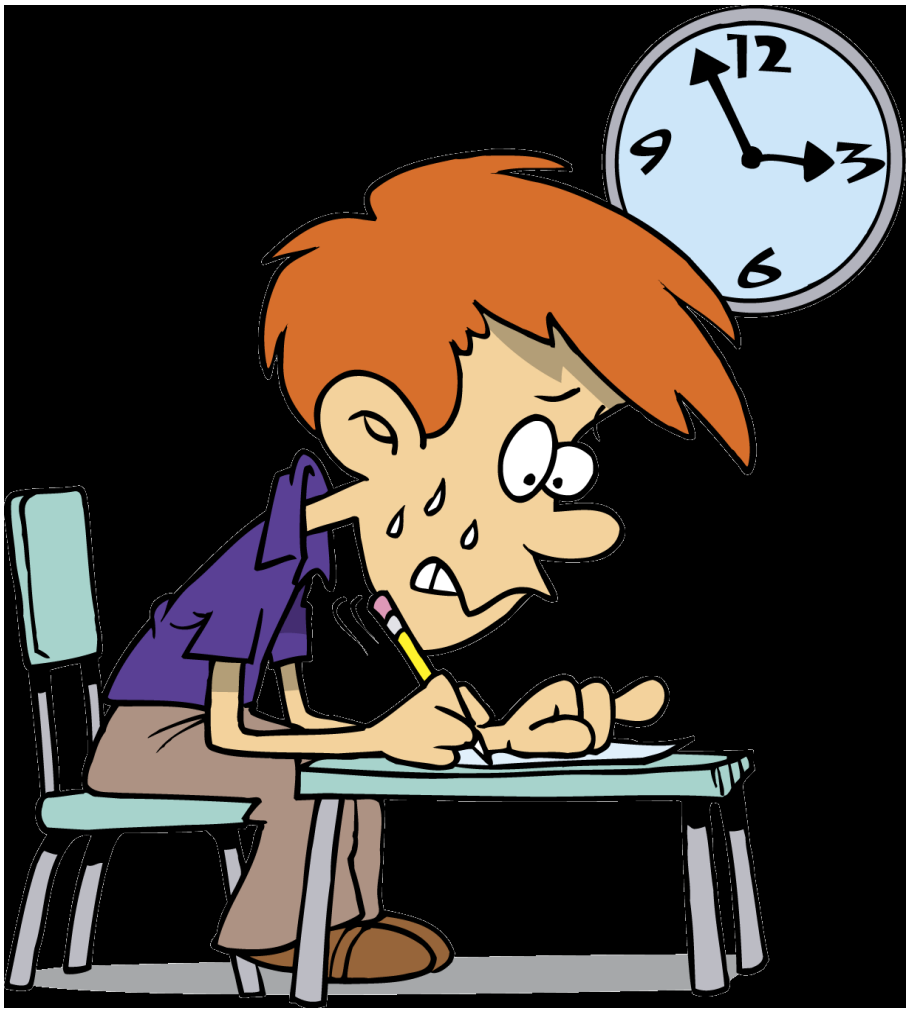
“Exam development”



Koning Willem I College
’s-HERTOGENBOSCH

Koning Willem I College
's-HERTOGENBOSCH

Content

1. Introduction

Constructing exams includes more than just the construction of an exam, both in practical and theoretical terms. It forms the backbone of the educational curriculum and with that it is a significant factor in defining the educational quality. Examination formulated appropriately clarifies which learning results should be achieved and with that it guides the educational content.

To know how a proper exam should look like, you need to be familiar with the necessary knowledge and skills. In this short course you will be provided with the theory in this field.

Aim of the course "Exam development"
The aim of this course is to develop the expertise in the field of the construction of exams, in particular the quality requirements on exams.

2. The aim of examination

Exams are used to determine whether a student achieved the required level and is promoted to or has failed/passed the next year. What is asked in the exam depends on the qualification requirements of the educational curriculum. The exams are linked to elements from the curriculum. Having clarified that, you can compose an exam.

3. Quality requirements for exams

When developing exams, in whatever form, certain guidelines are formulated. These guidelines are tools which make a quality exam product. These quality requirements are essential to all types of exams, for example theoretical, practical and skills exams. Does the assessment measures, in practice, what we want to measure? Is it possible that candidates show different results than you expected? And is that competence at a proper/correct level within the given situation?

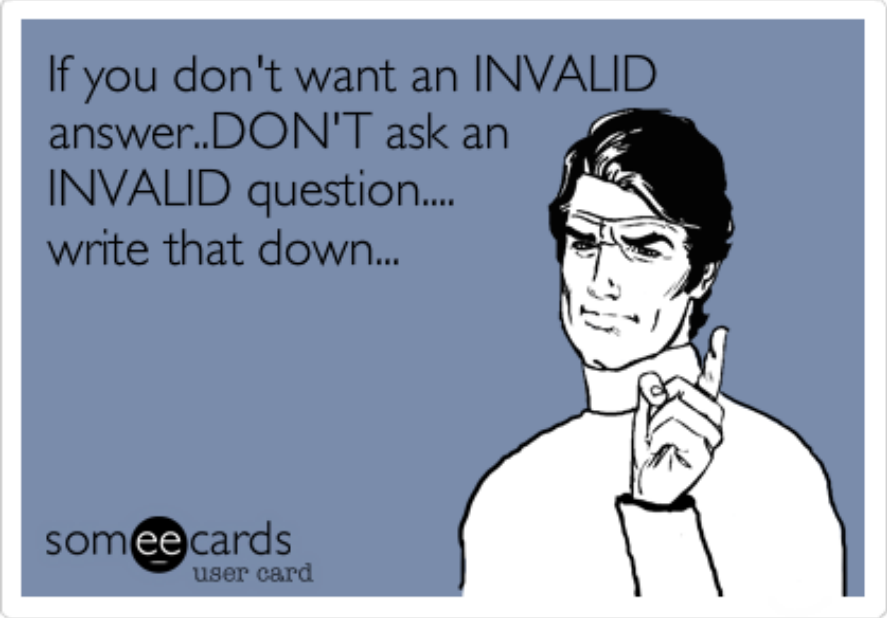
The aim is to make a quality exam. There are several quality requirements which you can incorporate, but we focus on the three most important requirements, validity, reliability and transparency. Per quality requirement we explain the degree, give some examples and, lastly, we test your knowledge.

Validity

What is validity?
The validity of an exam has to do with the significance, practicability and accuracy which are drawn from the exam results (Van Berkel & Bax, 2014). Therefore the question is whether the exam results gives us information in which we are interested and or, based on these results, the desired decisions can be made (Wools, 2011). The exam should include the proper exam material, with the proper complexity.

The question with validity is, do the exam results correspond to what we hoped for? In other words, do we measure what we want to measure (Valcke, 2010)?

An example of a non-valid end result is to give a math test and ask questions about fractions and afterwards give opinion on arithmetic skills of the student.



Types of validity
Content validity is the extent to which the exam as a whole represents the measurements. In other words, every important subject of the curriculum should be addressed in the exam.
Concept validity is the extent to which the exam result can be seen as a measurement for the intended theoretical concept. For example, it is non-valid examining insight by asking about memorised definitions.

Validity depends on the aim of an exam and therefore it is possible that an exam can be valid on the one hand and not on the other hand.

Validity can be improved by:
- checking the items/exams by other people;
- checking the content structure;
- selecting a representative subset from the desired learning objectives (curriculum) and, based on this, constructing an exam (tool: specification table /exam matrix).

Examples

Example 1(correct example of validity)

Exam Listening Skills English.
In the listening comprehension test English the student will hear (a part of) a foreign radio show. This exam is divided into two fragments. After every fragment the student needs to answer a multiple-choice item about what he/she has heard. The student has about 20 seconds for each question, after which the next fragment follows. With the help of multiple-choice questions it is measured whether the student listened properly or not.

Explanation:

The validity of this exam is correct for the following reasons:

- It is intended that the results of this exam say something about the listening skills of the students who sit this exam. We interpret the exam result as a measurement for listening skills. Next we use the result to determine whether a student has sufficient listening skills to pass the exam English Listening.

Example 2 (incorrect example of validity)

Knowledge exams about the anatomy of the foot and hand. The knowledge exam should test actual and comprehensive knowledge.

The exam indicates that the questions are not limited to the anatomy of the hand and foot. There are also questions about other parts of the human body. Furthermore, the exam questions only address actual knowledge.

Explanation:

The validity of this exam is not correct for the following reasons:

- the exam is not limited to the required content, we do not measure what we are supposed to measure. The exam questions should be limited to the anatomy of the hand and foot;
- the exam is limited to the examination of actual knowledge which causes insufficient measurements. Here the comprehensive knowledge is not being examined.

TEST YOUR KNOWLEDGE ON VALIDITY

ANSWERS: WHAT DO YOU KNOW ABOUT **VALIDITY**

VALIDITY			
1.	Validity is that you measure what you want to measure	O True	O Not true
2.	You can express the extent of validity in a number	O True	O Not true
3.	A valid exam is a correct exam	O True	O Not true
4.	A multiple-choice exam as exam to obtain a navigation licence as captain of the boat, is non-valid	O True	O Not true
5.	There are more types of validity	O True	O Not true
6.	The validity of the English exam increases when speech skills are included	O True	O Not true
7.	A medical thermometer is a valid measuring instrument	O True	O Not true
8.	An exam which everyone passes, is non-valid	O True	O Not true
9.	The validity of a (professional) exam increases when taken in an authentic situation	O True	O Not true
10.	It is valid to examine professional competence by means of a written exam with multiple-choice items	O True	O Not true

1. TRUE
   This is indeed an often used definition of validity. Although the explanation is short and simple, in practice it proved to be difficult to remain consistent with the aim of the exam and the learning goals to be measured, when making the exam items.

2. NOT TRUE
   There is no number to express validity, but the extent of validity is always something you should consider in relation to the aim you want to pursue with your measurement. And because validity is a flexible concept, this subject will often lead to discussion.

3. NOT (NECESSARILY) TRUE
   An exam should be valid. That is a condition. That is true, but validity is not the only criterion for a correct exam

4. TRUE and NOT TRUE
   You can only really know if someone is capable of handling and navigating a boat when you let the person actually steer. A multiple-choice exam is applicable to determine whether someone knows the rules and regulations of sailing. And if that is a sufficient condition to safety on the water, then you can deem the exam to be reasonably valid. Luckily for safety on the road, we have chosen a driving test as a more valid exam, whereby you indeed need to demonstrate whether you are sufficiently able to drive.

5. TRUE
   There are more types of validity. Content validity is the best-known type. With that we mean whether the exam covers the content to be measured. A second type is concept validity. Is the exam adapted to the content which we want to measure? There are more types, which we will not include right now.

6. TRUE
   Speaking English is of great importance to further the skills in English. Therefore, an oral exam should be part of the English exam.

7. NOT (NECESSARILY) TRUE
   It depends on what you want to measure. If you want to measure whether a patient has a temperature, it is a valid instrument. When you use it to weigh the patient or measure the temperature in the desert, it is not a valid instrument.

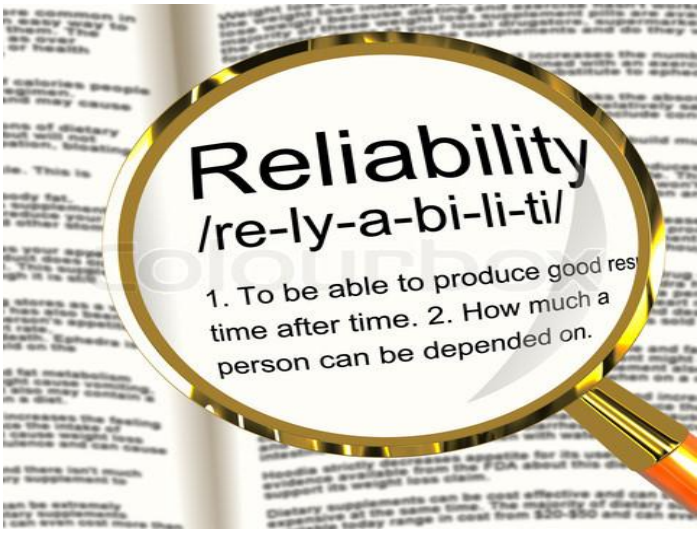
8. NOT TRUE
   The number of people passing an exam is not a correct indicator for the validity of an exam. However, it is advisable to analyse the exam questions (maybe the questions were too easy?).

9. TRUE
   Validity is measuring what you want to measure. If you want to measure whether a person is capable of acting professionally, it is best to do so in an actual professional situation. For example, if you want to measure whether a person is capable of giving injections correctly to a patient, it is best to do so with a real patient instead of a doll. Indeed, a doll cannot react the way a real patient does.

10. NOT TRUE
    Assessing professional competence is only possible by letting the students perform the acts. A written exam alone is not sufficient.

Reliability

What is reliability?
What is the predictive value of the achieved result of an exam? How great is the probability that the student achieves the same result with a second measurement? The more measurements, the more reliable the information is.
If someone else would evaluate, will that person come to the same conclusion? If I myself would evaluate the learning process again, would I come to the same conclusion? A higher reliability can especially be achieved by using very clear criteria, asking sufficient questions and/or relying on a sufficient large collection of observations of learning behavior. The use of multiple observers, which we call 'the four eyes principle', is a trusted technique with qualitative assessments (for example, based on observation).



The link between reliability and validity is comparable to the link between a foundation and the house on top of it. An observer only sees what is visible, the exam and the questions. The foundation (reliability) and everything affecting it, such as material, water level, depth, strength, remains invisible. Only after a soil research (exam analysis) the quality of the foundation will show. If the base is quicksand (the exam is unreliable), it is no use building a house (the exam), even though the architects did their best to design a sophisticated house. It is also possible to reason backwards. Namely, laying a hard foundation, thus very reliable, with on top of it a monster of a house. For example, a house with only bathrooms. That 'house' will still be there in a hundred years, but it cannot be named house because it does not meet the criteria of a house (Van Berkel 2014).

Reliability can be improved by:
- raising the number of items in an exam;
- objective and measurable assessment criteria;
- multiple moments of measurements;
- increasing expertise of examiners;
- setting conditions of exam should be equal to all participants;
- alignment between all examiners involved in the exam in advance;

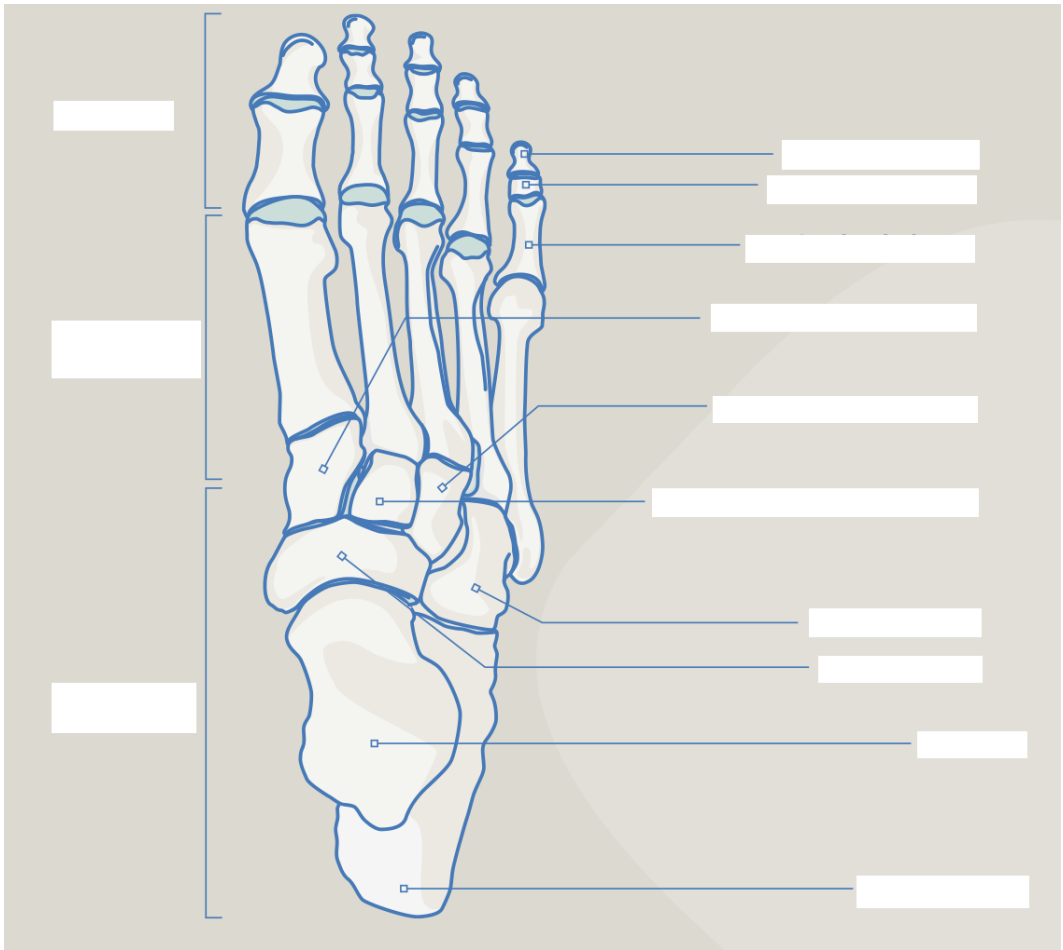
Examples

Example 1(correct example of reliability)

The human body is part of the curriculum. To gain insight to the level of knowledge of anatomy, the students will be presented with a knowledge exam. One of the questions is about the anatomy of the foot.

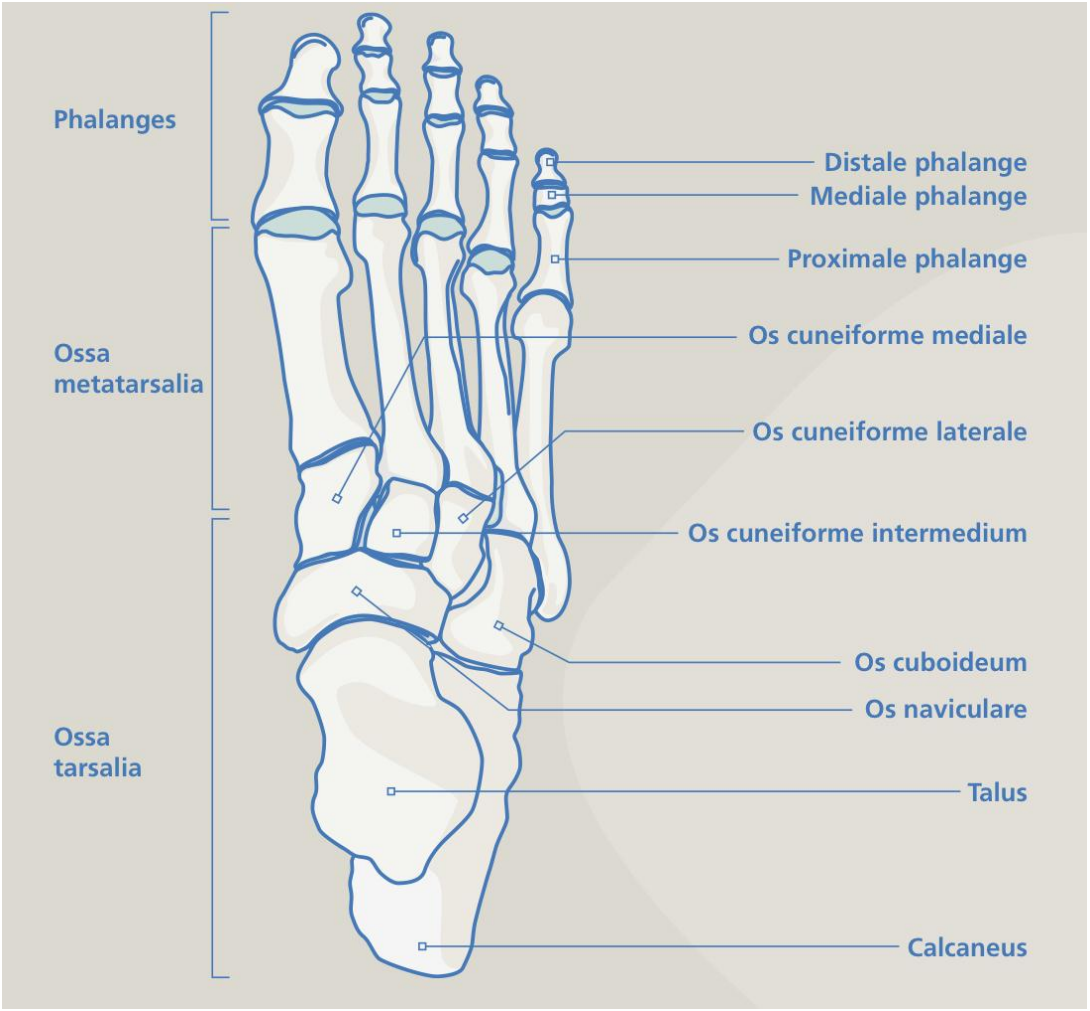
Exam question:

1. Write down the correct Latin name of the relevant part/bone in the foot in the illustration below.
   This assignment counts for the maximum of 13 points. Each correct Latin name 1 point. Divergent answers count for 0 (zero) points.



Answering model:

Exam question 1.
This assignment counts for the maximum of 13 points. Each correct Latin name 1 point. Divergent answers count for 0 (zero) points.



Explanation:

The reliability of this question is correct for the following reasons:
- exam conditions apply to all candidates;
- the assessment criteria are objective and measurable;
- clear correction regulations.

Example 2 (incorrect example of reliability)

Candidate A works at Johanneswerk, a care and nursing home for the elderly in La Nucia (Spain). Candidate B works at the first aid in East Talinn Central Hospital (Estonia). Both candidates take the exam below.

Situational picture:
Context in which the exam has been written (based on the curriculum).
The nurse gives personal care by supporting the patient with regard to eating and drinking, excretion, mobility, sleep-wake rhythms and monitoring vital functions. She observes, notes and monitors continuously changes in the health status and the well-being (cyclical process). She continuously applies risk identification to ensure the patient safety. She interprets the collected data, consults experts, if needed, makes agreements on the policy to be followed and carries it out or initiates it. She reports the collected data and findings and adjusts the nursing plan, if needed.

The exam takes 1 day. The exam candidate carries out all the tasks above as a nursing employee.

Explanation:

The reliability of this question is incorrect for the following reasons:

- you should consider whether the conditions of exams are equal to candidate A and B.
  Do the workplace, the activity and the workload at the first aid not look completely different than at a care and nursing home for the elderly;
- if the answer is yes, the conditions for exam candidates are not equal and this will affect the reliability of the assessment.

TEST YOUR KNOWLEDGE ABOUT RELIABILITY

RELIABILITY

1.	The assessment regulations should have guidelines which lead to an independent assessment.	O True	O Not true
2.	The assessor should be professionally competent.	O True	O Not true
3.	The available time for an exam in the professional practice may vary per candidate.	O True	O Not true
4.	The examiner should be competent in the field of assessments.	O True	O Not true
5.	The exam and the assessment in the professional practice may take place whenever the examiner wants.	O True	O Not true
6.	The reliability is correct if the exam “Management skills” is taken in a one-man-business.	O True	O Not true
7.	Correct assessment guidelines increase the reliability of the exam result.	O True	O Not true
8.	In a multiple-choice exam the reliability increases as the number of questions rise.	O True	O Not true
9.	If the practice supervisor also assesses the exam, it increases the reliability of the assessment.	O True	O Not true
10.	An instruction for all examiners increases the reliability of the assessment.	O True	O Not true

ANSWERS: WHAT DO YOU KNOW ABOUT RELIABILITY

1. TRUE

The assessment regulations should include such guidelines, it is possible to randomly deploy (professional competent) examiners for the assessment. Based on the guidelines it should not make any difference which examiner is being deployed (not person dependent).

2. TRUE

The assessor should be professionally competent to act as an examiner. Only if you are professionally competent, you can assess whether shown behavior meets the requirements. Only in case of a multiple-choice exam the assessor cannot be professionally competent.

3. NOT TRUE

The same rules and exam conditions are applicable to all candidates. Therefore the starting and finishing moments should be the same for all candidates. It is possible that some candidates finish the exam quicker than others.

4. TRUE

The assessor should be professionally competent to act as an examiner. Besides being professionally competent, he should also be competent in the field of assessing.

5. NOT TRUE
The assessment cannot depend on accidental factors, time among others. The assessment should include clear information on the method of assessment, assessment criteria, exam conditions and the exam location and time. The conditions, assessment criteria etcetera should be equal to all candidates. These should be clear to the candidate beforehand.

6. NOT TRUE
In this situation the exam “Management” cannot be taken. Who are you managing?

7. TRUE
The clearer the assessment rules (assessment criteria, method of assessment), for the examiner, the bigger the chance the exam results are close together.

8. TRUE
The more questions, the less chance , the more reliable the result. Here we assume well formulated, clear questions.

9. NOT TRUE
The assessor should be independent (meaning: not involved in the candidate’s guidance process).

10. TRUE
If the instruction is clear and the examiner exactly knows what is expected of him, the bigger the chance the measurement of results is reliable.

Transparency

What is transparency?
Transparency of an exam refers to the (lack of) clarity among students about what to expect . (Van Berkel 2014)
In order to safeguard the transparency, it is necessary to properly and fully inform the students. Beforehand the student should have no further questions about the preparation, execution and assessment of the exam. A transparent exam is not only of importance to the student but also to the examiner and any other persons involved.

The transparency of an exam is assessed by the extent to which the persons involved are being informed and instructed. Transparency can be improved among other things by providing the following information:

- beforehand it is clear what the exam is about, what kind of questions will be asked or what kind of assignments will be given;
- beforehand it is clear what is expected of the student, with regard to their level of knowledge and skills;
- beforehand it is clear whether failed elements can be compensated with passed elements;
- beforehand it is clear what the result of the exam can be, for example: passed, partly passed, redo certain elements, failed (resit the complete exam). The final score can also be expressed in marks;
- beforehand it is clear how much time there is available to make the exam;
- beforehand it is clear when the exam starts and ends (starting time and ending time, including date and location);
- beforehand it is clear which materials and tools the student can use during the exam;
- it is clear to the student what performances are expected of him in the questions and/or assignments in the exam;

- for the execution of the exam, it is of importance to the examiners and student that the exam is provided with exam conditions;
- beforehand it is clear how the candidate is able to score;
- beforehand it is clear how many points the student can score;
- the minimum score the candidate needs to attain to pass the exam is clear (cutting score);
- it is clear whether there is a scale score;
- it is clear which score corresponds to which final score (standardisation).

Examples

Example 1: (correct example of transparency)

Exam question from a knowledge exam of the Caregiver programme:

1. There are six guidelines described about lifting techniques in order to prevent/discuss back complaints. Write down four of the six guidelines.

   1. ----------
   2. ----------
   3. ----------
   4. ----------

   Maximum of 8 points for this question. 2 points per correct guideline. Divergent answers count for 0 (zero) points. Only the first four answers will be assessed.

   Answer:
   1. Do not lift more than 20 kilos
   2. Wear flat shoes
   3. Always stand in a small spread position when lifting
   4. Keep your back straight when lifting
   5. Bend your knees while lifting
   6. Keep your arms as close to your body as possible

Explanation:

The transparency of this question is correct for the following reasons:

- to the candidate it is clear which knowledge aspect accompanies the question;
- to the candidate it is clear what performance is expected;
- to the candidate it is clear the maximum of points he can score for this question;
- to the candidate it is clear that there is a scale score and how he can score;
- to the examiner it is clear which answers are considered correct and which answers are considered incorrect.

Example 2: (incorrect example of transparency)

Exam question from a knowledge exam of the Nurse programme:

1. Describe two (2) possible risk factors and two (2) possible causes of a Cerebral Vascular Accident (CVA).

Answer:

Possible risks
- Smoking
- High blood pressure
- A high cholesterol
- Diabetes mellitus
- Overweight
- Physical inactivity

Possible causes

A CVA can occur when:

- there is a temporary blockage of an artery in the brain (brain infarct)
- a leak in an artery in the brain (brain haemorrhage). This leads to the loss of blood flow to a part of the brain.

General causes of a CVA:

- Atherosclerosis: causes arteries to narrow and the artery walls to weaken.
- Blood pressure: the artery wall becomes damaged because of the continuously high blood pressure, which may lead to an infarct/bleed.

Specific causes of an infarct:

- Thrombosis: narrowing and damage in the arteries can cause a blockage in little blood vessels
- Embolism: a dislodged blood clot in another part of the body can travel through the bloodstream and cause a blockage in the brain.
- Composition of the blood: when there are relatively too much coagulation factors, the blood will easily coagulate and the chance of a blockage in the arteries is bigger.

Specific causes of a bleed:

- When it becomes weaker, the artery wall can rupture
- Congenital artery defects (because of a different structure, arteries can easily rupture)
- Brain traumata (subdural hematoma)
- A skull fracture which pierced a large artery (epidural hematoma)

- Bleeding in a vascular tumour.

Source: http://www.home.zonnet.nl/sailbert/CVA/ziektebeeld.htm

Per cause and per risk factor 1 point. This question counts for 4 points.

Explanation:

The transparency of this question is not correct for the following reasons:

- to the candidate it is unclear how to answer (oral or written description);
- to the candidate it is unclear the maximum of points he can score fort his question. This is clear from the answering/correction model;
- to the candidate it is unclear whether there is a scale score and how he can score subsequently;
- the question is too broadly formulated, because of that there are more answer possibilities than the answering/correction model shows. These answers can also be considered correct. Because of this it is unclear to the candidate what is expected. The question should be formulated more concretely (frame working). In this way it is easy to lead to the desired result/answer. It is a good idea to use correct and incorrect examples from which the candidate has to choose;
- to the examiner it is unclear which performances should be considered correct and which should considered incorrect.

TEST YOUR KNOWLEDGE ABOUT TRANSPARENCY

TRANSPARENCY			
1.	Assessment criteria should only be clear to the examiners.	O True	O Not true
2.	To the candidate only the total amount of points is of importance at the exam.	O True	O Not true
3.	It is of importance to the examiners and students that the exam is provided with exam conditions.	O True	O Not true
4.	Each exam should include an instruction for the candidate and an instruction for the examiner.	O True	O Not true
5.	80% of the assessment criteria should be provided with a quality standard (objective and measurable).	O True	O Not true
6.	The available time for taking the exam should not require any mentioning.	O True	O Not true
7.	Beforehand it should be clear what the exam is about, what kind of questions will be asked or what kind of assignments the student should expect.	O True	O Not true
8.	The exam should include several instructions with regard to the execution of the exam.	O True	O Not true
9.	It should be clear to the student what the minimum score is to pass the exam (cutting score).	O True	O Not true
10.	It should be clear to the student which score corresponds to which final score (standardisation)	O True	O Not true

ANSWERS: WHAT DO YOU KNOW ABOUT TRANSPARENCY

1. NOT TRUE

The assessment criteria should be clear to the candidate as well as the assessor. These can/may be described differently for the assessor (content wise they should be the same). Example candidate: you should act in accordance with the rules and regulations with regard to the safety. Example assessor: the candidate has acted in accordance with the rules and regulations with regard to the safety.

2. NOT TRUE

The candidate should be given more information than just the total number of points to be scored. The candidate also should be informed about the cutting score, but also how many points he can score per assignment. If there is a partial score (scale score), it should be clear how the partial score is applied.

3. TRUE

The exam condition and instructions for the execution of the exam should be clear to all persons involved in the examination.

4. TRUE

With every exam, regardless of the form of examination, there should be an instruction for all persons involved in the conduct and assessment. It means that an instruction should always be provided for the candidate and examiner. If there are

also other persons involved, for example a simulant, there should be an instruction available.

5. NOT TRUE
All assessment criteria included in the exam should be provided with a quality standard. The quality standard should be objective and measurable. Non-objective and/or none or hardly measurable assessment criteria make the exam unreliable.

6. NOT TRUE
The available time should be clearly mentioned to the student as well as the examiner. During the exam the examiner should adhere strictly with the time made available.

7. TRUE
Beforehand the student should be informed about the content of the exam, the kind of questions or assignments he can expect. This information is crucial to help him prepare for the exam.

8. NOT TRUE
For the execution of the exam it is of the utmost importance to the examiners and student that the exam is provided with a clear and complete set of instructions of the exam.

9. TRUE
To the student it should be clear the maximum of points he can score and the minimum score he needs to attain to pass the exam (cutting score).

10. NOT (necessarily) TRUE
It is not mandatory to provide the candidate insight into which score corresponds to which final score. However, it is mandatory to clarify the cutting score. As a service, this information may always be provided.

Bibliography

Inspective van het Onderwijs (2012)(Education Inspectorate) Toezichtkader BVE 2012 (Supervisory Framework). Utrecht: Inspectie van het Onderwijs.
Van Berkel, H., Bax, A. & Joosten-ten Brinke D. (2014). Toetsen in het hoger onderwijs (Higher Education Examination). Houten: Bohn Stafleu van Loghum.
Sanders, P. (2011) Toetsen op school (Examinations at school). Arnhem: Cito.
Valcke, M. (2010). Onderwijskunde als ontwerpwetenschap. Een inleiding voor ontwikkelaars van instructie en toekomstige leerkrachten (Didactics as design science. An introduction for developers of instruction and future teachers). Gent: Academia Press.

Internet resources

Schoolexamensvo.nl
www.toets.nl