goals will be to gain support from students, colleagues, superiors, and teachers for your plans.

### Acknowledgement

We sincerely thank Dr. Götz Fabry (Freiburg, Germany), Anja Görlitz (Munich, Germany), Dr. Robert L. Hulsman (Amsterdam, The Netherlands), and Tanja Pander (Munich, Germany) for critically reviewing the paper.

### References

• Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise, Acad Med 1999; 74:1129-34.
• Hulsman RL. The art of assessment of medical communication skills. Pat Educ Couns 2011;83:143-1444.


Further reading
Numerous studies have been made on how and why to assess competencies in communication. We give only a few references, mainly reviews or consensus statements, from the vast literature on this subject.


Wass V, Van der Vleuten CPM, Shatzer J, Jones R. Assessment of clinical competence. The Lancet 2001; 357:945-949,