

REVIEW

Assessment in anatomy

Erich Brenner^{1,5}, Andy R.M. Chirculescu^{2,5}, Concepción Reblet^{3,5}
and Claire Smith^{4,5}

¹Division of Clinical and Functional Anatomy, Medical University of Innsbruck, Innsbruck, Austria, ²Department of Anatomy and Embryology, Faculty of Medicine, Carol Davila University, Bucharest, Romania,

³Departamento de Neurociencias, Facultad de Medicina y Odontología, Universidad del País Vasco (UPV/EHU), Vizcaya, Spain, ⁴Anatomy, Brighton and Sussex Medical School, University of Sussex Campus, Falmer, United Kingdom, and ⁵Members of the Trans-European Pedagogic Anatomical Research Group (TEPARG)

SUMMARY

From an educational perspective, a very important problem is that of assessment, for establishing competency and as selection criterion for different professional purposes. Among the issues to be addressed are the methods of assessment and/or the type of tests, the range of scores, or the definition of honour degrees. The methods of assessment comprise such different forms such as the spotter examination, short or long essay questions, short answer questions, true-false questions, single best answer questions, multiple choice questions, extended match questions, or several forms of oral approaches such as viva voce examinations.

Knowledge about this is important when assessing different educational objectives; assessing educational objectives from the cognitive domain will need different assessment instruments than assessing educational objectives from the psychomotor domain or even the affective domain.

There is no golden rule, which type of assessment instrument or format will be the best in measuring certain educational objectives; but one has to respect that there is no assessment instrument, which is capable to assess educational objectives from all domains of educational objectives.

Whereas the first two or three levels of progress can be assessed by well-structured written exami-

nations such as multiple choice questions, or multiple answer questions, other and higher level progresses need other instruments, such as a thesis, or direct observation.

This is no issue at all in assessment tools, where the students are required to select the appropriate answer from a given set of choices, as in true false questions, MCQ, EMQ, etc. The standard setting is done in these cases by the selection of the true answer.

Key words: Assessment – Knowledge – Skills – Attitudes – Written exams – Practical exams – Structured observation – Portfolio – Extended matching questions – Spotter tests

INTRODUCTION

A series of cooperation programs focused on education were organized and financed by EU (1990-2013: Tempus, Erasmus-Socrates, etc.), to elicit and develop bilateral mobilities of teaching staff and students (undergraduate and PhD levels), in order to facilitate contacts and develop partnerships, at both levels. They were designed to evaluate the competency and professionalism, as a concrete starting point for ensuring a minimal compatibility of the education standards, between EU members and candidates and a common basis for a valid and equitable European Credit Transfer System (ECTS), generally accepted in all European countries, especially for medics and dentists. This raises questions of core curriculum or benchmark, organization of courses in units, methods of

Corresponding author: Erich Brenner. Division for Clinical and Functional Anatomy, Medical University of Innsbruck, Müllerstrasse 59, 6020 Innsbruck, Austria. Phone: +43 512 9003 71121; Fax: +43 512 9003 73112. E-mail: erich.brenner@i-med.ac.at

Submitted: 25 November, 2014. Accepted: 29 December, 2014.

teaching and assessment.

Assessment must ensure a good compatibility between different schools, in both the same country and in different countries, as well. The use of external examiners (largely used in UK and Ireland) has to ensure a higher rate of objectivity and a more constant level of exigency, as well as better dialogue and opinion exchanges between institutions.

From this perspective, a very important problem is that of assessment, for establishing competency and as selection criterion for different professional purposes.

In fact, assessment may be directed toward three main tasks:

1. the evaluation of achievements in educational objectives acquired during a whole term or a short-term intensive course;
2. the evaluation of achievements in educational objectives acquired during self-learning / training, according to a recommended curriculum;
3. the evaluation of efficacy of a lecture, conference, workshop, seminar or whatever – tested before and after the session.

Assessing the students' progress in anatomy does essentially not differ from assessing in other disciplines. Assessing in anatomy has to obey the same general parameters as there are objectivity, validity, and reliability.

Objectivity is of course important. All students must be assessed the same way, the same difficulty, and without distinction of person. An objective assessment will be widely accepted by the students. Objectivity becomes a topic of concerns when different assessors are satisfied with different levels of knowledge, skills, etc. The magic word for objectivity is standardization. The more the questions or tasks for the assessment are unified throughout the assessors and the more the – expected – answers, actions, or even behaviours are standardized, the more objective is the assessment. As a matter of fact, there is no totally objective assessment as each assessment will contain inherent biases built into decisions about relevant subject matter and content, as well as cultural (class, ethnic, and gender) biases.

Validity is an often stressed and repeated term, which means that an assessment must measure what was intended to measure with this assessment. Validity refers to how well the assessment instrument actually measures the underlying outcome of interest; validity is not a property of the assessment instrument itself, but rather of the interpretation or specific purpose of the assessment instrument with particular settings and learners (Sullivan, 2011). This is important when assessing different educational objectives; assessing educational objectives from the cognitive domain will need different assessment instruments than assessing educational objectives from the psycho-

motor domain or even the affective domain. Validity itself can be split to four different aspects. The first of it is the content validity, which means that the assessment's content must measure (pre-) defined educational objectives. The second aspect is the criterion validity, which means that the assessment's outcomes (scores) should correlate with an external reference, or is predictive. The third aspect is the construct validity, which means that the assessment corresponds to other significant variables, for instance varying backgrounds. The fourth and final aspect seems quite simple, the face validity, which means that the item or theory behind makes sense, and it is seemingly correct to the expert.

Reliability is often intermingled with and misinterpreted as objectivity. Reliability means that an assessment consistently achieves the same results with the same or a quite similar cohort of students when administered repeatedly (Sullivan, 2011). There are several issues influencing an assessment's reliability such as ambiguous questions, inappropriate or incomplete instructions, or untrained test-takers. This is quite obvious in practical exams, but is even true in simple-looking multiple choice questions-exams. A reliable assessment is temporal stable when its repetition within a short time with the same or a very similar cohort will produce similar results. A reliable assessment possesses form equivalence when the performance of test-takers is equivalent on different forms of an assessment based on the same content. A reliable assessment has internal consistency when the responses are consistent across different tasks or questions.

Some other aspects of the students' activity are now considered as important to be included into the final marking. Among these additional aspects, which might be assessed, there are manual skills, presentation skills (display), writing skills, oral communication, ethical and professional development, position to science philosophy, experimental design / analysis of scientific literature, data analysis and interpretation, team work ability, individual initiative, originality, IT competence skills and many others, which completely lack from some schools marking criteria for assessment in anatomy. The use of a Personal Development Planning (PDP) system and of academic and personal tutorials appears as most recent and attractive approach. Anyhow, it can be included as one of the agreed methods of evaluation across EU. Establishing an Honours degree classification is considered very stimulatory to increase students' motivation for learning performances.

The selection of an assessment instrument is mainly depending on what the students should have learned. Indeed it has long been established that assessment drives learning (Biggs and Tang, 2011). Among the collection of more or less elaborated instruments there are some, which are suitable

ble for testing the students' achievements in the cognitive domain, and there are others, which can be used for testing items in the psychomotor domain. The vast number of instruments does also contain formats suitable for an assessment of educational objectives from the affective domain. There is no "golden rule", which type of assessment instrument or format will be "the best" in measuring certain educational objectives; but one has to respect that there is no assessment instrument, which is capable to assess educational objectives from all domains of educational objectives. Adopting the best assessment instrument is crucial, in that it needs to link together the aims and learning outcomes and test appropriately the right domain, this is termed constructive alignment (Gibbs and Habeshaw, 1989). To ensure alignment a blueprint is constructed in which the learning outcomes are matched to the examination mode and the questions asked. Biggs' (1999) model of constructive alignment suggested that when assessment is designed, a three-stage model should be used:

- 1) Identify clear learning outcomes that students can achieve and demonstrate (i.e. what is achievable for that point of study and from previous experience).
- 2) Design appropriate assessment tasks that will directly assess whether each of the learning outcomes has been satisfactorily met (can MCQ or EMQ questions assess if a student has the knowledge to identify an object?)
- 3) Design appropriate learning experiences or opportunities, so that the assessment tasks are achievable (formative assessments).

In the following, we will use a very simple example, the glenohumeral joint (GHJ). We will try to establish educational objectives from the cognitive domain, the psychomotor domain, and even the affective domain.

GENERAL PROBLEMS IN ANATOMICAL ASSESSMENT

The main question in planning anatomical education and therein assessment is, what do medical courses need to assess in students. The main question is, in fact: "How is the content of anatomical education defined that a certain medical school expects its students to learn?" Possible answers include specific objectives given to the students, general goals from which students must produce specific objectives for themselves, and a curriculum composed of core portions and/or extended (or elective) portions (Chirculescu et al., 2007).

An evaluation of the students' "core" knowledge requires an agreed "core" syllabus. For each type of assessment, this is a compulsory condition. In this respect, ASGBI produced and published a proposed outline of the core knowledge of anatomy, as a benchmark that its Education Committee felt

should be expected of medical students and young doctors (McHanwell et al., 2007). Another approach was made in the AMEE guide No. 41 (Louw et al., 2009).

All thinking processes must be based on underlying information and the question whether the students have that underlying information. When this question is answered satisfactorily, the next question is if the students can use that information in a logical, deductive, correlative manner in order to arrive at diagnoses, treatments, management plans, etc. Finally, the question has to be, whether the students can show that they know the evidence on which their thinking is based. In relation to assessment, this means that different tools are needed for evaluating core knowledge, the extension of knowledge beyond core, establishing an Honours degree classification, for establishing competency, or as a selection criterion for different professional purposes.

Assessing knowledge – the cognitive domain

The cognitive domain is the best known and most assessed domain of educational objectives (Anderson et al., 2001; Bloom et al., 1956; Krathwohl, 2002). This cognitive domain comprises at least two dimensions, the dimension of (possible) contents and the dimension of the progress, what a learner should be able to do with the content. [As a matter of fact, Krathwohl defined a "knowledge dimension" and a "cognitive process dimension", which constitute the cognitive domain (Krathwohl, 2002). In order to use these dimensions for all three domains, we redefined the "knowledge dimension" as the "content dimension" and the "cognitive process dimension" as the "progress dimension".]

The **content dimension** contains – of course – *factual knowledge*, i.e. in our case the knowledge of anatomical facts and the – medical/anatomical – terminology. But without knowing what to do with these facts, it will be impossible to assess the students' learning progress. Another item of the content dimension is the *conceptual knowledge*. Besides the factual and conceptual knowledge we find also *procedural knowledge*, i.e. the knowledge of methods or procedures. Finally, the content dimension of the cognitive domain contains *meta-cognitive knowledge*, i.e. knowledge of the principles and generalizations, theories, structures, and abstractions in a certain field.

Using the glenohumeral joint, examples of contents could be:

- *factual knowledge*: the ligaments of the GHJ
- *conceptual knowledge*: the principle(s) of a spheroid joint, with the GHJ as an example
- *procedural knowledge*: the process to evaluate the range of motion of the GHJ
- *meta-cognitive knowledge*: an overview of resources on the GHJ (textbooks, websites, models, etc.)

It is obvious that these examples per se cannot be assessed. It is necessary to define, what the student should be able to do with these contents. These activities are described by the **progress dimension**. Very basically, a student should be able to *remember* different contents. This can be assessed by any type of assessment targeting on the recall of facts. Nevertheless, anatomical education should or even must not stop with rote memorization (which is an often heard indictment against anatomy as teaching object). Thus, students should for instance – as the next item on the progress dimension – *comprehend* their knowledge, i.e. demonstrate their understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating the main ideas. Of course, comprehension is better than mere memorization, but there are even higher levels of progress. The next step is to *apply* or use the knowledge, i.e. solve a problem in new situations by using acquired knowledge, facts, techniques and rules in a different way. The next step is to use the acquired knowledge to *analyse* a certain problem or question, i.e. examine and break information into parts by identifying motives or causes, make inferences and find evidence to support generalizations. The ability to analyse is a major prerequisite for clinical reasoning, which itself is a major general educational objective in medical education. A successful analysis enables the students as the next step of progress to *synthesize* all available information, i.e. compile information together in a different manner by combining elements into a new pattern or proposing alternative solutions. Besides the synthesis there is also the task to *evaluate* the (available) knowledge, i.e. present and defend opinions by making judgments about the quality of information, validity of ideas or quality of work, in general based on a set of criteria. Such evaluations are important steps towards evidence based medicine.

Using the glenohumeral joint, examples of progress could be:

- *remember*: recall those ligaments, which are part of the GHJ
 - *comprehend*: compare the range of motion of the GHJ with the hip joint
 - *apply*: identify those structures (ligaments, bones, etc.) of and around the GHJ, which inhibit the abduction of the humerus beyond 90° and mechanisms involved in continuing abduction over that limit, i.e. understanding the three steps of arm abduction: initialization, GHJ abduction proper, increasing amplitude by rotation of scapula
 - *analyse*: analyse the cause(s) for the much higher incidence of anterior dislocations of the GHJ (vs. other types of dislocation)
 - *synthesize*: propose a program for physical training in order to prevent anterior dislocations of the GHJ
- *evaluate*: evaluate current treatment options for dislocations of the GHJ in terms of the patients' age and sex; correlate a GHJ-dislocation with fractures of clavicle or the humeral neck, differentiate between paralysis of accessory, axillary or suprascapular nerves; evaluate pain related to local damage or referred pain (e.g. gallbladder, heart)

It is obvious that these different levels of progress in learning cannot be assessed by the same examination. Whereas the first two or three levels of progress can be assessed by well-structured written examinations – such as multiple choice questions (MCQ), or multiple answer questions (MAQ), other – and higher level – progresses need other instruments, such as a thesis, or direct observation.

Furthermore, one has to consider the factor time. There are – and have to be – differences, when students are tested. Thus assessments might have three different aims (Sinclair, 1965):

- a. to establish the capacity of the students to acquire knowledge (assessment immediately on completion of the program);
- b. to determine their ability to retain knowledge (surprise assessment two years later); and
- c. to ascertain their capacity to recover their knowledge in a specified time (assessment of post-graduate students).

Assessing skills – the psychomotor domain

Similar to the cognitive domain, the psychomotor domain has a content dimension and a progress dimension (Harrow, 1972; Simpson, 1971). The **content dimension** comprises skills or competencies such as manual skills, perceptive skills, and psychosocial (or communicative) skills.

Using the glenohumeral joint, examples of content could be:

- *manual skill*: perform a closed reduction of an anteriorly dislocated GHJ, or local anti-inflammatory / anaesthetic injection / infiltration
- *perceptive skill*: assess the GHJ and its surrounding bursae for crepitus
- *communicative skill*: document a case of successful reduction of an anteriorly dislocated GHJ and write an appropriate physician's letter

The **progress dimension** starts with the *perception*, i.e. the process of becoming aware of objects, qualities, or relations. The perception is followed by the *set*, a preparatory adjustment or readiness for a particular kind of action or experience. The next step is the *guided response* as an early step in the development of skill and comprises both imitation and trial and error. The guided response is followed by the *mechanism*, where the learned response has become habitual. Whereas the mechanism addresses mainly simple skills, more complex ones can be addressed as *complex overt*

response, which resolves uncertainties. Finally, a skill is at the level of *automatic performance*. Miller reduced this progress dimension for clinical skills assessment to four steps: “knows”, “knows how”, “shows how”, and “does” (Miller, 1990); actually with the first two steps similar to the progress dimension of the cognitive domain of procedural knowledge. Another simplified progress dimension comprises only three steps, (1) *imitation*, (2) *control*, and (3) *automatism* (Guilbert, 1998).

Using the glenohumeral joint, examples of progress could be (note that the terms of the progress dimension have been transformed to verbs):

- *imitate*: imitate a puncture of the GHJ
- *has control*: perform a closed reduction of an anteriorly dislocated GHJ
- *does automatically*: assess the range of motion of the GHJ routinely in different subjects

Obviously, educational objectives from the psychomotor domain cannot be assessed by multiple choice questions tests. When testing skills, the candidate has to perform the skill and the assessor has to judge the quality of performance, best by using a standardized score-card. Another option would be to assess the final outcome of the skill, i.e. the “product”. When assessing the skill of performing a subcuticular suture, for instance, either the process of suturing itself can be observed (in a structured manner), or the suture itself can be assessed.

Another example of assessing communicative skills is given at Brighton and Sussex Medical School, where a written communication project that is directed at writing for a lay audience, was incorporated as the main coursework element (Evans, 2007). For this summative assessment, all students are ‘commissioned’ by the editor of a fictitious newspaper or popular magazine to write a short lay statement of no more than 500 words in response to a letter from a member of the public regarding a defined clinical condition associated with reproductive or loco-motor anatomy. Statements were assessed against a defined set of marking criteria, which focus on aspects such as interest, readability and presentation and each component was classified against an A-E grading scale.

An element unique to anatomy is its three dimensional nature and this ties together the cognitive and the psychomotor domain of the subject.

Spatial ability is the ability to perceive, retain and recognize or reproduce three-dimensional objects in their correct proportions when they are rotated in space, translated, juxtapositioned, projected, sectioned, re-assembled, inverted, re-orientated or verbally described.

In anatomy spatial ability has been found to be a good predictor of students’ success in learning anatomy and examination performance. Spatial ability might be even more important than the type of educational materials that are studied (Garg et

al., 2001). It is understood that spatial ability in anatomy can be trained (Fernandez et al., 2011), in particular the ability to judge distance (metric ability), including depth perception. In attempting to explain this unique element of anatomy the term ‘Touch mediated perception’ (Smith and Mathias, 2010) has been created. Current thinking has not yet explained how spatial ability and understanding the three dimensional nature of anatomy can best be assessed. Having said that three dimensional understanding is essential in applying knowledge, as is demonstrated by damage to underlying structures in surgery where this awareness is lacking (Ellis, 2002).

Spatial ability is a skill that substantially helps professional training and exercise of any health-care career. Spatial ability can be strongly developed with dissection of human cadavers complemented with the observation of radiological images and other diagnostic images. Indeed, all these tasks if practiced together create mental images of the structure of the body in the person making. These mental images are essential not only for surgeons but even more to perform physical exams and interpret radiological images. We consider that the best method in order to assess the spatial ability and three dimensional understanding of the students is by structured practical exams and if possible the exam in the lab of dissections made by them, which should be explained by the dissector during *viva voce*. Portfolio is also a good option.

Assessing attitudes – the affective domain

The affective domain is the most frightening for assessors. Nevertheless, also the affective domain can be analysed for its content dimension and progress dimension (Anderson et al., 2001; Krathwohl et al., 1964) and therefore prepared for assessment. That might be particularly important for all examinations a future physician has to pass, from the very beginning until her or his final qualification.

First of all, there are definition problems for the **content dimension**. In general, the content dimension describes the way people react emotionally and their ability to feel another person’s pain or joy. Affective objectives target the awareness and increase in attitudes, beliefs, mental images, emotions, feelings, and related behaviours.

In contrast to the content dimension, the **progress dimension** is easier to define. First of all, there is the level of *receiving*, i.e. a phenomenon of awareness, willingness to hear/see/feel, and paying selected attention. Without receptivity no learning can occur. As the next step follows the level of *responding*, i.e. there is an active participation, where the learners attend and react to a particular phenomenon. This is superseded in the next step by *valuing*, where student attaches a value to an object, phenomenon, or piece of infor-

mation. The next step is *organizing*, where different values, information and ideas are put together and the student accommodates them within her/his own mental representations; this is important for being reflective. Finally, one is *internalizing* the educational objectives, when the student holds a particular value or belief that now exerts influence on his/her belief and behaviour, so that it becomes a characteristic, thus building up a value system – including the targeted educational objectives – that controls her/his behaviour. A simplified version of the progress dimension is limited to the steps “receptivity”, “response” and “internalization” (Guilbert, 1998).

Examples, not directly related to the glenohumeral joint, could be:

- *receives*: listen to others with respect
- *responds*: assists patients to express their concerns towards surgery
- *values*: is sensitive towards cultural differences (shows value diversity)
- *organizes*: organizes values into priorities by contrasting different values
- *internalizes*: displays a professional commitment to ethical practice on a daily basis

As it is the case with the psychomotor domain, the affective domain can't be assessed by simple written examinations such as a multiple choice questions test. This domain can be assessed by either structured observation formats or feedback formats.

Assessing professionalism

Professionalism, as manifested through a commitment to carrying out professional responsibilities, adherence to ethical principles, and sensitivity to a diverse patient population is coming to the forefront as one of the six major requirements in all residency programs (Escobar-Poni and Poni, 2006). Gross Anatomy laboratories, based on cadaver dissections, seem to provide the more implicit skills to develop the basic elements of professionalism that are evaluated during clinical rotations (Escobar-Poni and Poni, 2006); they must be recognized and consecutively taught in order to evaluate them. These authors compiled a list of elements of professionalism commonly accomplished and evaluated during gross anatomy courses with cadaver-instruction (Table 1). They also suggested a set of tasks and activities, which can be included as learning objectives that can be taught, evaluated, corrected, and awarded. These tasks and activities comprise the creation of a “clinical anatomy chart”, the compilation of “progress notes”, the establishing of a “box of observations and suggestions”, the formal establishing of the “dissecting team”, “peer-teaching”, “commemorative services”, “interaction with the relatives of the donors”, “peer review evaluations on professionalism”, and finally the development of “self-reflection and journals/portfolios”.

Another example of assessing the development of the attributes of professionalism was developed at the School of Medicine and Dentistry of The Queen's University of Belfast (Heyns, 2007). An educational strategy based on role-playing was developed to engage all students around the dissection table. Students received comprehensive background reviews on professionalism, its attributes and the identification of such attributes in the context of the dissection room. Roles, with specific duties attached, were allocated to each team member. Circulating academic staff members directly observed student participation and gave formative feedback. Students were given the opportunity to reflect on their ability to identify the attributes and reflect on their own and their peer's ability to develop and practise these attributes.

OPERATIONALIZATION – FROM EDUCATIONAL OBJECTIVES TO ASSESSMENT TASKS

Educational objectives consist of an **object** (usually a noun) and an **action** (or verb). The action generally refers to the targeted progress dimension, whereas the object describes an item of the targeted content dimension, which students are expected to acquire or construct (Anderson et al., 2001). In other words, educational objectives are a function of an action over an object. In order to make educational objective more readable, each objective should be preceded by a – more or less – general statement such as “the successful candidate will be able to ...” or “the student will be able to ...”.

Our previous examples will therefore be stated in the following manner.

- “The successful candidate will be able to ...*
- ...recall those ligaments, which are part of the GHJ;*
- ...compare the range of motion of the GHJ with the hip joint;*
- ...identify those structures (ligaments, bones, etc.) of and around the GHJ, which inhibit the abduction of the Humerus beyond 90°;*
- ...analyse the cause(s) for the much higher incidence of anterior dislocations of the GHJ (vs. other types of dislocation);*
- ...propose a program for physical training in order to prevent anterior dislocations of the GHJ;*
- ...evaluate current treatment options for dislocations of the GHJ in terms of the patients' age sex and profession;*
- ...imitate a puncture of the GHJ;*
- ...perform a closed reduction of an anteriorly dislocated GHJ;*
- ...assess the range of motion of the GHJ routinely in different subjects;*
- ...listen to others with respect;*
- ...assist patients to express their concerns towards surgery;*
- ...be sensitive towards cultural differences*

Table 1. Elements of Professionalism commonly accomplished and evaluated during gross anatomy courses with cadaver-instruction (Escobar-Poni and Poni, 2006)

-
- Accepts constructive feedback and modifies behaviour appropriately
 - Adapts style and content of communication appropriately when regarding each cadaver
 - Adheres to institutional policies and procedures
 - Adheres to local dress code
 - Admits errors and assumes personal responsibility for mistakes
 - Advocates for colleagues
 - Advocates for societal health issues
 - Arrives on time for scheduled activities and appointments
 - Attributes ideas and contributions appropriately to others
 - Balances personal needs and patient care obligations
 - Completes assigned share of team responsibilities
 - Conveys information and answers questions honestly and tactfully
 - Demonstrates appropriate boundaries for inter-professional relationships
 - Demonstrates sensitivity to power asymmetries in professional relationships
 - Discusses colleague-issues without using inappropriate labels or comments
 - Discusses donor-issues without using inappropriate labels or comments
 - Displays compassion and respect for all cadavers even under the most difficult circumstances
 - Engages in informal teaching and learning activities with colleagues as appropriate
 - Facilitates conflict resolution
 - Fulfills all laboratory and non-laboratory responsibilities in a timely manner
 - Improves team effectiveness through motivation and facilitation
 - Intervenes immediately when unprofessional behaviour presents clear and present danger
 - Maintains a positive attitude amidst increased and unanticipated additional work
 - Maintains composure during difficult interactions with colleagues
 - Maintains confidentiality of the cadaver information in public areas
 - Maintains thoroughness and attention to detail
 - Makes valuable contributions during class, rounds, or meetings
 - Offers advice when appropriate
 - Optimizes cadaver comfort and privacy when conducting dissection
 - Provides cadaver information to team members in a timely and effective manner
 - Reacts appropriately to other's lapses in conduct and performance
 - Requests help when needed
 - Responds appropriately to help a distressed or impaired colleague
 - Serves as knowledge or skill resource for others
 - Signs over and ensures coverage of cadaver duties when unable to fulfil responsibilities
 - Solicits and values input from colleagues when appropriate
 - Takes on extra work when appropriate to help the team
 - Takes steps to prevent repetition of error
 - Teaches and emphasizes tenets of professionalism when appropriate opportunities arise
 - Transmits accurate and detailed information for optimal transition of care
 - Uphold ethical standards in research projects and other scholarly activities
-

(shows value diversity);

...organize values into priorities by contrasting different values;

...display a professional commitment to ethical practice on a daily basis";

An example of intermediate educational objectives for an anatomical dissection class is given in Table 2 together with their classification and suggestions for assessing.

Of course there has to be a decision: which of these objectives are valid anatomical objectives, and which ones are clinical objectives to be allocated to several clinical specialties or disciplines.

This – political – decision should best be made collaboratively, e.g. by the educational committee, but when this is not attainable, it must be made within the discipline of anatomy itself.

Operationalization of educational objectives is the process of defining a fuzzy concept so as to make the educational objective clearly measurable and to test it in terms of empirical observations, i.e. examinations. In other words, operationalization is the process of strictly defining variables into measurable factors. In other words, an educational objective describes **who** (addressed audience, students, etc.) **does** (activity, behaviour) **what**

(subject matter, content).

When transforming educational objectives into assessment tasks we have to add at least a criterion, describing the targeted level of quality: “who does what **how well**”. The best would be also to describe the targeted setting, i.e. the location, time, amount, equipment, etc.: “who does what how well **in which setting**” or “who does what how well **under which circumstances?**”

When we want a student to be familiar with the basic terms of orientation (to be able to use these terms in a regular manner routinely; “in a regular manner” = how well, “routinely” = in which setting) we should not set up questions like “What does ‘distal’ mean?” or “Which axes define the frontal plane?”, but we should simply use these terms in the examinations in a regular manner routinely.

One of the most difficult steps in operationalization is the scaling of the criteria as indicators of the targeted level of quality. As an example we will use the following educational objective: “*The student will be able to perform a closed reduction of an anteriorly dislocated GHJ.*” A possible criterion would be a demonstration in front of the examiners (assessors). Thus, the assessment task would for instance be “*The student performs a closed reduction of an anteriorly dislocated GHJ in an observed setting.*” Now we have to define the levels of quality, for instance as a nominal scaled “correctly”, “partially correct”, or “incorrect”. The examiners’ rating sheet would therefore contain a statement like:

The student performs a closed reduction of an anteriorly dislocated GHJ:

- Correctly
- Partially correct
- Incorrect

We could use an ordinal scale when we want to test for higher levels of progress, i.e. “very often”, “often”, “occasionally”, “rarely”, and “never”. Transferred to our example we have to test of course our student repeatedly.

The student performs a closed reduction of an anteriorly dislocated GHJ correctly:

- Very often
- Often
- Occasionally
- Rarely
- Never

Of course we could also establish an examination setting with a defined number of repetitions and can establish a ratio scale, for instance “N-times out of 10”.

The student performs a closed reduction of an anteriorly dislocated GHJ correctly in N times out of 10 attempts.

Only when this operationalization has taken place, one can decide on the appropriate assessment instrument. As a matter of fact, exactly this process is not observed routinely. Very often, the

assessment instrument is selected before the operationalization and the operationalization is omitted.

Assessment instruments

It is obvious that the students’ learning is influenced and directed by the assessment they will have to take. The selected assessment strategy defines, to a huge amount, what students are willing to learn, or as Flexner (1910) stated, “However the teacher teaches [...] the way in which the student studies is largely influenced by the examinations” (shortened reprint as: Flexner, 2002). As for instance multiple choice questions (MCQs) foster factual knowledge within the cognitive domain, learning of educational objectives from other domains will be set aside by the students and no one will ever be able to judge whether they have achieved the necessary progress in these objectives.

Assessments can motivate students by at least four distinct factors: (1) the relevance of the assessment, (2) the content of the assessment, (3) enthusiastic teachers, and (4) group-dynamics.

As a general rule, for each domain an adequate assessment strategy will have to be used. This rule is not one without exemption; there are educational objectives from one domain, which should be assessed by the principal method of another domain (Brenner et al., 2003).

Screening the educational literature reveals a vast number of assessment instruments, some of them quite similar to other formats, some others seem quite exotic (Table 3). On the other hand, anatomists seem to be quite traditional when assessing their students using “specific written questions”, “(practical) spotter examination”, and “multiple choice question” examinations, no matter whether they teach in a traditional curriculum, a problem-based curriculum, or a system-based curriculum (Heylings, 2002). “Integrated case-based papers” and “specific written anatomical questions” are seldom used formats, as is the case with “vivas”, “worksheets”, and “self-directed projects”. The most interesting result of this study was a very limited number of complementary usages of different formats. Furthermore, assessment is mainly targeted for identification of anatomical structures, much less assessments address the applying or analysing levels.

Besides the general limited selection of assessment instruments, several other tools have been developed, for instance the “Think-Tank” (Peel, 1998), or the “three-dimensional multiple choice test” (Schubert et al., 2009).

Standard setting

Assessment tools may also be classified by a major difference as there are tools requiring the (active) production of an answer, a behaviour, or a product, e.g. written paper(s), structured oral ex-

Table 2. Intermediate educational objectives for anatomical dissection (Brenner et al., 2003).

Domain	Educational Objective	Classification & Assessment
Cognitive	Students will describe the spatial construction of the human body.	In order to reach this objective, students will have to apply both factual and conceptual knowledge. Dissection will be used as educational activity, and students will be assessed by a structured oral examination.
	Students will describe and explain the integration of an organ, organ-system, etc., into the appropriate higher system.	Students should be able to analyse their systematic factual knowledge in the context of the topographical specimen they are creating. This implies that they are also able to apply their conceptual knowledge of topography. The educational activity will be dissection, and students' achievements will be assessed by a structured oral examination.
	Students will delimit the topographical regions and cavities of the human body from each other.	This objective calls for analysis and evaluation of factual knowledge. Additionally, appropriate concepts of limitation of regions and cavities contribute to this objective. This objective will be taught best by dissection, and assessed by a structured observation.
	Students will demonstrate the ability to differentiate and exclude based on concrete examples/specimens.	This one demands for the evaluation of factual and conceptual knowledge. Dissection will be the appropriate educational activity and students will be assessed by a structured oral examination.
	Students will describe and demonstrate the inner and outer surface-relief of the human body.	Similar to the previous educational objective, this one fosters the evaluation and application of factual knowledge. Students should learn this objective by dissection and they will be assessed by a structured oral examination.
	Students will recognize and name prosected entities/structures of the human body.	This targets the evaluation of factual knowledge. Whereas it seems that this objective is imminent in anatomical dissection, it has to be stated explicitly. In order to achieve this objective, students will use dissection as proper educational activity and they will be assessed by a structured oral examination.
	Students will describe and demonstrate the layer-sequel in the construction of the regions and cavities of the human body.	The description of the layer sequel of the human body is a clear application of a conceptual knowledge. Students should achieve this educational objective by dissection, and present their achievements by a structured oral examination.
	Students will discuss systematic fundamentals, especially course and function, in the synopsis of the structures of a region or cavity.	We map this to the application of conceptual knowledge. The proper educational activity will be dissection and the appropriate assessment will be a structured oral examination.
	Students will describe and explain the interaction of organ-systems.	In order to achieve this, students will have to apply and analyse a profound conceptual knowledge of the function of organ systems. Dissection should be the proper educational activity, and assessment will be made by a structured oral examination.
	Students will demonstrate as well as describe the different regions and cavities of the human body, their content, their function, and their borders.	In contrast to the first objective, here an analysis of conceptual knowledge is demanded. Students will have to integrate concepts of systematic anatomy, topography, function, etc. The educational activity used will be dissection, and assessment will be made by a structured oral examination.
	Students will describe and demonstrate important 'spreading-streets' of infections.	Achievement of this necessitates the analysis of conceptual knowledge. The educational activity will be dissection, and achievements will be assessed by a structured oral examination.
	Students will recognize, understand, demonstrate and document the individual peculiarities and variations in the construction.	This comprises both the evaluation of the conceptual knowledge of anatomical variability, and – to some extent – a distinct communicative skill. Students should achieve this educational objective by dissection, and their achievements will be assessed by their portfolio.
	Students will learn by mutual learning and explaining each other.	Mutual learning and explaining address the meta-cognitive knowledge of learning processes. We want the students to gain the competence to evaluate these processes or knowledge, respectively. Students therefore will have to work as a group together – of course with supervision by the teacher. Assessment will be done by structured observation.

Table 2 (Cont). Intermediate educational objectives for anatomical dissection (Brenner et al., 2003)

Domain	Educational Objective	Classification & Assessment
	Students will find and demonstrate anatomical structures within a region or cavity, without damage to other structures within.	Whereas others may consider this educational objective to map into the cognitive domain, we want the students to apply their factual and conceptual knowledge in the sense of a manual or tactile skill: students should acquire control of the clinical skill to find and demonstrate anatomical structure by dissection. We will assess the students by a structured observation.
	Students will demonstrate selected medical techniques, such as several forms of local anaesthesia, coniotomy, trepanation, tooth extraction, punctures, and palpations.	The human cadaver used in anatomical dissection is a good tool for training several medical techniques. Students should at least gain control for these clinical skills. Selection of techniques has to be done carefully and only techniques should be used, which are relevant for a general practitioner. The selected techniques will be taught by initial demonstration and following exercise during dissection. The students will be assessed by a structured observation.
	Students will handle selected (surgical) instruments appropriate.	Similar to the previous objective, this objective addresses a clinical skill. Again, we want the students to gain at least control of this manual skill. Therefore, the usage will first be demonstrated and then trained by application in the process of dissection. This skill will be assessed by structured observation.
	Students will acquire tactile skills.	It targets a professional competence and maps clearly to the psychomotor domain. Any enhancement of the students' tactile skills can address the reproach that nowadays medicine uses laboratories and machines too extensively. We want the students to acquire real automatism in this manual (and also perceptive) skill. The proper educational activity should be dissection and assessment will be made by structured observation.
	Students will handle specimens and instruments carefully.	Besides the necessary attitudes, we focus on the manual skill of handling instruments and specimens. It should be learned by dissection and will be assessed by structured observation.
	Students will observe, feel and describe.	The main target of this educational objective is the clinical skill of perception. We want the students to achieve a high level of automatism in this important professional competence. It goes beyond the mere tactile skill addressed previously. It will be taught by dissection, and assessed by a structured observation and the students' portfolio.
Psychomotor	Students will describe and demonstrate the spatial relationships of anatomical structures to each other within a certain region or cavity, but also beyond their borders.	Whereas the previous one requires a manual skill, this educational objective targets the control of a – verbal and non-verbal – communicative skill. The appropriate education activity is dissection, and students will be assessed by a structured observation.
	Students will implement an appropriate documentation of their own activities.	This utilizes meta-cognitive knowledge, but mainly targets the psycho-social – communicative – skill to document all activities of the students themselves. Therefore, the students will have to manage a portfolio as a product of group-work, which will be finally assessed.
	Students will manage and implement an exact documentation.	This one is quite similar to the previous objective. The major difference is the content to be documented: whereas in the former the content of documentation is the students' own activities, this objective targets all other entities, such as special findings, external influences, etc., which can contribute to a full imagination of the cadaver dissected. As a patient's file is compiled by different physicians, etc., students are intended to set up and manage a portfolio as a group.
	Students will incorporate themselves into a team by appropriate work, especially by transmission of knowledge, cooperate, lead and solve conflicts.	This is a quite difficult one to classify as it comprises factual, conceptual, and meta-cognitive knowledge of group-dynamics and knowledge-transfer, as well as the psycho-social skills to apply these entities of knowledge. Seen from another point of view, it addresses several entities of professionalism such as the commitment to professional competence and responsibilities. As for this paper, we classify this educational objective to the domain of professionalism. Students should learn by dissection and will be assessed by structured observation.
	Students will recognize and document deviations from the healthy, 'normal' condition.	This educational objective is meant to be mainly a psychomotor skill, which uses of course conceptual knowledge. The educational activity will be dissection, and assessment is accomplished by the students' portfolio.

Table 2 (Cont). Intermediate educational objectives for anatomical dissection (Brenner et al., 2003)

Domain	Educational Objective	Classification & Assessment
Psychomotor	Students will express them precisely and clear orally as well as in writing by practical application of the medical and anatomical terminology in clinical and scientific context.	This communicative – psycho-social – skill is intended to the level of automatism as we consider it to be a crucial skill in the medical profession. The main educational activity will be dissection, accompanied by the students' group-work. This skill will be assessed by a structured oral examination for the oral part and by the portfolio for the written part.
	Students will respect the dignity of the dead person.	This objective addresses an attitude, which the students should internalize. Achievements should be gained both by dissection and group-work. This group-work should be the planning and implementation of any sort of commemoration. We will leave it to the students' decisions, which form of commemoration they will chose, a – religious – service, a recital, or whatever they want. Assessment will be made by structured observation, and the students' are asked to add the materials used for preparation and implementation, e.g. letters, programs, etc., to their portfolio.
	Students will demonstrate an appropriate commitment with death.	The demonstration of a commitment is behaviour from the affective domain. We target on the concept of death and dying, which we consider to be an important professional competence. By dissection we want the students at least to gain response to this educational objective. Assessment will be made by structured observation.
Affective	Students will show, keep and justify behaviours in the contact with the deceased.	Whereas the required behaviour from the previous objective targets death in general, this one focuses mainly on the behaviours with and towards the specific cadaver, with which the students work actually. The educational activity will be dissection. Additionally, students should compile in group-work a set of behavioural rules to be applied in the lab. Assessment will be accomplished both by structured observation and the students' portfolio.
	Students will demonstrate respect and honesty towards academic and non-academic personnel.	This can be seen either as attitude or as behaviour. In contrast to the next educational objective, which focuses on the inter-personal relations more generally, this objective addressed the reproach that "doctors" behave condescending against other health professionals. As the curriculum is output-oriented, we consider it as a behaviour, which – of course – must be taught mainly by demonstration and role-modelling and we will assess the achievements by structured observation.
	Students will obey and justify behaviours in the contact with patients, colleagues, etc., as for instance with respect, tolerance, cleanness, and (self-) discipline.	We want the students to internalize these behaviours, which are also crucial for medical professionalism. There will be no special education activity – beside dissection – and assessment is done by structured observation as well as the portfolio.

aminations, structured observations, viva voce, etc., whereas others require (just) the selection from given answers (e.g. true-false, MCQ, EMQ, etc.). These two forms may be combined. This means that in some circumstances, both the practical & theoretical steps of an examination might be first discursive answers, and secondly structured assessment tools, or vice versa.

This distinction is important in terms of standard setting. Standard setting is most important for gaining reliability. In other terms, all assessors must agree on a common, consolidated set of rating and scores for the answers, behaviours, or properties of the students' products. It is in effect demarking the cut off between acceptable and not acceptable (Norcini, 2003).

In anatomy standard setting is frequently employed to define the pass mark of anatomical knowledge. However, it is not suitable for associat-

ed skills and attitudes. As described by Cohen-Schotanus and van der Vleuten (2010) in a detailed analysis of the examination results in medical schools in the Netherlands, there is no gold standard in standard setting. In anatomy, criterion-based methods have tended to be preferred. Both the Angoff (1971) and Ebel (1972) methods are commonly used in medical education (Bandaranayake, 2008), although the suitability of such a specialized panel of anatomists for Angoff has not proved to be as reliable as a modified Ebel (Cohen-Schotanus and van der Vleuten, 2010). A further problem with both these methods is the gathering together of a suitably large cohort of people sufficiently well informed about how anatomy fits into the overall course structure and aspirations. Anecdotally (and there does not appear to be any published information on this) the variance among the expert assessors is sufficiently great to

Table 3. (incomplete) survey of different assessment instruments used in medical education

Paper-based	Patient-based	Teacher-/Peer-based
	<ul style="list-style-type: none"> • 360° Evaluation Instrument • Application Test • Appraisal Technique • Assessment of Performance <ul style="list-style-type: none"> * Checklist Evaluation * Global Rating * Assessment Panels 	<ul style="list-style-type: none"> • Peer Evaluation: <ul style="list-style-type: none"> * Medical Team * Multi-disciplinary • Practical Examination • Rating Scales • Recall (Oral) Examination: <ul style="list-style-type: none"> * Chart-Stimulated * Video-Stimulated
<ul style="list-style-type: none"> • (Modified) Essay Questions Test • Multiple Choice/Answers Questions Test • Tag/Spotter Test • Thesis and Project • Publications/Conference Papers • Progress Test • Patient Surveys • Short Answer Questions 	<ul style="list-style-type: none"> • Direct/Practice (structured) Observation • Direct observation of procedural skills (DOPS) • Key Feature Problems / Examinations • In-Basket Exercise • Objective Structured Clinical Examination • Objective Structured Long Examination Record • Simulations and Models • (Standardized) Oral Examination: <ul style="list-style-type: none"> * Patient based long case * Patient based short case • Simulation based Assessment: <ul style="list-style-type: none"> * Simulated Patient Examination * Standardized Patient Examination * Simulated Surgeries • Patient Management Problem 	<ul style="list-style-type: none"> • (medical) Record (case note) Review • Record of Practice: <ul style="list-style-type: none"> * Critical Incidence Technique * Portfolios * Procedure, Operative, or Case Logs • Structured Interviews • Structured Trainer's Report • Triple Jump Exercise • Video Assessment • Vivas: <ul style="list-style-type: none"> * Unstructured (Question firing) * Structured Oral Examination • Clinical evaluation exercise (mini-CEX)

render the whole process of questionable value. Data gathered at the Winter Meeting of the Anatomical Society in 2011 at Cardiff, in a workshop conducted on anatomy assessment with 71 anatomists participating, when asked about their use of standard setting it was revealed that the predominant method used by anatomists was Angoff (38%) followed by Hofstee (15%) and the Fixed Method Percentage (12%), though the new Cohen method (Cohen-Schotanus and van der Vleuten, 2010), which uses the performance of a top group of students as a measure of the relative difficulty of the assessment, was used in Dutch schools and in one UK school.

This is no issue at all in assessment tools, where the students are required to select the appropriate answer(s) from a given set of choices, as in true-

false questions, MCQ, EMQ, etc. The standard setting is done in these cases by the selection of the "true" answer(s). Problems arise, when there is not a single best answer; and even more, when the second best answer is also true and was not intended as "true" answer by the question's author, but found post-hoc as "true" answer by the examinees.

Nevertheless, these are minor problems in comparison to the problems arising from unstandardized assessment items, where the examinees are required to produce an answer, behaviour, a product, or something else actively. For comparison, a multiple answer question and an identical question for a (structured) oral examination is given:

Therefrom several problems arise, which need

	MAQ	Oral examination
Question	Which of the following optional structures can be found in the glenohumeral joint? (a) Articular disc (b) Articular recess (c) Capsular ligaments (d) Labrum (e) Meniscus	Which optional structures can be found in the glenohumeral joint?
Expected Answer	Which of the following optional structures can be found in the glenohumeral joint? <input type="checkbox"/> (a) Articular disc <input checked="" type="checkbox"/> (b) Articular recess <input checked="" type="checkbox"/> (c) Capsular ligaments <input checked="" type="checkbox"/> (d) Labrum <input type="checkbox"/> (e) Meniscus	The optional structures of the glenohumeral joint are an articular recess, capsular ligaments, and a labrum.

for consent:

- In the oral exam question, **which answers are allowed for being true?** In this case, the student might not respond “articular recess”, but might use the term “axillary recess” or even “recessus axillaris”. The “capsular ligaments” might be replaced by “glenohumeral ligaments”, and even the labrum might be responded as “glenoid labrum”. Obviously all these answers are true for the GHJ. All assessors will easily agree that this set of expected answers [articular recess|axillary recess|recessus axillaris], [capsular ligaments|glenohumeral ligaments|Ligg. glenohumeralia], and [labrum|glenoid labrum|labrum glenoidale] will be rated as “correct”. In other cases, it might be more difficult to get consent on a standardized set of expected answers.
- How are true and false answers rated?** This question has to be answered for both the MAQ/MCQ and the oral question. When MCQs/MAQs are administered in a generalized setting, these issues are predefined by the administration. For the written version with predefined answers, there is general agreement that each answer contributes to the total score. Thus an un-marked “true” answer is equally wrong than a marked “false” answer, whereas a marked “true” answer is equally right than an un-marked “false” answer. In oral exams, this gets more complicated, because of the fact that all expected answers are only “true” answers. Thus, consent has to be achieved how unspoken missing “true” answers should be handled and rated. Additionally, consent has to be achieved whether the assessor is allowed for inquiring for a missing answer (“Don’t you think that there is something missing?”) or not.
- How should the whole question be rated?** In the written setting of MAQs/MCQs exams, this issue is often predefined by the admin-

istration, or the software for analyse the scoring sheets. There is general agreement that one question can be rated either “correct” or “incorrect” as a whole. The question arising here is, if “incomplete” answers, i.e. a mixture of answers rated “true” and “false”, might be rated “partially correct”. With this “partially correct” scoring, there is a logical problem: (a) when a “true” answer is marked, it is logically “true”, (b) when a “true” answer is not marked, it is logically “false”, (c) when a “false” answer is not marked, it is logically “true”, and (d) when a “false” answer is marked, it is logically “false”. Logically, “true” and “false” is “false”. Thus, if even one answer is marked wrongly, the whole conjunction or construct is “false”. In MCQs, this is quite simple: the question is “true” only, when only the expected “true” answer is marked, and not more (as these are obviously “false”). In MAQs, it is also simple from a logical point of view: the question is “true” only, when all expected “true” answers are marked, neither more nor less. Nevertheless, there are several assessors, who are willing to give partial ratings for incomplete answers, with neglecting that each missing “true” answer is also “false”. This might be administered based on a general consent of all assessors involved.

- In oral examinations, exactly this issue is heavily discussed between assessors and scrutinized by the examinees. Having no common consent on this question among the assessors bates the assessments reliability. The logically recommendation would be: (a) “when the student responds incompletely (with or without inquiring), i.e. doesn’t reproduce all expected ‘true’ answers, the whole question is rated ‘false’”, and (b) “when a student includes a ‘false’ answer to her or his response, the whole question is rated ‘false’”.
- How should the rated questions be scored?** There is a difference between rating and scoring, as rating means that the assessor rates the answer either “true” or “false”. By this, he gets an answer whether question itself, the logic conjunction, is answered “true” or “false”. Finally, this rating has to be transformed to scores.
In MCQ/MAQ-examinations, in general there is such a high number of questions that all questions are scored identically with exactly one score (or point). Of course, there is the possibility of weighting, i.e. a subset of questions is considered more important than other ones, and is therefore awarded with higher scores. In this case, of course no partial scores should be assigned. In oral examinations, this step is not always done, especially when the whole assessment’s result has to be transformed to grades (depending on the

grading system used; A-E, 1-5, 1-6, etc.). Nevertheless, it is as simple as with MCQs/MAQs, and each question is assigned a score. All questions together result in an overall examination-score, which can afterwards be transferred or recalculated to grades. There is just one limitation: the whole exam must at least contain as many questions as grades ...

Standardisation for oral examinations is therefore quite easy, when all steps are made in accordance and in common agreement of all involved assessors. Of course, even more sophisticated and structured forms of standardized oral examinations might be developed (see below).

Written exams

Even in written exams, questions might match different levels of the progress dimension. A good description of different types of questions is given by Chollet et al. (2009):

“Four question types were possible: memorization, pathway, spatial, and analytical (Question Type). Memorization questions were those questions that required a student to recite a fact (e.g., True/False: CN I is responsible for olfaction). Pathway questions required students to trace the path of blood or a nerve (e.g., trace the path of blood from the appendix back to the heart). Spatial questions required the student either to describe the spatial relationship of one structure to another (e.g., True/False: The subclavian vein is anterior to the anterior scalene muscle) or to identify the indicated structure in a diagram. A question was classified as analytical if it required the student to solve a problem that required significant integration of material. Analytical questions were often presented as clinical scenarios (e.g., A patient presents with shoulder pain. Explain how this might be related to inflammation of the gallbladder).”

Multiple choice questions (MCQ)/ multiple answer questions (MAQ)

Literature on multiple choice questions is legion. Thus we will strictly present a short overview of the structure.

An optimal MCQ starts with a **stem**. This stem introduces the candidates with the topic of the question, and gives all, but not more than, the information necessary. This stem can also be an image, or a clinical vignette. Images could be schematic drawings, photographs, images from different imaging modalities such as ultrasounds, x-ray/ plane radiographs, CTs, MRIs, or diagrams, for instance an electrocardiogram. The clearer the information given is, the less misinterpretations of the proper question will occur.

The stem is followed by a genuine **question** (with a question mark!), or a distinct **request**. The question or request has to make unmistakable clear, what the candidate should do with the infor-

mation given in the stem supported by his own knowledge. In the case of the MCQ, the candidate will have to make the decision to mark the appropriate answer.

The question is followed by a set of possible **answers**. In the case of MCQs there is just one “correct” choice, the “single best answer”, the others – in general four – “wrong” choices are distractors. In the case of multiple answer questions (MAQs), there are more than one “correct” choice; on the other hand, MAQs should at least contain one “wrong” choice.

The construction of the “wrong choices”, the distractors, is very important for the quality of the question. The term “possible” is the first and most important feature of an answer or distractor; this very choice has to be a *possible* answer to the question asked. Statements such as “none of the above” are inappropriate choices, as they aren’t answers to the question asked. Furthermore, the answers should be *similar* in terms of provenance, for instance the same system, topographical region, or they are all clinical treatment options, diagnoses, etc. All choices, the “correct” and the “wrong” ones, must be of *similar length*. Questions writers tend to elaborate the “correct” choice while disregarding the distractors, resulting in the fact that the longest answer is very often the correct one. Another cues for wrong answers are absolute statements (“always”, “never”, etc.), therefore such terms should be avoided. The authors should be aware that even grammar or the gender of terms might be cues. In general:

1. Avoid interpretation problems (semi-quantitative terminology, mingled cause and effect, ambiguous terminology)
2. Test facts, avoid testing opinions
3. Test only one aspect per question
4. Avoid hints and cues
5. Use short and clear sentences
6. Use semantically un-ambiguous terms
7. Avoid negations
8. avoid nonsense alternatives (fillers)
9. avoid two-in-one alternatives
10. make sure alternatives are mutually exclusive

Now some examples derived from the GHJ.

Synovial joints may have different optional structures.

Which of the following optional structures can be found in the glenohumeral joint?

- a. Articular disc
- b. Articular recess
- c. Capsular ligaments
- d. Labrum
- e. Meniscus

The expected correct answers for this MAQ would be (b), (c), and (d). The choices (a) and (e) are in fact also optional structures of a synovial joint, but can’t be found in the glenohumeral joint. Of course this question asks for mere rote memorization.

Mr. Doe shows up in your office. He complains

that he feels some pain in his left shoulder. The onset was some days ago when he had to work in his garden and cut off some branches of a tree.

When you examine Mr. Doe's left shoulder, you find that he is unable to abduct the left arm fully. The measured range of movement (add/0/abd) is 30°-0°-65°.

Which of the following muscles of the glenohumeral joint might be damaged?

- infraspinatus muscle
- subscapularis muscle
- suprascapularis muscle
- teres major muscle
- teres minor muscle

Extended matching questions (EMQ)

Extended matching questions are a set of questions (two or more), which use a common list of answers, which is not restricted in length. This format is now used on many licensing and certification examinations, as well as intramural tests at many medical schools. In contrast to other question types, this list of common answers precedes the individual questions. It is indispensable that all these common answers are completely unique and free of any cueing. Their sequence starts with a general introduction or theme, followed by the question or lead-in (mostly formulated as request) that gives the students instructions on what to do. These parts are followed by the list of answers. The response option includes one correct answer for each question, and other possible responses as distractors. It is the main characteristics of EMQs that this list of possible answers can pertain to any number of independent test items or questions. There is no restriction on the number of times a given answer may be correct. Several stems, in most cases more or less complete clinical vignettes, complete an EMQ. Therefore they are quite difficult to construct; furthermore, there is only few examination software capable to use this question format.

The following patients all suffer from muscular weakness and sometimes from pain. Please choose the most likely affected nerve from the list below:

- suprascapular nerve
- axillary nerve
- radial nerve
- ulnar nerve
- median nerve

A 54 year old patient complains about difficulties to turn a key. He can't remember an accident but reports an extensive tennis match.

A 42 year old woman complains that she has difficulties to dress her hairs.

An older patient complains that he increasingly drops his pint.

EMQs shall overcome some of the criticisms levelled at the use of multiple-choice questions to test factual and applied knowledge. Among the ad-

vantages to using EMQs:

- Themes aid the organization of the examination.
- As questions are written in themes or general topics it allows for writing many questions for that theme.
- Good questions provide a structure designed to assess application of knowledge rather than purely recall of isolated facts (which is not only true for EMQs!).
- The approach to writing these questions is systematic, which is very important when several people are contributing questions to one exam.
- The extended list of answers allows the inclusion of all relevant options, and therefore reduces the opportunity for students to 'guess' the correct answer.
- EMQs were found to be more discriminating than two and five option versions of the same questions resulting in a greater spread of scores, consequently resulting in higher reliability.

Medical students previously familiar with the one-from-five type of question (MCQ) do relatively poorly the first time they are faced with the new-style EMQs, probably because at least some of them are still working in the "eliminate the distractors" mode. On the assumption that students will adopt whatever mode of working gives them the best chance of passing the examination, EMQs that require the students to use and evaluate data with the aim of making a deduction, i.e. using (clinical) vignettes, then this is what they will do.

Spotter/tag tests

In a tag or spotter test, the students have to pass different stations with pre-dissected/prosected specimens, where several structures are "tagged", and to work on relating tasks. The questions concerning to the stations and their tagged structure have to be answered on an exam-sheet or its electronic equivalence. There are used SAQs (short-answer questions) usually, occasionally MCQs are used. Very often the tag or spotter test is used simply to assess the identification of the tagged structures, thus being an assessment of remembering factual knowledge (Winkelmann, 2007). Besides the identification, sometimes a further question is added to the targeted structure. This may be derived from the system, topography, function, development, clinical reference, etc. (Adamczyk et al., 2007). Besides the mainly simple form of a spotter, more sophisticated versions have been developed (Schubert et al., 2009). Justifications for using spotter style assessments include that they:

- are aligned to three dimensional spatial understanding

- can test relations of structures to each other
- can test ability to identify similar structures e.g. nerves
- are aligned to learning through touch
- can be related to clinical applications
- can test an appreciation for anatomical variation
- may be linked to imaging
- can require production of knowledge

When seen in its simplest form, the spotter/tag test is just a modified written exam with either MCQs or SAQs. The tagged specimens serve as initial trigger or stem for the corresponding questions. The real specimen can be replaced by its image. One could imagine that two tests with literally the identical questions, one presenting the specimens as stems and the other presenting the specimens' images as stems, will achieve identical results; when not, the assessment does not assess what it is intended to measure. Any differences in results might be biased by the setting itself.

Spotter questions can also be included as a station in an Observed Structured Clinical Examination (OSCE), or integrated into a wider practical assessment including other disciplines. As well as OSCEs, Objective Structured Practical Examinations (OSPEs) have been proposed as an effective modality for assessing the practical aspects of anatomy (Yaqinuddin, 2013). It is suspected that different models of this have emerged under alternative names as the next generation of 'spotters'.

The spotter has been criticized for testing low levels of knowledge (Yaqinuddin et al., 2013). Although this may be true for a spot such as 'Identify this structure', where testing is at the level of simple recall, modifications could move such a question into a higher taxonomy order. Of particular interest is that key action words associated with anatomy are used in spotter style assessments, and that these can be aligned to the taxonomy orders. The word 'Identify' is frequently associated with the *Remembering* level. 'Describe the function' is associated with *Comprehension*, whereas phrases such as 'Damage to X causes...' are associated with *Applying* and 'What clinical condition is associated with...?' with *Analysing*. There are few examples of 'create' (synthesize) but as this involves bringing components together, it may be more suited to clinical scenarios and OSCEs; further work will elucidate if anatomy and clinical skills are aligned and truly integrated. Anyhow, it is often simple to increase or decrease a level as appropriate.

Oral exams

Structured oral examination (SOE)

A structured oral examination (SOE) is a special form of a Viva (JCEM, 1997). SOEs are individual assessments. Their main targets are educational objectives from the cognitive domain, except ob-

jectives assigned to the meta-cognitive knowledge (e.g. "learn by mutual learning and explaining each other"). A structured oral examination should be structured by both the number of questions and the rating scale for each item. Of course, for each question the appropriate steps of standard setting must be passed through.

An example of the first author's institution should enlighten this. In the topographical dissection course, students have to take – among other items contributing to the final score – five structured oral examinations (SOEs). Each of these SOEs comprises three items. It is proposed that one item should deal with the recent major topic (e.g. intestine, heart, etc.), another item with the examinee's specimen; whereas the topic of the third item will be open to the assessor's decision. Each item itself consists of three questions. The first question of each item is either to name a shown structure or show a named structure. This initial question is considered mandatory, such a student not able to show or name the structure under question, has failed this whole item. When the initial question is answered correctly, the first score is assigned and followed by two consecutive questions, one of them dealing with the anatomical context (vascularisation, innervation, topographical relations to other structures, etc.) and the other dealing with either anatomical or clinical contexts. Each question is assigned, if answered appropriately, one additional score. Thus, each item can be awarded with up to three scores, and the whole SOE sums up to a maximum of nine scores, with a pass/fail level of six scores. This seems quite complex, but the assessors are supported by a convenient rating sheet:

Table 4.

Assessment X	Date:				Assessor:					
Item 1:	0	1	2	3						
Item 2:	0	1	2	3						
Item 3:	0	1	2	3						
Total Score:	0	1	2	3	4	5	6	7	8	9

All five SOEs will result in a maximum of 45 scores (and a minimum of 30 scores); with a total maximum of 72 scores for the whole course. The other scores can be achieved by the structured observation of the student's collaboration (max. 6 scores), the structured observation of the dissected specimen (max. 15 scores), and the portfolio (max. 6 scores).

These proposed SOEs should not be mixed up with a standardized oral examination, which is defined as a type of performance testing using real or at least very realistic patient cases with a trained physician examiner questioning the examinee (ACGME and ABMS, 2000). The SOEs were selected for their advantage of checking that a stu-

dent's verbal and mental ability matches the ability demonstrated in written form (JCEM, 1997), such as in the summative and integrative assessment at the end of the second term, but also in the portfolio. As these students sat for a written exam (MCQs) on systematic anatomy already, we chose this form of an oral examination, which also allows for following up a student's answers and assessing the depth of knowledge as well as assessing attitudes (JCEM, 1997). On the basis of the selection of the qualification-profile ('The student is able to communicate precisely and clearly in clinical and scientific context orally as well as written'), and under consideration of the facts that the 'big examinations' are written assessments and the written expressiveness should be examined in the dissection lab by the portfolio, at least one oral assessment must be implemented. The students must be able to express themselves clearly also in 'stress-situations'.

The reliability of this assessment strategy should be quite good as the initial questions of each item are standardized, the answers are likely to be within a specified range and the number of questions in total is quite high for an oral examination form.

The validity of SOEs hinges upon the precise level of 'structure' used. The more structured and more closed the questions, the less will be the opportunity to assess the student's attitudes and knowledge as a coherent whole (JCEM, 1997).

Online assessments

An example is the Oxford "live" experiment: a qualification intensive course 8 hours a day for 3 weeks of Principles of Clinical Anatomy, each week ended with an evaluation test. The online-assessment of the students' progress led to the following conclusions (Morris and Chirculescu, 2007):

1. Students have repeatedly commented that, by the end of the course, they felt that they had acquired the anatomical knowledge necessary to examine patients with confidence.
2. In each year, 4-6 students have scored >90% in each of the three weeks.
3. With on line assessment, inter- and intra-examiner variability is completely eliminated.
4. Marks for the cohort of students and also for each individual part of each question are available within minutes of the end of the examination. This means that 'rogue' questions can be immediately spotted to facilitate decisions on re-sits/vivas.
5. More importantly, the relative difficulty of each part of each question is quantified and a bank of questions of known difficulty will enable the creation of more robust assessments in the future.
6. Student feedback was very positive in all the 3 years passed (2004-2006), despite the fact that the course comes just one week after

their final preclinical Medical Science examination.

7. Having seen the results, external examiners have requested that on-line assessment is used for all preclinical core examination.
8. Always, the best score was related to the 1st week, and the poorest to the 2nd week examination test.

Practical exams

By exposing the spotter / tag tests as a special form of a written exam, where the tagged specimen forms the stem of the question, there is a shortcoming of real practical examinations, the more when we are searching for a structured version. Searching the literature for "anatomy" and "practical exam" reveals only few usable descriptions of practical examinations. In most cases, these "practical exams" were mere spotter / tag tests in some quite interesting modifications (eg. Cundiff et al., 2001).

We can only suppose that this is due to the fact that no appropriate educational objectives have been formulated. When the focus is only on knowledge, the worst on a low step on the process dimension (mere memorization), there is indeed no need for a practical examination.

A practical examination is an examination, in which the examinee is asked to do, to perform a certain task, especially manually. When we, for instance, want our students to comprehend the composition and structure of a certain bone, then a task for a practical examination could be to re-assemble a scattered bone. When you ask, if this clinically relevant, ask your traumatologist ... Anatomy could provide a "sheltered workshop" in advance of clinical reality. When we want our students to apply their knowledge of the range of motion of the GHJ, the task should be an ROM-test of a subject, either a co-student or a simulated patient (or even a real patient). Of course, such a task should be accompanied by appropriate verbal analysis of normal and/or pathological findings.

For dissection, the most obvious task for a practical examination would be the request to dissect a certain structure within a certain time.

Structured observation

Structured observations are individual assessments and can observe (1) the student's active contributions (e.g. Heyns, 2007), (2) their work's product, the dissected anatomical specimen, (3) selected medical techniques, or for instance the anatomical basics of selected clinical skills, on a predefined and easy rating scale. The reliability of structured observations depends on the clear definition of the rating scales used. A high number of single observation-events will contribute to a high reliability. The validity of this instrument in total should be acceptable as both process and outcome (product) are assessed.

In the first author's setting, structured observations are made on the student's active collaboration and their specimen. In general, the student's active collaboration is continuously observed by the staff. The structured observation is recorded once a fortnight in the student's file with a score of 0 or 1 for each observation.

Even the student's product in the dissection lab, the dissected specimen, can be monitored for assessment purposes. For this structured observation, both the dissection-technique of the student, and – essentially – the product, the specimen itself, should be evaluated. Of course, the respective difficulty of the region dissected, but also the quality of the specimen has to be accounted for.

A rating scale could for instance comprise of four levels: 0–1–2–3, with 0 being "insufficient", 1 "satisfactory", 2 "good", and 3 "outstanding". Of course, these levels have to be defined more clearly by common consent of all assessors. "Insufficient" dissection can be defined as a specimen, where the student has not at all exposed all those major structures (such as the main nerves of a plexus, a major vessel, etc.), which should have been dissected according to the dissection plan, or has cut apart one or more of these major structures. The specimen might be "satisfactory", when the major structures are exposed in general, but incomplete in respect of the target region itself or adjacent regions. A "good" specimen can be defined by a complete exposure of all major structures and most of the minor structures within the target region and appropriate connection with other adjacent regions. Finally, an "outstanding" specimen should go (far) beyond the common requirements for a student's dissection, such as a complete exposure of all major and minor structures within the target region and appropriate connection with other adjacent regions, combined with a complete removal of connective tissue where appropriate; the specimens should look like an image of a photographic atlas. Of course, there has to be a list of major and minor structures ...

Portfolio

The portfolio could be either a single student's product or a group assessment where all students working on one cadaver will have to contribute. The portfolio documents are distinct products of the students' work. It can be used for direct presentation of performance. The portfolio can assess the students' skills in team-work and documentation, the usage of different sources of information (old and new media), ethical aspects, as well as self-assessment (reflection).

Based on the selection of different items of the general educational objectives – under consideration of the respective parts and key-requirements – we recommend that all students working on one cadaver should produce and manage a common portfolio as group.

The portfolio documents distinct products of the students' work. It can be used for direct presentation of performance. The portfolio combines technical as well as general achievements, for example excerpts from books, learning-diaries, other elements chosen by the students and/or negotiated with the teacher, project-reports, case studies, etc.

The portfolio for the dissection lab should have to comprise the following obligatory items:

- An initial and ongoing description of the corpse, with an exact description of the findings without diagnosis, and without influence by 'norms'. This will train the students' skills in observation, examination and documentation. This item is quite similar to the "clinical anatomy chart" (Escobar-Poni and Poni, 2006).
- A documentation of the dissection itself (who dissected what when with which result; "progress notes" as proposed by Escobar-Poni and Poni, 2006);
- A documentation of the implementation of selected techniques (clinical skills, such as spinal puncture, etc.);
- A documentation of the usage of instruments;
- A documentation of the individual peculiarities (norm-variations, varieties, abnormalities [either congenital or acquired] etc.), and eventual pathological changes (tumours, fractures, etc.) of the corpse, supplemented by findings from 'new' and 'old' information-technologies (part of the "clinical anatomy chart" proposed by Escobar-Poni and Poni, 2006); and finally
- An ongoing self-evaluation of the student's own performance (in the sense of a reflection). The students can include not only their own behaviour in a specific circumstance, but can also write about the behaviours of others (Escobar-Poni and Poni, 2006).
- Optionally, the students might insert further documents, for instance the documentation of their commemoration service. Optional supplements will of course be honoured with additional scores.

The appraisal takes place for the entire group together. Each participant of the group will get the same score. The main focus of the evaluation is set towards tidiness, wholeness, and correctness regarding content and terminology. The good news is that it might stimulate mid-level and border-line students to improve their learning performances. The bad news is that it can either encourage some students to stay apart and passively benefit of the others' work or make the performant and ambitious students, if indulgent - to work harder or if exigent - to exclude some members of the team from doing, not to spoil nor to decrease the level of the group's work, in order to avoid decrease of the team performance and global evaluation.

The reliability of this instrument in general is quite low, since the range and individuality of dif-

ferent portfolios will decrease the level of reliability between different assessors. Therefore, this instrument should be limited to about 10% of the maximum score. Nevertheless, such a portfolio is high on face and content validity.

Students' Self-directed Learning and Assessment

'Doughnut Rounds'

'Doughnut Rounds', are self-directed learning approaches, which involve four to seven medical students and/or residents. They choose the reading material as a group and are expected to formulate 12 questions based on the week's readings. They pose these questions to their colleagues in a 'game show' format. In this scenario, the attending teacher/lecturer has only to bring the doughnuts and to act as a moderator during the session. (Fleischer et al., 1997)

Donut rounds are at least as good as lectures in imparting factual knowledge and may provide a selective advantage to weaker students (Bulstrode et al., 2003).

CONCLUSION

There is much diversity concerning the assessment of anatomical knowledge, skills, and attitudes in medical and dental courses and there needs to be agreement about the best strategies and methodologies to be pursued to ensure consistency, reliability, (clinical) validity, and standardisation across the European medical and dental sectors. Therefore, the development of a European-wide clinical- and research-oriented integrated curriculum of Anatomical Sciences (Gross Anatomy, Histology and Embryology) for medical students is a major need, which should help also to achieve a unitary evaluation system.

ACKNOWLEDGEMENTS

We gratefully thank Tracey Wilkinson, Jo Bishop and John Morris for their valuable input on standard-setting.

REFERENCES

- ACGME AND ABMS (2000) *Toolbox of Assessment Methods. A product of the joint initiative of the ACGME Outcome Project of the Accreditation Council for Graduate Medical Education (ACGME), and the American Board of Medical Specialties (ABMS)*. <http://www.acgme.org/Outcome/assess/Toolbox.pdf>. accessed 11-06-2003.
- ADAMCZYK C, HUENGES B, MÜLLER-GERBL M, PUTZ R (2007) Das Fähnchentest als neue Prüfungsform im Fach Anatomie an der Ludwig-Maximilians-Universität München. *GMS Zeitschrift für Medizinische Ausbildung*, 24(3): Doc152.
- ANDERSON LW, KRATHWOHL DR, AIRASIAN PW, BLOOM

- BS (2001) *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. Longman, New York, Toronto.
- ANGOFF WH (1971) *Scales, norms, and equivalent scores*. In: THORNDIKE RL (ed). *Educational measurement: Theories and applications*. American Council on Education, Washington, DC, pp 508-600.
- BANDARANAYAKE RC (2008) Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. *Med Teach*, 30(9-10): 836-845.
- BIGGS J (1999) What the Student Does: teaching for enhanced learning. *High Educ Res Dev*, 18(1): 57-75.
- BIGGS J, TANG C (2011) *Teaching for quality learning at university*. Open University Press
- BLOOM BS, ENGLEHART MD, FURST EJ, HILL WH, KRATHWOHL DR (1956) *Taxonomy of Educational Objectives: Handbook I, Cognitive Domain*. David McKay, New York.
- BRENNER E, MAURER H, MORIGGL B, POMAROLI A (2003) Classification of intermediate educational objectives of the dissection lab according to the domains in medical education. *Ann Anat*, 185(Suppl. 98): 229.
- BULSTRODE C, GALLAGHER F, PILLING E, FURNISS D, PROCTOR R (2003) A randomised controlled trial comparing two methods of teaching medical students trauma and orthopaedics: traditional lectures versus the "donut round". *Surgeon*, 1(2): 76-80.
- CHIRCULESCU A, CHIRCULESCU M, MORRIS J (2007) Anatomical teaching for medical students from the perspective of European Union enlargement. *Eur J Anat*, 11(S1): 63-65.
- CHOLLET MB, TEAFORD MF, GAROFALO EM, DELEON VB (2009) Student laboratory presentations as a learning tool in anatomy education. *Anat Sci Educ*, 2(6): 260-264.
- COHEN-SCHOTANUS J, VAN DER VLEUTEN CP (2010) A standard setting method with the best performing students as point of reference: Practical and affordable. *Med Teach*, 32(2): 154-160.
- CUNDIFF GW, WEIDNER AC, VISCO AG (2001) Effectiveness of laparoscopic cadaveric dissection in enhancing resident comprehension of pelvic anatomy. *J Am Coll Surg*, 192(4): 492-497.
- EBEL RL (1972) *Essentials of educational measurement*. Prentice Hall, Oxford.
- ELLIS H (2002) Medico-legal Litigation and its Links with Surgical Anatomy. *Surgery*, 20(8): i-ii.
- ESCOBAR-PONI B, PONI ES (2006) The role of gross anatomy in promoting professionalism: a neglected opportunity! *Clin Anat*, 19(5): 461-467.
- EVANS DJ (2007) The role of the anatomist in communicating Anatomy to a lay audience. *Eur J Anat*, 11(S1): 79-83.
- FERNANDEZ R, DROR IE, SMITH C (2011) Spatial abilities of expert clinical anatomists: comparison of abilities between novices, intermediates, and experts in anatomy. *Anat Sci Educ*, 4(1): 1-8.
- FLEISZER D, FLEISZER T, RUSSELL R (1997) Doughnut rounds: A self-directed learning approach to teaching

- critical care in surgery. *Med Teach*, 19(3): 190-193.
- FLEXNER A (1910) *Medical education in the United States and Canada. A report to the Carnegie Foundation for the Advancement of Teaching*. Carnegie Foundation for the Advancement of Teaching, New York.
- FLEXNER A (2002) Medical education in the United States and Canada. *Bull World Health Organ*, 80(7): 594-602.
- GARG AX, NORMAN G, SPEROTABLE L (2001) How medical students learn spatial anatomy. *Lancet*, 357(9253): 363-364.
- GIBBS G, HABESHAW T (1989) *Preparing to teach. An Introduction to effective teaching in higher education*. Technical & Educational Services Ltd, Bristol.
- Guilbert J-J (1998) *Classification of professional tasks into three domains: practical, communication and intellectual skills*. In: GUILBERT J-J (ed). *Educational Handbook for Health Personnel*. World Health Organisation, Geneva, 1.50-51.54.
- HARROW AJ (1972) *A Taxonomy of the Psychomotor Domain*. David McKey, New York.
- HEYLINGS D (2002) Anatomy 1999–2000: the curriculum, who teaches it and how? *Med Educ*, 36(8): 702-710.
- HEYNS M (2007) A strategy towards professionalism in the dissecting room. *Eur J Anat*, 11(S1): 85-88.
- JCEM (1997) *The Good Assessment Guide. A practical guide to assessment and appraisal for higher specialist training*. Joint Centre for Education in Medicine, London.
- KRATHWOHL DR (2002) A revision of Bloom's taxonomy: An overview. *Theory Pract*, 41(4): 212-218.
- KRATHWOHL DR, BLOOM BS, MASIA BB (1964) *Taxonomy of Educational Objectives: Handbook II, Affective Domain*. David McKey, New York.
- LOUW G, EIZENBERG N, CARMICHAEL SW (2009) The place of anatomy in medical education: AMEE Guide no 41*. *Med Teach*, 31(5): 373-386.
- MCHANWELL S, DAVIES D, MORRIS J, PARKIN I, WHITEN S, ATKINSON M, DYBALL R, OCKLEFORD C, STANDRING S, WILTON J (2007) A core syllabus in anatomy for medical students - Adding common sense to need to know. *Eur J Anat*, 11(S1): 3-18.
- MILLER GE (1990) The assessment of clinical skills/competence/performance. *Acad Med*, 65(9): S63-67.
- MORRIS J, CHIRCULESCU A (2007) Structure and assessment of a short intense clinical anatomy course shortly before clinical studies. *Eur J Anat*, 11(S1): 95-98.
- NORCINI JJ (2003) Setting standards on educational tests. *Med Educ*, 37(5): 464-469.
- PEEL S (1998) An innovative problem-solving assessment for groups of first-year medical undergraduates—Think Tanks. *Med Educ*, 32(1): 35-39.
- SCHUBERT S, SCHNABEL KP, WINKELMANN A (2009) Assessment of spatial anatomical knowledge with a 'three-dimensional multiple choice test' (3D-MC). *Med Teach*, 31(1): 13-17.
- SIMPSON E (1971) *Educational objectives in the psychomotor domain*. In: KAPFER MB (ed). *Behavioral Objectives in Curriculum Development: Selected Readings and Bibliography*. Educational Technology Publications, Englewood Cliffs, NJ, 60-67.
- SINCLAIR DC (1965) An experiment in the teaching of anatomy. *Acad Med*, 40(5): 401-413.
- SMITH CF, MATHIAS HS (2010) Medical students' approaches to learning anatomy: students' experiences and relations to the learning environment. *Clin Anat*, 23(1): 106-114.
- SULLIVAN GM (2011) A Primer on the validity of assessment instruments. *J Grad Med Educ*, 3(2): 119.
- WINKELMANN A (2007) Anatomical dissection as a teaching method in medical school: a review of the evidence. *Med Educ*, 41(1): 15-22.
- YAQINUDDIN A (2013) Problem-based learning as an instructional method. *J Coll Physicians Surg Pak*, 23(1): 83-85.
- YAQINUDDIN A, ZAFAR M, IKRAM MF, GANGULY P (2013) What is an objective structured practical examination in anatomy? *Anat Sci Educ*, 6(2): 125-133.