Competence-oriented oral examinations: objective and valid

Karl Westhoff & Carmen Hagemeister

Abstract

The Decision-Oriented Interview (DOI) takes account of behavioral regularities in interviewing and is therefore suitable for all types of interview, including oral examinations. First we describe how all effective oral examinations examine the content of a certain part of a course of study or the requirements of practice after graduation, and show that these requirements can always be arranged into a generally valid hierarchy. In an oral examination it is best to start by examining the simplest requirement and then move on through the hierarchy. In order for an examination to be fair and valid one needs a description of the basic set of all questions for the specific subject. This makes possible a criteria-oriented measurement of the candidate’s competence. The characteristics of the DOI as an oral examination technique are described, and an example given to show what the sequence of questions in the subject “Psychological Assessment” can look like. Empirical results from far more than 1000 oral examinations have shown that the use of the DOI leads to objective and valid assessments. In the conclusion checklists for the DOI as an oral examination procedure are given and explained.

Keywords: oral examinations, decision-oriented interview, requirements, hierarchy of requirements, criterion-oriented measurement, checklists, objectivity, validity

1 Correspondence concerning this article should be addressed to: Prof.em. Dr. Karl Westhoff, Wiesenstraße 63a, 51371 Leverkusen, Germany; email: mail@karl-westhoff.de

2 TU Dresden, Germany
1 Introduction

Oral examinations are taken in many fields. For decades they have been the subject of criticism (summarized e.g. by Birkel, 1984), one argument being that agreement between different boards of examinations, i.e. the objectivity of evaluation, is low. In psychology, different results have been found concerning the objectivity of oral examinations: Trimble (1934) and Preiser (1975) found that oral exams in psychology are relatively objective. Dienst and Armstrong (1988) found a moderate objectivity, while Engvik, Kvåle and Havik (1970) assessed the objectivity as non-satisfactory. Novy, Kopel and Swank (1996) were the first to study the psychometric features of oral exams for licensing in psychology. All these studies taken together lead to the conclusion that oral exams in psychology are not, by nature, highly objective. An earlier study (Westhoff, Hagemeister & Eckert, 2002) showed that a high degree of agreement between examiner and assessor and sufficient agreement with the examinee after the exam can be reached when explicit evaluation rules are used and the exam is properly structured. Hagemeister and Westhoff (2010) replicated these results and demonstrated the content validity of oral examinations and their corresponding results. Nowadays even forms of examinations that have a long tradition, such as oral examinations in general and doctoral vivas in particular, should be shown to be evidence-based (McAdams & Robertson, 2012).

We would like to report here an extract of our efforts over a few decades to communicate to students not only knowledge of a subject but also the mastery to apply that knowledge competently (Westhoff & Hagemeister, 2014). Our main aim was always to put psychology students into a position where they were able to write psychological reports and were proficient in all the necessary steps therefor. The requirements of a psychologist in an assessment practice were always the criteria under which the performance of a student was measured in their (oral) examinations. We have always examined in a criteria-oriented and not a norm-oriented fashion in written and oral examinations.

Before the first author became an examiner in Psychological Assessment he had used the examination system presented here for the oral psychology examinations of education students taking education science as a subsidiary subject. The chairmen of the examinations were head-teachers or deputy heads of secondary schools. The examinations in Psychology were always held in tandem with an examination in Education Science, which was taken by a professor of education science chosen by the student. A mark was given for each part of the exam, and these were subsequently averaged to give an overall mark. Whereas the marks in education science were often discussed with the chairmen and changed by consensus, this was never necessary for the marks in psychology.

Upon receiving the authorisation to examine in the subject of Psychological Assessment, the first author introduced the documentation of the proposed marks from the minutes-taker and from the examiner, and the mark that was finally awarded, as well as the student’s self-assessment of their performance. The remarkable concordance between the mark awarded and the student’s self-assessment was the subject for discussion among our colleagues. Thus arose the proposal from Günter Debus, which was implemented from that moment onwards, that the degree of concordance between the student’s self-assessment before the exam and the mark awarded be taken as a criterion of validity.
In the course of the following 25 years we have collected these and other pieces of information, examined them with regard to their predictive power, and published them (Westhoff, Hagemeister, & Eckert, 2002; Hagemeister & Westhoff, 2010). The “tool kit” developed by the first author, “The Decision Oriented Interview (DOI) as a Selection Interview” (Westhoff, 2009) can be helpful in oral examinations and markedly reduce the stress and workload for examinees, examiners and minutes-takers. Furthermore, a procedure according to the DOI leads to objective, valid and thus fair oral examinations (Westhoff & Hagemeister, 2009). The procedure to be presented here can be applied to all orally examined subjects.

Establishments that desire advice in the development of a competence-oriented teaching and examination system can receive this, as well as training for the teachers and students, from the first author.

2 An oral examination examines requirements from practice

In an oral examination, the examiners examine the knowledge and mastery of examinees in the presence of an expert who takes the minutes. They assess how the examinees measure up to previously defined requirements of knowledge and mastery.

These requirements for knowledge and mastery result from the following sections of the studies or training or from the subsequent practical employment. There is no fundamental difference between the two, then the core of the matter is that the examinees should be able to use their knowledge and mastery in the following section of their studies or in their subsequent practical employment.

Whereas in a course of academic study a student can learn and be taught only by means of examples that are as instructive as possible, an apprentice learns and practices on all the tasks that will typically present themselves in their trade or vocation. The corresponding examination discussions are therefore subject to fundamentally different requirements.

In learning by example the knowledge to be acquired cannot in principle be complete. Correspondingly, the translation of knowledge into solutions for practical problems can only be demonstrated, exercised and examined exemplarily. If examinees are able to do this then one assumes that they will be able to acquire the necessary knowledge and translate it into mastery in the future. An examination within the framework of a vocational training is also only possible as a test of randomly chosen examples of knowledge and mastery, but this for practical reasons.

In a vocational training, the realities of the subject, car mechanics for example, prescribe what tasks a qualified person must be able to perform. For that reason one would probably find no differences between Bavaria and Lower Saxony in the expectations of what a car mechanic must be able to do.

In a course of academic study the expertise of the responsible professor sets the standard for what the students must attain because here there is normally no obviously definable set of facts or knowledge to be learned like in the training of a car mechanic. It is there-
fore of crucial importance “what” one has studied “under whom”. In the most favourable situation the employers of such students then know what is to be expected of a new employee.

The subjects taught at universities that cover topics with a very high reality content, such as medicine, have a large overlap in the knowledge that has to be conveyed. The less unambiguous the requirements of reality are, the greater the scope for differing emphasis in the teaching. The overlap of the knowledge conveyed by different lecturers in a subject is then smaller than in the first case.

2.1 Hierarchy of requirements in practice

If one looks at the requirements in practice, one finds everywhere the following hierarchically listed requirements for knowledge (1. and 2.) and mastery (3. and 4) for a qualified person:

1. The presentation of fundamental specialist information;
2. The presentation of extensive detailed specialist information;
3. The delineation of commonalities and differences in various specialist procedural methods;
4. The presentation of one or more approaches for solving a practical specialised problem.

From qualified persons one can at least expect that they know the basic facts and information in the subject and can delineate these in an understandable fashion. In practice most questions can be answered satisfactorily with such knowledge. There are, however, questions in practice that require that a qualified person can depict comprehensive and detailed knowledge correctly and understandably. This clearly goes beyond the requirements in the case first described.

The first two requirements in practice merely require the reproduction of knowledge. But that does not provide satisfactory answers to all questions in practice. Questioners are perfectly within their rights to expect from qualified persons that they can explain accurately and in detail the advantages and disadvantages of certain technical procedures in such a way that questioners can decide themselves which of these to choose. Not just knowledge but also mastery is clearly required here.

One cannot imagine that qualified persons could be able to offer one or more approaches to a solution for a problem without first making a systematic comparison of different procedures. For this the description of an approach to a solution is composed of far more than just the description of commonalities and differences of various approaches. An analysis of the starting situation, the planning of a sequence of steps, and a time schedule, for example, are necessary. In an individual case the pros and cons of different approaches must also be analysed.
2.2 Requirements increase in difficulty

In this hierarchy of requirements candidates must first demonstrate their knowledge and then their mastery in two stages. Logically, candidates can only tackle a following stage when they have successfully completed the previous stage. For this hierarchy the following stage is always more challenging and demanding than the previous one. In principle, an oral examiner can put questions in an arbitrary order from the various stages of difficulty, but this has serious disadvantages both for the candidate and the course of the examination if the candidate does not have a competent command of the subject, which is naturally often the case.

If candidates realise that they cannot answer a question properly, this puts additional pressure on them and can lead to a mental block that prevents them calling up knowledge that would otherwise be at their disposal. Such a procedure can lead to an underestimate of the candidates’ potential and thus to an unjustified bad mark. Questions that are too difficult for the candidates’ level of knowledge and mastery can lead to their attempting to answer by “blathering”, which leads to a waste of examination time in which they could have been demonstrating what they really know and master.

It therefore suggests itself that the candidates be examined orally in a sequence of four stages of difficulty. The examiner can move to question at the next stage of difficulty if a candidate has adequately and correctly answered sufficient questions at the current stage. And the candidates have the advantage that, for the stages that require mastery, they have already activated the necessary knowledge in various aspects and have it now more readily available.

2.3 Describing a basic set of questions

If one wants to put questions from practice that simulate a real life situation as closely as possible, then one can orientate oneself in general terms on the rules of a Decision-Oriented Interview (DOI) and formulate questions that accord with the four stages of difficulty proposed above. One has then determined “how” one wants to question.

For a fair examination to be performed it should also be made known beforehand “what” is going to be examined. One should describe explicitly the subject of the examination in all detail relevant in practice and therefore also relevant in the examination. The examination is thus an appraisal of the specialist practical competence, which is composed of knowledge and mastery.

The basic set of all possible questions derives from the combination of the subject of the examination and the manner of questioning (Klauer, 1987).
2.4 Content-valid examination makes possible a criteria-oriented measurement of competence

If an examiner uses a certain quota of questions from the basic questions set then the examination is content-valid. An examination is content-valid when the questions, decided in detail previously, are asked in the manner also decided previously.

Examiners can put questions with regard to each of the four general requirements described above concerning the competence of the examinees by starting at the lowest level. If it is sufficiently obvious from their answers that the examinees are competent at this level, the examiners can then ask questions at the next level of difficulty. The examiners carry on, moving up to the next levels in sequence until the candidates have achieved their highest level or the examination time is used up. The last level that the candidates have reached gives their level of competence. In practice it is found that most candidates who are familiar with the examination scheme and have comprehensive knowledge at their command can improve independently from “fair/pass” to “satisfactory”.

The four levels of specialist competence correspond to the marks “fair/pass”, “satisfactory”, “good” and “very good”. They can be upgraded or downgraded with “plus” and “minus” by 0.3 marks, with the exception that a mark of 1.0 cannot be so upgraded and of 4.0 cannot be so downgraded. Candidates who fall below 4.0 have simply failed the examination.

Requirement-based definitions of marks or grades make possible a criteria-oriented assessment of competence in a subject. For examinations these are fairer than a norm-oriented determination of marks based on frequencies in a so-called normal or Gaussian distribution, where one assumes competencies to be distributed according to a Gaussian curve within the group of examinees and then requires that the marks also be distributed accordingly. Viewed empirically however, different distributions can occur – such as when one is dealing with a particularly competent or a particularly incompetent group of candidates.

3 Characteristics of the Decision-Oriented Interview (DOI) as an examination procedure

With DOI as a selection interview procedure (e.g. for filling a job vacancy) one is appraising the characteristic behavior and experience of candidates. In contrast, with DOI as an examination procedure the aim is to assess the knowledge and mastery of candidates, i.e. their competence with regard to a certain subject matter.

Examiners are not interested in having the behavior and experience of candidates in certain defined situations described to them. They want to establish what candidates know and can do in a certain specialist area. The examiners therefore avoid all questions of the sort that are concerned with the behavior and experience of candidates.
“Why?” and “what is the reason?” are valuable questions when one is examining the knowledge and mastery of a candidate. In a selection interview such questions are not informative because they lead to the interviewee making causal attributions rather than describing behavior and experience.

3.1 The Examinee Must Know the Subject Matter and the Manner of Questioning

For an oral examination the candidates must have learnt and revised something so that they can show what they know and can do. So that they are able to do this, the subject matter that they will be examined in should have been described previously in all relevant detail. The candidates must also know what type of questions will be put to them in the examination. This does not mean that each precise question must be known to them in advance, but merely that the candidates must know what manner of questions the examiner will put to them so they can suitably prepare themselves mentally for the examination.

The scaling of the questions in order of increasing difficulty also belongs to the manner of questioning. This has advantages for the course of the examination and the best possible estimate of the candidates’ competence, as we have already described above. Furthermore, the examiners can allow the candidates to choose the starting point for the examination. This has the advantage, as a rule, that the candidates’ start in the oral examination is eased. This initial success heartens the candidates and allows them to cope better with the examiners’ subsequent questions.

The candidates know before the examination that they can choose a starting point. This allows them to choose a topic in which they are particularly interested or of which they have a good command. In addition they are informed beforehand that their task with the starting topic is solely one of reproduction, and that the examiners will choose subsequent questions at the level of “good” and “very good”. By this means the examiners counteract the tendency of some students to try to steer the oral examination or to want to enter a topic straightaway at the “very good” level.

4 Example of an examination in psychological assessment

What do you want to start with? (partly standardized procedures, decision-oriented interview)

From partly standardized procedures

Level 4:

What is an interview? (definition)

What do you know about the Decision-Oriented Interview (DOI) (Presentation of DOI)
Level 3:
(Here are put questions that cover precise information about DOI. These questions cover important information about DOI and depend on what the examination candidates themselves volunteer. For example:
Why can one describe the DOI as a toolkit?
What do you know about the reliability of the DOI?
What do you foresee as difficulties with the DOI as an in-depth interview technique?)
Do you know an example of an interview procedure that is largely standardised? (CIDI)
Level 2:
What do CIDI and DOI have in common? Where are the differences?
Level 1:
As a young psychologist you have to make suggestions for improving the admission interview procedure in a clinic. What are your considerations? How do you go about it?

From standardized procedures
Level 4:
What is a test? (definition)
Describe the Adaptive Intelligence Diagnosticum version 2. (AID 2 description)
Level 3:
(Here follow questions that cover precise information on AID 2. These questions include important information on AID and are dependent upon what the candidates themselves volunteer. For example: How is adaptive testing realised in AID?
What difficulties are inherent in this?
What preconditions must one check out in order to construct an adaptive test?)
Level 2:
Describe the commonalities and differences between AID 2 and WIT 2.
Level 1:
A 40-year old lorry driver has lost an arm in an accident and should be retrained. You have to examine him and as part of this make an assessment of his intelligence. How do you go about this?

From judgment and decision in psychological assessment
Level 4:
Describe Brunswik’s Lens Model.
Level 3:
What type of model is the Lens Model?
How is it that the Lens Model is a paramorphic model of judgment formation?
What would be an isomorphic model?
Do you know an example of an isomorphic model? (HYPAG)

Level 2:
Where do you see commonalities and differences between the Lens Model and HYPAG?

Level 1:
After your examination you are employed in an institution of your choice and are asked to improve the admission assessment. What do you consider?
How do you go about this?

5 Concerning the objectivity and validity of DOI as an examination technique

The DOI has proven to be a highly objective form of oral examination technique provided the examiner and the minutes-taker keep to the rules described above, i.e. as a rule the minutes-taker and examiner have a very high degree of agreement in their assessment. In cases of doubt the carefully recorded minutes can help in reaching a decision.

In over two decades we have conducted far more than 1000 oral examinations according to this procedure, and the agreement between the assessors was between 0.9 and 0.99. The DOI has proved itself to be highly objective as an oral examination procedure.

Upon requesting the candidates to assess their own performance after the conclusion of the oral examination we also found their opinion to be in high agreement (0.78) with the mark or grade achieved. This is remarkable in that it is naturally the task of the candidate in an oral examination primarily to achieve a good performance in answering the questions and not to indulge in self-observation. This was achieved so well because, inter alia, the candidates could well recognise the level of difficulty of the questions actually put to them. This is an essential precondition for experiencing an examination as fair.

As we have demonstrated above, configuring an oral examination according to the rules of DOI is content-valid, i.e. the construction is reproducibly free of errors or faults.

A positive correlation of 0.48 between the candidates’ estimates of their own knowledge and mastery before the examination and the grade or mark they were awarded also speaks for the empirical validity of DOI as an oral examination procedure. This coefficient of validity has the same dimension as with other structured selection interviews or good tests.
6 DOI checklists for oral examinations

The checklists for DOI as an oral examination procedure are presented below and are pretty much self-explanatory. The basis can be found at Westhoff (2009). Here only a few indispensable explanations should serve to aid understanding. After the presentation of each checklist some clarifications follow.

6.1 Checklist basis of DOI

1. Fundamental preconditions: Honesty and respect
2. Preparation: Recognition of one’s own negative or positive bias
3. Any bias has to do with the examiner’s experience of life
4. A positive bias causes problems
5. Initially the bias is positive or negative but never absent
6. To examine how well the candidate meets the requirements

The examiners and minutes-takers are called upon to treat the candidates with respect, to behave fairly towards them, and only to ask questions that concern the topic to be examined. A positive bias can cause problems if the examiners start off with questions that are pitched too high. A negative bias is also not uncommon but the examiners must always stick to the topic to be examined and formulate the questions fairly. A business-like, objective and friendly attitude should always be maintained. If examiners find this difficult then they will be unable to examine fairly.

6.2 Checklist characteristics of favorable questions

1. Clarify the frame of reference
2. Principle of “one thought/idea” per question
3. Short, pertinent questions
4. No suggestive questions
5. The examiners must also ask questions that they themselves may find “awkward” or “embarrassing”.

For each question it must be clear what it is referring to, and only one thing at a time should be asked. It is a bad academic habit to first hold a short lecture and follow it up with a sequence of questions. Short questions to the point are the hallmark of a professional. Examination questions should not be suggestive; such questions are unfair (see below). Some examiners find it “embarrassing” to confront excellent students with questions pitched at the “fair/pass” level. However there is no objective reason for this, then every fair examination starts at the simplest level in order to give every examinee the same chance, and also the opportunity to “warm up”. When the students know the procedure then they also know that some of the questions can be very simple and have no reason to fear that the examiners are laying a trap for them with a simple question.
6.3 Checklist suggestive questions

Questions in a discussion are suggestive when the answer “desired” by the examiner can be detected from the question because at least one of the following conditions has been met:

1. Prior information indicates the desired or expected answer;
2. The question already includes an evaluation of the subject of the question;
3. Something has been assumed as a given that cannot be assumed to be a given because it could be (have been) different;
4. Alternative answers have not been fully enumerated;
5. For a complete enumeration of alternative answers, or for “yes/no” answers one of the answers is more self-evident for the examinee;
6. The question contains leading filler words like “certainly” that give an indication of the expected answer.

As a rule open questions are good questions. Such questions do not prescribe how they are to be answered. If examination questions are suggestive then they are unfair, either towards the current examinees because they lead them onto thin ice, or towards other examinees who do not receive any suggestive help or hindrance from the examiner. Instead of using suggestive questions, the examiner can better examine certain issues in a miniature role play.

6.4 Checklist formulation of most efficient possible questions

1. Questions as short and apposite as possible
2. Active rather than passive voice
3. No unusual words
4. As few technical terms as possible
5. “Simple” language
6. Filter questions
7. Factual questions do not raise people up or put them down.

As in any good assessment interview the amount the interviewers (examiners) talk should be less than that of the interviewees (examinees). If at all necessary, questions should be given a short introduction. If during the course of the examination a conflict of opinion should occur between examiner and examinee, a discussion of this should be postponed until after the end of the examination.

The most efficient questions are those that test knowledge and mastery in the subject and not questions that primarily test verbal skills. The number of technical terms should be reduced to just the term that is being tested because otherwise it is not clear what the stumbling block was for the candidate. This is particularly important for examinees whose mother tongue is not the language of the examination. Language difficulties should not superimpose on difficulties in the subject. With a filter question, that is a closed question to which the answer is either “yes” or “no”, it is possible to ascertain whether the candidate has even studied the topic. If “yes” then open questions can be put
concerning the topic. If “no” then the examiner should not dig deeper but assess the candidate as failed in that topic.

6.5 Checklist hierarchy of requirements in practice

1. Presentation of basic specialist knowledge
2. Presentation of comprehensive and detailed specialist knowledge
3. Presentation of commonalities and differences of various specialist approaches
4. Presentation of one or more approaches for the solution of a practical specialist problem

The hierarchy of requirements in all practice is formulated generically. For each subject one must define what is to be learned and examined, i.e. what is to be understood in each case. Further above we have described how one arrives at a basic set of questions that cover a subject. What is to be taught, learned, mastered and examined is therefore defined explicitly for all the parties involved.

7 Outlook and discussion

The Decision-Oriented Interview (DOI) procedure presented here can be applied as a model for all oral examinations. It is based on general requirements for the knowledge and mastery to be expected from a qualified person or specialist, and provides a rule framework for the development of an integrated teaching and examining concept. The procedure in oral examinations has been optimised under psychological aspects to the point where students can prepare themselves optimally for an examination and are able to present their knowledge and mastery adequately in that examination. For examiners and minutes-takers DOI provides a basic structure that has been proven to lead to highly objective and valid examinations, and simultaneously frees examiners and minutes-takers (as well as examinees) to a large extent from avoidable strains and handicaps.

This system is also suitable for enabling minutes-takers to give relatively inexperienced examiners, after the examination, feedback on their performance. The rules are explicit, and so the minutes-takers can give the examiners feedback based on the examination minutes. The requirement that further questions should follow-up from what examinees have just said can be difficult for examiners in their first examinations. With a bit of practice examiners very soon get used to this.

The procedure for DOI as an oral examination technique is easy to learn. It is applicable to all university subjects and to all practical and vocational training programmes. With its high objectivity and validity values it is an ideal basis for allaying the fears that (some) students have of oral examinations. The students can concentrate on learning and revising and the examiners can establish in a friendly and fair atmosphere how the examinees measure up to the requirements of the course.
References


