The power of digital assessments

**Step by step to higher
quality and ease of use**

Brigitte Hellings

Colophon

The power of digital assessments
Step by step to higher quality and ease of use
Copyright © 2020 - Optimum Assessment

Updated edition
The power of digital assessments
Step by step to higher quality and ease of use
Copyright © 2020 - Optimum Assessment
Author: Brigitte Hellings

Title of original publication
Digital Assessment? Help!
Copyright © 2014 - Andriessen & Partners
Author: Manon Bonefaas

Editors
Natasja Kroon (Optimum Assessment)
Jantina Dorreman (Optimum Assessment)

Design
OOGT - Merkontwerpbureau

Printing
Q-Promotions

The Power of Digital Assessments is a publication of:
Optimum Assessment
Almystraat 10A
5061 PA Oisterwijk (The Netherlands)

+31 (0)13-5286363
info@optimumassessment.com
www.optimumassessment.com

Foreword

Focus on personal growth and professional development is increasing. As a result, assessment is also becoming more and more important for quantifying, recognising and appraising acquired knowledge and skills. We believe that intuitive assessment software is a powerful tool that facilitates growth in a smart, flexible and efficient way.

Over the past 20 years, we have been able to support and advise a large number of organisations in the field of (digital) assessment and examination. As a partner in digital assessment, Optimum Assessment is continuously innovating and further evolving its products and services. We work in close collaboration with our clients, who each stipulate specific requirements for assessments. The development of our assessment software platform has expanded enormously and, in the future, we also want to continue to devote significant attention to devising new (customised) technical solutions that better meet the widely diversified wishes of our clients.

In this publication, we share our vision of and enthusiasm for digital assessment. 'The power of digital assessment' is an update of the original 2014 publication entitled 'Digital Assessment. Help!'. This renewed version has been written from a practical perspective and, besides guidelines for organising the logistics of digital assessment and a checklist for making the transition from written to digital exams, it also describes in detail the quality requirements for assessments, the criteria that organisations can use when searching for suitable assessment software and the latest innovative developments in the field of digital assessment.

I am pleased to present this booklet to you so that we can share our knowledge. I hope you enjoy reading it!

Martin van den Berge
Oisterwijk, January 2020

Contents

1 Introduction

This booklet is intended for everyone who is considering starting up digital assessme nt or who wishes to 'vamp up' their digital assessment processes. Chapter 2 conceptualises digital assessment and sets out various reasons for conducting assessments digitally. Chapter 3 outlines the widely ranging choices that you can make when maki ng the transition from written to digital examinations. Le t it inspire to you to design digital assessment processes th at meet your and your organisation's requirements. Chapter 4 explains how you can pay attention to the quality of di gital assessments. Chapter 5 lists various factors that are i mportant in choosing a digital assessment system. It will probably help you to be able to decide what questions you should ask suppliers of assessment software systems. Lastly, we pay attention to innovation in assessment.

After reading this booklet, you should have a clear picture of what digital assessment entails, how it can be applied and what it can deliver in practice. You will notice that making assessments digital vastly improves their quality and, at the same time, significantly simplifies the processes involved. Especially in the long term, you will notice advantages in terms of efficiency. Use this booklet for inspiration and motivation to start working with digital assessments.

Why digital assessment?

This chapter explains what digital assessment precisely entails. We refute a number of misconceptions, give concrete insight into the practical advantages and conclude the chapter by outlining the added value of using digital methods to conduct different types of assessments.

assess

2.1 What is digital assessment?

Digital assessment is a procedure for taking an on-screen exam (on a computer or tablet) using purpose-built software. The exam questions are stored in a database in one location that is accessible via a secure connection. Digital assessment is a slightly broader term than online assessment because it does not necessarily require an internet connection. Digital assessment also involves more than digital examination because 'examination' refers to a formal form of testing linked to a certificate or qualification. This is not necessarily the case with assessments, which can of course be performed for a variety of reasons. For more information about the difference between assessment and examination, please see paragraph 2.5.

## 2.2 Reasons for making assessments digital

What makes examination boards and educational institutes want to switch from paper-based exams to digital assessments? The reason often stems from the wish to modernise exams in terms of content and hosting method, and to provide greater flexibility for when and where candidates can take exams. The obvious way to achieve this is to 'go digital'. Another frequently heard reason is the wish to simplify the marking procedure. For example, digital assessments make sending exam papers by post redundant and, with closed exam questions, eliminate the costs of marking all together. Lastly, e-learning is another common reason for digital assessments: the learning process is digital and interim progress tests form a component thereof.

The reasons for digital testing may be diverse, but so are people's concerns as to its effectiveness. The decision-making

regarding whether to adopt digital assessments and which system to use generally exposes all manner of misconceptions and false assumptions, the most common of which are listed and disproved below.

2.3 Misconceptions

There are various 'prejudices' against digital assessments. False assumptions can lead to (unjustified) concerns about the positive effect or added value of digital assessment. Perhaps you recognise them:
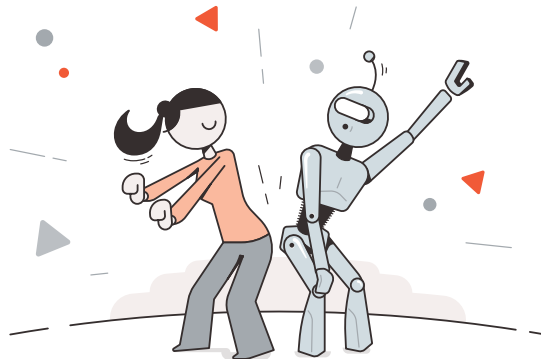
**○ Everything is marked by a computer, which means digital assessment lacks a human dimension;**

This assumption stems from the idea that a digital exam can only consist of multiple-choice questions. In any case, many different types of closed questions can also be used as described and explained in Chapter 3. There are also systems that have options for including open as well as closed questions in exams.

With open questions, candidates have to formulate their own answer. These questions can be marked 'in the old-fashioned way' by an assessor or teacher, or even by several if necessary.

● **With multiple-choice questions, everyone just guesses the answers, which makes exams far too easy;**

Open questions are in fact often easier than closed questions. This does, however, depend on the quality of the questions. Being able to 'guess' the answers is therefore certainly not a guarantee for obtaining a good grade or pass mark. Why not? Closed questions contain an element of probability, i.e. there is a chance that candidates will guess, but not actually know the right answer. 'Guesswork' can, however, usually be taken into account by adjusting the pass mark.
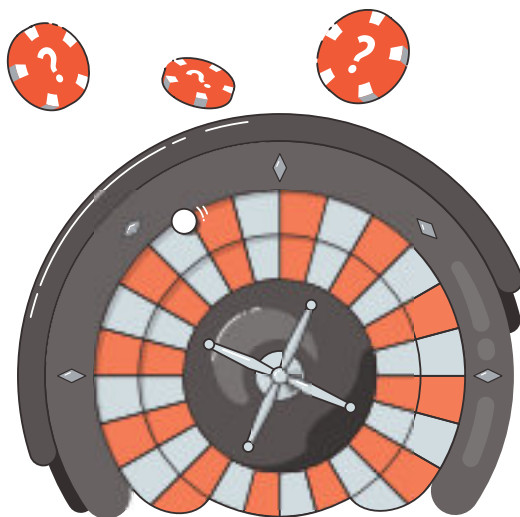
In addition, closed questions require the ability to interpret the question and the alternative answers accurately. Lastly, closed questions - provided they are asked correctly - require a thorough and, in some instances, a more abstract understanding of the subject matter.

The initial investment quickly pays off.

The chance of guessing the right answer to closed questions is also comparable with the chance of 'bluffing' open questions: even if you do not know precisely the right answer, you can still bluff and formulate an answer that is nearly correct. An assessor awards points to the formulated answer, even if it is not exactly the same as the intended one.

An example of the degree of difficulty of closed questions is an examination board that switched from digital assessment with multiple-choice and open questions to using only multiple-choice questions. What happened? The candidates had so much difficulty with the new multiple-choice questions that the passing percentage halved. It was then decided to use the old form of assessment – a combination of multiple-choice and open questions, as a result of which the passing percentage started to rise again. Research (Jordan & Mitchell, 2009) described on the weblog of Sally Jordan (2009) also confirms this).

● Digital assessment forces you to use a wide range of question forms;

Once you step over to digital assessment, you might think that candidates want a 'fun' and innovative exam complete with great graphics and videos. However, it is more important to set a proper exam, in which the types of questions are representative of the subject matter.

● Digital assessment is expensive;

Potential users of an online assessment system tend to see only the necessary initial investment costs. Setting up a database with test and exam questions is by no means inexpensive, but the same investment in a database has to be written off after conducting a small number of paper-based exams (see the assumption below about the size of the database). The content can only be stored, kept confidential and suitable for reuse with digital exams. The initial investment quickly pays off. In addition, you can also gradually build a database with test and exam questions in phases. More information about this is given in Chapter 2.

● Digital assessment imposes high demands on the available facilities;

It is a misconception to think that you need a large computer room to conduct digital assessments. The ideal thing about digital assessments is that they are not bound by time or place. Somewhere with an internet connection and (with summative assessments) adequate supervision in one form or another

(one-on-one invigilation, in-person proctoring at an exam location or remote supervision with a camera and telephone connection) is in fact all you need. Digital assessment gives candidates much more flexibility for taking exams and it requires only basic facilities. An internet connection is not even necessary if you opt for holding exams offline or a hybrid form of paper-based exams.

○ You need training before you can start digital assessments;

Taking a training course will of course help you to get the most out of the assessment software. But… you do not have to know every little detail. Define specific roles before you start training, for example, super user, data input specialist or exam supervisor. It makes a huge difference in the training time needed to implement digital assessment.

○ The larger the question bank, the better;

Your question bank must of course be representative for the subject matter you are examining and it has to contain sufficient 'versions' when you start digital assessment. However, within these guidelines, you should try to keep the question bank as small as possible.

In order to be able to use statistical data, for example, the percentage of candidates who give the right answer to a question, that question has to be asked very frequently. If the question bank is overfull, it can take a long time for the same question to come around again. In addition, maintaining a large question bank is labour-intensive and therefore also costly.

⚬ If there's no internet, exams will have to be stopped;

Some people are bound to say that if there's no internet, exams will have to be stopped. Good assessment software can, however, save all the answers candidates gave until the internet stopped working. Holding exams offline is another possibility. Incidentally, it is also always possible to generate a paper-based exam. This can of course be done in the event of an internet failure, but also in places where there is no internet connection at all, for example, in prisons. A digital exam can therefore also be converted into a written version.

⚬ With digital assessment, candidates cannot access details of their results;

After taking an exam, the entries submitted by candidates who wish to see details of their results can be retrieved from the database, including the given answers, the

correct answers and any feedback after the marking process. Exam data can in any case always be reproduced.

o Dyslexic or poor eyesight? Digital assessments cannot accommodate people with a disability.

Similarly to written exams, the amount of time allowed for digital assessments can be adjusted for each individual candidate. Extra time can be allowed, for example, for dyslexic candidates. A larger font can also be set for visually impaired candidates and, if necessary, the software can read the exam questions out loud.

Hopefully, we have been able to dispel any doubts about digital assessments. Something that we have not yet mentioned that is nevertheless a major (if not the main) advantage of digital assessment is that the quality of exams increases and is easier to improve. The quality of exams is discussed in more detail in Chapter 3. This and all the other advantages of digital assessment are explained in the following paragraph.

2.4 Advantages of digital assessment

This paragraph outlines a number of the advantages of digital assessment, viewed from the perspective of different user groups: candidates, educational institutes, businesses and examination boards.

From the perspective of candidates:

- Flexibility for taking exams in terms of time and location;
- No worries about (illegible handwriting);
- The results can be announced more quickly - marking usually takes less time and, with closed-question exams, candidates can be informed of the results immediately after they have finished;
- Under specific conditions such as remote supervision, candidates can also be allowed to take exams at their own home.

From the perspective of educational institutes, businesses and examination boards:

- Convenient storage of exam questions in a database structure in one location;
- Reusable exam questions;
- Insight into the quality of questions through fast and user-friendly analysis tools;
- Variation in the form of questions is possible (but not mandatory);
- Open questions can be incorporated in exams;
- Simplified marking procedure;
- Reduced costs for developing exam questions;
- Possibility for continually upgrading questions and improving their quality;
- The exam matrix evens out variations in difficulty level;
- Accessibility to and oversight of results;
- Secure examination method and contents;
- Insight into the quality of training and education programmes;
- Insight into the professional competence of teachers, markers and exam makers;
- The methods for administering and taking exams can be customised;
- Exam supervision can under specific conditions take place remotely

- Feedback can be given directly and easily after taking progress tests during the learning process.

2.5 Formative compared with summative assessments

One area in which digital assessment has distinctive added value is interim, non-official stude nt progress tests. This is usually referred to as formative assessment. Formative or diagnostic assessment focuses on testing acquired knowledge and has the aim of improving learning abilities. This type of assessment gives trainers and educators insight into shortcomings in courses and programmes that have to be addressed.

Formative assessment is often used solely for evaluation purposes and is commonly associated with the following types of tests and exams:

- entry level tests;
- diagnostic tests;
- preliminary knowledge tests
- mock exams;
- progress tests;
- self-tests.

The objectives of formative assessment are:

- to monitor students' progress;
- to facilitate student self-testing;
- to aid studying;

- to collect information to serve as input for learning-process design.

Summative assessments are intended to reflect the (final) learning outcome. Summative assessments have:

- … consequences: they prove that a candidate can or cannot do something. Summative assessments are usually selective. A good example in this respect is obtaining a driving licence. After passing the theory and practical tests, you are allowed to drive a car;

- … supervision. In this way, you can demonstrate that an exam is properly invigilated and you can prevent fraud as much as possible;;

- … a formal status. If you pass, you can obtain an official qualification or certificate;

- … (usually) nationally recognised. This can be the case if a training course is subsidised or if training/educational institutes hold exams in accordance with established (professional and/or sector-specific) standards and examination programmes. Students benefit because of the increased value of the particular qualification they want to obtain!

- … strict confidentiality. There's no point in taking an exam if its contents are known beforehand. What does it measure?

Digital assessment in practice at the Stichting Vakbekwaamheid Horeca (SVH) (Foundation for Professional Competence in Catering Services)

The SVH has conducted digital assessments alongside written exams since 2002. From 2014 onwards, the same techniques used for digital assessment have also been applied to written exams. This entails, among other things, that written exams and digital assessments are compiled from the same database. In addition, the marks for and results of written exams are digitally processed and analysed.

The advantages of using digital techniques with written exams

There are three main advantages that improve the quality of exams and the service level for candidates:

1. (Quality.) In this way, each exam is unique and does not have to be compiled manually. The system does this automatically on the basis of the exam matrix. Because each exam is unique, the risk of fraud decreases, and, because no student is given the same exam, the chance of them knowing the answers to questions in advance is much smaller.

2. (Quality.) SVH uses multiple-choice questions in its digital assessments and the written equivalents thereof. After candidates have taken a written exam, their answers are scanned. In this way, another advantage of digital assessment can be exploited: the assessment software is used to analyse the questions, which means that there is continuous quality control.

3. (Service.) There is no need to know who will turn up to take an exam in advance, which gives candidates flexibility. Even if unregistered candidates want to take an exam, they are able to do so because additional written copies are always available. Candidate's names only have to be linked to their results after completing the exam.

Combination of written and digital exams

Because the SVH holds both written and digital exams, the organisation has a large amount of flexibility with regard to when and where exams can be taken by candidates. Digital exams require facilities, such as laptop computers and a reliable internet connection, that are not needed when conducting written exams. The combined use of written and digital exams means locations can be spread further apart, which also improves the service level given to candidates and trainers.

Consistent exams and exam location quality

Because the exams are generated by a matrix, the content is weighted and, although each exam is unique, there are no differences in difficulty level (between written or digital versions). The use of different locations for conducting exams is also subject to quality standards. The quality of the written exams is also monitored by freelance invigilators and administrators. Both forms of exam are also externally supervised to ensure that set procedures are followed.

The advantages and a key focus point of digital assessment

Taking an exam on a computer is a pleasant experience for candidates: it avoids writer's cramp when answering open questions, improves legibility for markers and candidates can also easily change their answers during the exam. Lastly, the answers are stored centrally so they are always available. Human error is also automatically prevented in processing the answers. Information security is a key focus point when conducting digital assessments. Its importance should not be underestimated and the SVH gives it the highest priority.

Making the transition from written to digital exams

- In this chapter, we illustrate how processes will change after making the transition to digital exams. We also outline the steps towards implementing digital exams and the decisions that have to be taken. An important step, in which digital exams differ the most from paper-based equivalents, is the method for setting up question banks. Lastly, we look more closely at the possibility of introducing digital exams in phases.

digital

3.1 Digital exam logistics

We define exam logistics as the entire process from preparing and making an exam ready to be taken by candidates to announcing the results and analysing the exam questions. In this paragraph, we show what digital exam logistics entails. To do this, we separate the process into three stages: before, during and after the exam.

Before the exam

A number of factors related to written and digital exams have to be determined beforehand. The main difference is that these factors in principle only have to be configured once with digital exams. In this respect, the main choices that have to be made are as follows:

- exam design;
- exam compilation;
- number of questions;
- the pass mark;
- exam format;
- enrolment or registration method;
- choice between fixed or flexible moment.

Exam design
To design an exam, you use an exam matrix - an overview of how the subject matter is divided over the questions candidates have to answer. The matrix is usually linked to attainment targets, teaching objectives or examination objectives. It indicates which questions cover each teaching objective and how they are weighted. The matrix ensures that all exams have the same structure. This is no different with digital exams. You use the matrix to design the

complete exam and structure the question bank (the database). The proverb a good start is half the battle won is also true for digital exams.

Questions can be reused and maintenance is easier

Exam compilation: different versions or questions selected at random

The great thing about digital exams is that you can configure the software to generate an exam automatically and let the system pick the questions. This is called an at random exam. You do, however, still have to tell the system how to 'prepare' the exam. With at random exams, the questions are placed separately in the database according to a specific structure that corresponds, for example, with the examination objectives. It is therefore not the case that 40 questions are simply drawn at random from a depository of 200. The distribution of subjects in an exam should follow the predetermined design. The main advantages are that questions can be automatically reused and are easier to maintain. After initially opening the question banks, at random exams enable cost reductions and efficiency improvements.

If you find this step too difficult, you can opt to set up fixed (versions of) exams. In this case, the database contains entire exams consisting of questions, exactly the same as written exams. The system picks

a complete exam for a specific student. This variant also clearly reduces costs compared with written exams: because the exam content remains confidential, the questions can also be reused to compile new exams.

An additional option with at random exams is to place questions that belong to different exams in the same question bank (database) This is done when there is an overlap between questions in different exams. An example of two overlapping exams: an exam on species of birds found in the Netherlands and one on birds that migrate to and from the Netherlands. Setting exams with slightly divergent contents for different target groups is another option. An example of this is setting two exams on migratory birds in the Netherlands, one for first-year students and the other for second-year students. Using these options also improves efficiency. If you want to include multiple exams for various target groups in the database, you can label the exam questions so that the assessment software only selects questions for a particular exam or target group.

Paragraphs 2.2 and 2.3 contain more information about at random compared to fixed-version exams.

The number of questions
We are often asked what is the minimum number of questions that a question bank should contain. This depends on a number of factors:

- Questions tend to 'wear out'; is the system security adequate and what is the usage frequency? After questions x-have been placed in exams a number of times, candidates will become familiar with them (and the respective answers) and they can no longer be used effectively. The larger the question bank, the smaller the chance of questions wearing out.

- Standards have to be applied to have sufficient exam questions, also to cover resits and to remove poorly formulated questions from the database. With a question bank that has been recently initiated, you can assume it can be used to draw an exam three times. With fixed-version exams, you should allow a safety margin and be able to draw an exam four times. If you have an exam containing 30 questions, you will therefore need a minimum of 90 to 120 questions in the question bank to start with. For more information about starting digital assessment with limited-size database, see paragraph 2.3.

- Standards have been established for setting the right number of questions to cover the scope of the subject matter. With extensive subject matter, all the relevant aspects have to be included in the database. This means it has to be possible to ask questions about all the relevant aspects. Candidates must not be able to assume that questions about some topics will never be asked.

- Topics that should always be included in an exam. The content of an exam must of course be representative of the subject matter. If the questions only cover a small fraction of the subject matter, the exam will lose its validity. The more extensive the subject matter, the more questions an exam needs to have. The larger the exam, the more questions the database needs to have.

- The minimum number of measuring points an exam should contain. A general guideline is to have a minimum of 30 measuring points (items that can be used to grade candidates). However, this number should be matched to the probability of guessing the right answer and increased if necessary. If an exam contains, for example, a number of yes/no questions (with a 50% chance of guessing the right answer), the number of measuring points should also be higher. And the more measuring points, the more questions an exam needs to have. If an exam consists of, for example, open questions worth 2 points each, it must contain a minimum of 15 questions. An item bank with only

open questions worth 2 points each should therefore contain at least 45 to 60 questions. The standard for the number of questions the database should contain for closed-question exams is much higher.

Pass mark
You always have to specify a mark that signifies the difference between passing or failing an exam. Sometimes you do this in advance, sometimes afterwards. Most digital assessment software systems require the pass mark to be set in advance. A pass mark can be set in many ways. An easy one is 'half the answers must be correct'. Another method is to allow for the probability of candidates guessing the right answer even if they do not know it. For example, with multiple-choice questions with four alternatives, the probability of guessing the right answer is 25%. If a candidate guesses each question, there is a chance that 25% of their answers will be correct. If you add the minimum that a candidate is required to know, which is usually 50%, the pass mark is: 50 + 25 = 75%. In most situations, pass marks are calculated using formulas based on the number of questions, the number of alternative answers, or the largest common denominator of the two.

Exam layout
Anyone who constructs exams knows that making the layout is time-consuming and labour-intensive, and that it has to be done precisely. You want your exam to be representative and candidates not to be distracted by errors or irregularities in the layout. With digital assessment a layout is required for each question and for the exam as a whole. You can use all the features and functionalities of word processing software for the layout of exam questions. Most assessment software systems include an editor for formatting your exam questions.

The advantage of digital exams is that you only have to make the layout settings for all your exams once. You can of course give each exam a different layout if you want to. In this wat, you can always create exams with your own 'look and feel' and with much less effort than with written exams.

Registration

Digital and written exams have similar procedures for registering candidates. In this respect, it is important to note that assessment software sometimes can be expanded with modules for registering candidates and paying fees. The registration procedure for digital exams is ultra-efficient. Candidates can register individually online and select an available time slot. It is also possible to register candidates in groups, which can be useful if an exam takes place at a set time. Read more about this below.

Fixed or flexible moment

Before starting to build an exam, you have to decide whether it should be taken at a fixed moment or whether it can be held flexibly. A fixed moment is chosen if all the candidates are intended to take the exam simultaneously. One reason for doing this is if you want to hold an exam at a location where it is not possible to receive a continuous flow of candidates.
Don't be confused: holding a fixed-moment exam does not necessarily mean having to use fixed-version exam in your question bank! You can set fixed moment exams and configure the question bank to generate at random exams or vice versa: flexible time with fixed-version exams in the question bank. A set time with fixed-versions in the question bank most closely resembles a traditional written exam.

Holding exams

Summative assessment takes place under controlled circumstances, which means, among oth er things, that it is supervised. Supervision can be arranged for your own or a rented exam location. Via online p roctoring, 'remote supervision' is also possible. Onl ine proctoring enables each individual candidate to take an exam at their preferred location, for example, at their home or place of work in a familiar environment, usually at a p re-arranged time. The examination board or organisation does not have to pay the costs for the exam location. The exam can be supervised remotely by a proctor live or afte r it has been completed.

Digital assessment platforms usuall y have a range of tools to facilitate taking exams, such as calculators and panels for automatically displaying 'time remaining' or the number of questions left to answer and options for allowing additional time (in the event of an emergency). Various settings can also be made for special target groups, such as dyslexics and the visually impaired, for example, allowing more time, a larger screen display or voice support.

After the exam

Digital assessment is typified by its possibilities for marking closed questions automaticall y, providing feedback and informing candidates of their resul ts. You personally decide what you want to automate.

The following elements of the post-exam process are briefly discussed below:

- Marking;
- Results and analysis;
- Feedback on summative assessments;
- Feedback on formative assessments.

Marking

Digital assessment simplifies the marking procedure irrespective of whether you use closed questions that are marked automatically or open questions that need to be marked in person. In both cases, you can be sure that the correct answer model will always be used because the system only displays the most recent version. With open questions you can also assign one or more correctors to different sections of candidate's exam entries. It this situation, the system will automatically send an email to the respective marker to inform them, for example, that exam entries are waiting to be checked. The marker corrects digitally and their assessments are stored in the system so that they can always be retrieved and reviewed.

Results and analysis

With exams with questions that can be automatically marked, the results are immediately available to candidates. Candidates know straight away whether or not they have passed. Digital assessment software often has a functionality for generating provisional results. You can also use it to generate reports about the level of difficulty and the quality of the questions. In addition, analyses can be made per trainer/marker to gain insight into any differences. The analyses tell you whether and how your exam can be improved. This information can of course also be generated for written exams, but it involves a relatively large amount of work because the questions are only used once and analyses cannot be automated.

Feedback on summative assessments

Sometimes it is desirable to give feedback on summative assessments, for example, if it is not possible for candidates to review their exam entries. In this case, candidates benefit from feedback on the level of attainment targets. Candidates can be advised and informed about the possibilities of resitting an exam without giving them insight into the questions.

Feedback on formative assessments

Interim progress tests are formative and do not count towards a student's final grade. Be cause students are intended to learn from progress tests, they have to be given feedback. Di gital assessment simplifies the p rocedure for providing feedback.

After making the transition from written to digital exams, setting up a formative assessment is a piece of cake. Good quality feedback is, however, essential. Progress tests make candidates aware of what they do and do not know, and challenge them to study harder.

Formative assessments make it much clearer to candidates where they stand. Taking tests should be a pleasant and effective learning method. An additional advantage: the performance of candidates who suffer from a fear of failure starts to improve if they regularly obtain good marks in formative assessments. The feedback they receive acts as a positive endorsement and they no longer have to be afraid of failing an exam because they realise they have mastered the subject matter.

Four tips for giving feedback on formative digital assessments:

1. Decide in advance on what level you want to give feedback. Feedback can be given on different levels: attainment targets, learning strands, exam questions or answers. The various possibilities may be limited by the assessment software you use. Hattie & Timperley (2007) indicate that feedback can be given on different levels, including right/wrong answers as well as initiating reflection on a candidate's thinking process;

2. Decide how you want to present feedback. When feedback is given on the level of answer alternatives, there is a risk that it will be negatively formulated. Try to be constructive when giving feedback;

3. Decide when you want to give feedback:
   - Immediately after entering the answer to a question or completing an exercise;
   - Immediately after finishing an exam or;;
   - After candidates' marks have been finalised (when discussing or reviewing the answers);

   The possibilities for giving feedback also depend on the assessment system. Sometimes candidates can immediately change their answers (given a second chance). if they are allowed to do this, immediate feedback is required. Recent research suggests that delayed feedback may be more effective than immediate feedback given during an exercise (Van der Kleij, 2017).

4. Try to make the feedback you give as low-maintenance as possible.

   For example, do not refer to a page in a book if a revised edition that book is published every year. If you do this, you will have to update your feedback annually. It is better to refer to the subject that warrants closer study.

3.2 Steps towards implementing digital assessment

On the basis of the above information, you can already decide which steps you need to take to make the transition from written to digital assessments. In this paragraph, these step s, or the choices you have to make, are explained. A practical checklist has been appended at the end of this booklet that you can use as an aid when making the transition from paper-based to digital exams.

It is assumed that the following information is available:
- Teaching objectives, attainment targets and/or exam objectives;
- a global exam matrix showing the distribution of the subject matter over the questions and the weighting thereof;
- agreed starting point for calculating the pass mark.

Important aspects that may need further explanation before you can start working with the checklist are:
- The method for setting up question banks;
- At random compared with fixed-versions;
- The exam question design process.

Setting up question banks

Question banks have to be set up specifically for and are an important aspect of digital assessments. A large part of the checklist therefore focuses on question banks. We assume that is must be possible to conduct exams flexibly, i.e. at different times. Why is this assumption relevant for the checklist? Actually, only for one reason: Flexibility imposes requirements on exam parity. A candidate who takes an exam a week earlier or later may not be given a completely different set of questions to

other people. In paragraph 2.3 on phasing the implementation of digital assessment, flexible and fixed-moment exams are compared in more detail.

At random compared with fixed-version exams

The first section of the checklist covers the points that you usually decide when setting up an item bank. A distinction is made between at random and fixed-version exams. You can see from the checklist that using fixed-version exams requires more manual control, but you do not have to make as many decisions about the structure of the question bank in advance. For more information about the pros and cons of at random and fixed-version exams, see the table at the end of this paragraph.

The structure of a question bank for fixed-version exams only consists of complete sets of questions (named, for example, exam versions 1 to 10). The structure of the question bank for at random exams is developed separately and is shown in a question tree.
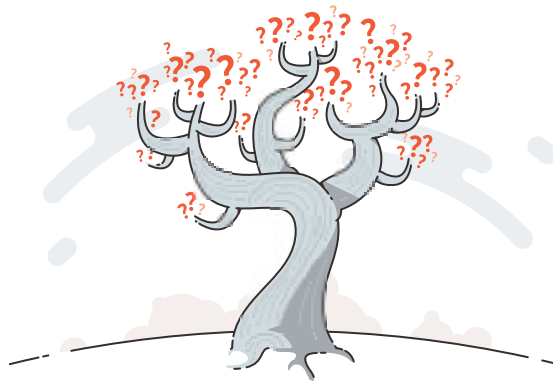
A question tree is a detailed representation of the structure of the question bank based on the exam matrix. A question tree contains information about:

- the minimum and maximum number of questions to be asked per subject;
- the number of points to be awarded per question (or the point range per question or set of questions);
- the proficiency level (understanding of concepts; ability to apply knowledge, see Bloom, 1956) required to answer a question;
- whether or not an exam contains illustrative case studies

You can see from the checklist (see Appendix 1) that the exam matrix and the question tree converge. The difference between an exam matrix and a question tree lies in how they are used and the level of detail. An exam matrix is widely used with a variety of different forms of assessment, including written and digital exams, and it indicates the exam structure. A question tree is used with the digital assessment of theoretical knowledge and it indicates the database structure.

The exam question design process

Lastly, we would like to point out that at random question banks and fixed-version question banks have different exam question design processes. With fixed-version exams, questions are generally designed per exam; in other words, one or two people jointly author a complete exam. With at random exams, different authors design sets of questions for the question bank. This is a significant advantage if the design work has to be completed before a specific deadline; many hands after all make light work.

At the same time, following this design process provides a greater safeguard. If questions are rejected for one reason or another, the entire exam does not have to be scrapped. The design process is not included i n the checklist. You can take a wide variety of steps that are different to the steps for designing written exams, but they are also organisation and project related.

The advantages of at random exams over fixed-version exams are summarised below:

At random	Fixed versions
✓ Easy automated exam preparation.	o Manual exam compilation
✓ Clear database structure.	o No clear database structure.
✓ Exams automatically follow the exam matrix.	o Whether exams follow the exam matrix checked manually.
✓ Exam construction is more flexible; possibility of assigning separate components to different exam builders.	o One or two exam builders work on a single-version exam. If it is discovered that an exam does not meet the requirements just before it is taken, a quality problem may arise
✓ All questions occur as a single instance in the database, which makes maintenance efficient.	o Questions can have multiple instances in the database, which makes maintenance more inefficient.

✓ Fewer questions needed at the start, for example, for drawing an exam three times.	○ More questions needed at the start, for example, for four-version exams.
✓ Questions are less recognisable or less easy to remember, because they are automatically shuffled.	○ Questions are more recognisable and easier to remember because they are presented in a fixed-version (having a sufficient number of versions eliminates this disadvantage).
✓ Feedback on the quality of exam objectives is possible: because the question bank is (partly) arranged according to and developed per exam objective, their quality can easily be verified.	○ Because the question bank is arranged according to exams and not exam objectives and also not developed per exam objective, the quality of the exam objectives cannot be directly verified.
○ When making the transition to digital assessment, the question bank structure does not need to be set in advance. This saves time when starting up digital assessments. In addition, because the question bank structure is not set this method gives exam builders more freedom in how they approach their work.	○ When starting digital assessment, it is not necessary to pay attention first to making choices about the arrangement of the question bank.

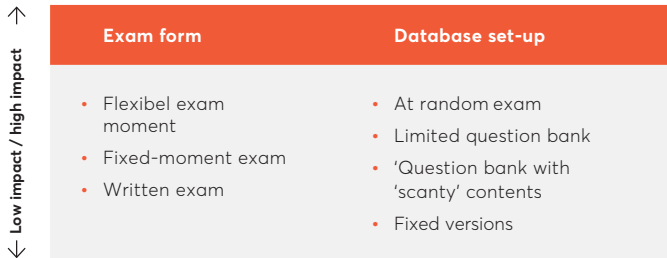*Figure: At random compared with fixed-version exams*

For more information about the design choices for question banks, see the checklist at the end of this booklet.

3.3 Phased digital assessment implementation

How do you ensure a seamless transition from paper to digital exams? You can regard the transition as a project, you can completely change the work processes or you can start off 'small'. How does starting digital assessment in phases work?

Several of the methods for starting off 'small' and limiting the (initial) amount of work and time needed to 'go digital' are discussed below:

Impact of digital assessment on existing processes

Exam form	Database set-up
• Flexibel exam moment • Fixed-moment exam • Written exam	• At random exam • Limited question bank • 'Question bank with 'scanty' contents • Fixed versions

(vertical axis label: Low impact / high impact)

Figure: schematic representation of the impact of digital assessment

Limited question bank

An easy way to keep things 'small' is to start off with a limited question bank. A distinction is made between at random and fixed-version exams. In a later stage when you wish to proceed to the next step, you can simply expand the question bank.

With at random exams

The starting point can, for example, be being able to draw an exam three times with different sets of questions. Why three times? If an exam is 'drawn three times', candidates can take the exam and two resits, and have to answer a set of completely new questions each time. If candidates are allowed to take a third resit, they might get questions that they have seen before, but the order in which the questions are presented can be switched around and 'shuffled'. The advantage of this is that the question recognisability, even with a third resit, remains low.

With fixed-version exams

The starting point can, for example, be four exams - one more than with at random exams. Why do you have to increase the standard number of exam questions in the database? Because there might be something wrong with one of the exams, as a result of which that particular exam cannot be used temporarily. Furthermore, candidates are presented with completely unique exams even if they take a third resit.

Question bank with 'scanty' contents

If you are reluctant to use a wide range of different types of questions when starting up digital assessment, you can of course decide to do this in a later stage without causing any problems. In this case, you should only use the same types of questions that you also set in written exams, for example, open questions or a combination of open and multiple-choice questions.

Fixed-versions

With fixed-version exams, the switch to digital assessment is relatively small. The reasons for this are that you do not have to finalise the database structure yet and you can leave the development process unchanged. See the checklist in Appendix 1. The procedure for fixed-version exams is actually similar to working with written exams.

Fixed moment

If you want to make the transition from written to digital even smaller, you can opt to hold exams at a fixed moment. Working in this way simplifies announcing and analysing the results in particular because the related procedures can be left unchanged. If you wish to change the grades on the basis of candidates' performance, you can do so, similarly to written exams, in groups. With flexible exams, you can also change the grades, but usually only after candidates have already received them. Don't be confused: Holding fixed-moment exams is not the same as using fixed-version exams. You can also hold fixed-moment exams using an at random question bank.

Printed exams

If you want your exam to resembl e a written variant more closely, you can allow candidates to take a (partly) paper-based version by printing copies of digital exams. The main advantage of this is that you can use the locations as you would for written exams. After al l, there is no need for computers or a stable internet connection. This enables you

to benefit from the advantages of digital assessment without having to change the processes i nvolved in conducting the exam. An additional advantage is that candidates' answers in a paper-based exam can also be digitally processed and analysed.

How do you ensure a smooth transition from paper-based to digital exams?

Pilot projects

In the transition to digital assessment, it is quite conceivable that new questions will be developed, or that a completely new exam will be constructed (or that the exam's 'look and feel' will in any case be perceived by candidates as new). How do you ensure a smooth transition? The following types of pilot projects can help ensure a smooth transition:

- pilot project by selecting fixed-moment exams;
- pilot project with trial period (flexible-moment exams);
- pilot project prior to opening the question bank.

Pilot project by selecting fixed-moment exams

How do you set up this type of pilot project? It can be done in different ways. The first and easiest way is to start with digital fixed-moment exams. You organise, for example, two or three fixed moments for exams. After each moment, you carry out a question analysis and, if necessary, adjust the grading. In this way, you directly test the question quality and alleviate any anxiety among candidates and other stakeholders.

Pilot project with trial period

Another form of pilot project, which is used with flexible-moment exams, is to set up a trial period of, for example, three months. You start the trial period with a low pass mark. You also communicate this to candidates. Then you evaluate the passing percentages and, if possible, increase the pass mark gradually in steps. This does, however, require the necessary tact and transparency to quell any unrest!.

Pilot project prior to opening the question bank

If you want to know for sure how well your questions will be answered, you can simply invite candidates to take exams. In this case, clear communication is required. What is their status if they pass? Will exam fees be refunded to candidates who do not pass? Etc.

The quality of examination questions

- The actual content of the question bank - the exam questions - is vitally important for the successful implementation of digital assessment. For this reason, this chapter discusses various quality criteria, the analysis of exam questions and types of questions and related grading possibilities.

qualit

4.1 Criteria for the quality of examination question

When is the quality of an exam good? The terms validity and reliability are used when determining exam quality. These terms cover a whole world of meaning, but here we will limit ourselves to their largest common denominator. For more information about the terms reliability and validity, please see the explanation given by Sluiter, Hemker and Eggen (2018). An exam is considered reliable if the outcome is not coincidental, which is the case, for example, if disruptive factors affect the way the exam is held or if the questions and the subsequent evaluation thereof contain errors. A valid exam measures what it is supposed to and is consistent with the purpose for which the outcome will be used. An example of the application of this criterion is exam questions aligned to specifically formulated teaching or exam objectives and the difficulty level. Complaints about validity of an exam are often related to how stakeholders interpret its results.

In short: a good quality exam means that a candidate who actually has mastered the subject matter passes and stakeholders believe that the exam is fair. Exam questions that meet quality criteria generally produce the most satisfactory results and meeting these criteria is usually relatively easy to achieve. Awareness of these criteria and the ability to apply them is therefore important. The technical exam criteria used involve:

1. Relevance;
2. Objectivity;
3. Specificity;
4. Efficiency.

1. Relevance

Does the question match the teaching and other objectives in the examination programme? Is it also about relevant knowledge or does involve detailed information that either nobody can possibly know or nobody will ever use? The question must be about the subject matter that is useful to a professional practitioner. An example: In an exam about product knowledge in the food industry, knowing the nutritional value of peanut butter by heart might not be considered important because this information can be found on the label of a jar.

2. Objectivity

Are answers to questions right al l the time, or are there also situations conceivable in which the 'right' answer is actually incorrect? Can other answers also be counted as correct? An objective question generally does not lead to any discussion. See the example of a non-objective question below.

> **What colours are in the Dutch flag? Tick all the correct answers.**
>
> A. Red
> B. Blue
> C. White
> D. Orange
>
> The right answer is: A, B and C

However, should answer D, orange, also be counted as correct? The Dutch flag should officially have an orange pennant. Answer D is probably the least relevant,

but it is not really wrong. In any case, this question could lead to discussion.

3. Specificity

A question should be asked in suc h a way that someone who has mastered the subject matter should be able to answer it correctly and someone who has not should not be able to do so. A specific question therefore distinguishes good and bad candidates. See the example below of a non-specific (open) question designed to be answered by adults.

Describe the leadership styles in commonly used management theory.

Answer:
The theory of Hersey and Blanchard outlines four leadership styles.

Description:
- Delegation: leaving tasks to employees, little control and little support;
- Support, consultation: helping employees, little control;
- Persuasion, motivation: large amount of task-oriented control and support;
- Order, instruction: large amount of control, little support.

Other answers can judged right or wrong at the discretion of the corrector.

The problem with this question is that it is not very focused; there are other management theories and models that are often applied in practice. In addition, it does not stipulate the requirements for a right answer, numerous answers could be counted as correct.

4. Efficiency

In order to meet the efficiency c riterion, limiting the information in a question to what is necessary for answering it is important. An example of this that we often see is a case study text that includes a long newspaper article as background information. It would be better not to include this type of information in an exam, but in the course material instead. Another example is a question with a double meaning that a candidate has to interpret correctly. Emphasise important information by using bold or italic text so that it draws the attention of candidates. The 'efficiency' criterion also covers language errors and complex sentence structures.

4.2 Analysis of exam questions

Quality criteria are used in exam construction and analytical values are used after an exam has been taken. In this section, we briefly illustrate some of the values, including their meaning, which assessment software automatically generates:

- p-value;
- a-value;
- rit-value.

The p-value is the percentage of candidates who answer a question correctly.

In other words, a high p-value indicates a relatively easy question. Almost everyone in the group of candidates has the right answer. An exam should not contain questions that are too difficult or too easy. Questions that are too easy fail to meet the exam objectives and distract good candidates, who will wonder if the exam should be so easy. A p-value from 0.3 to 0.8 is optimal for each item in the question bank.

The a-value indicates the percentage of candidates who select a diversionary answer.

The aim of diversionary answers is to give candidates who have not mastered the subject matter a plausible alternative. By analysing the diversionary answers, you can usually quickly see that multiple-choice questions only need to have three alternative answers. If a question has more than two diversionary answers (two diversionary plus the right answer makes three alternatives), you will often see that some of the diversionary answers are never selected. If this happens, you should consider scrapping alternatives that are never selected.

The rit-value indicates how accurately each individual item measures the same factors as the exam as a whole. In other words, how well each item fits in an exam.

The rit-value shows how an item segregates good candidates from poor candidates. The rit-value indicates the distinguishing capacity of an item.

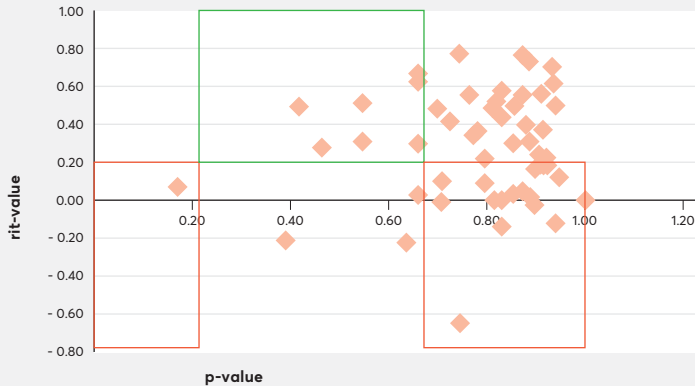**Combined question analysis (p and rit values)**

Figure: Combined question analysis

If the p-value and the rit-value are plotted against each other in a graph, you can easily see which items have to be critically assessed. The items in the red squares must in any case be examined. It is worth noting that this exam appears to contain a large number of easy questions (items with a high p-value). The standard for an acceptable rit-value varies. The following standards are generally adhered to in the literature (Veldhuijzen, Goldebeld & Sanders, 1993):

0.40 and higher very good
0.30 – 0.39 good
0.20 – 0.29 uncertain
0.19 and lower poor

4.3 Question types

Many types of questions can be used i n digital assessment. Do not be enticed by an assessment software supplier who offers a hundred different types of questions. The chance that you will use all them is small. With only eight types of questions, you can create a large amount of variation. A number of the types of questions that can be used in digital assessment are outlined below.

Questions that have to be checked or included in the marking logistics (if used):

1. **Open questions**
   Candidates have to formulate their answer in the answer field. Depending on the choices made by the exam builder, an editor is available to candidates.

2. **Upload questions**
   The candidates must enter their answers in a file attachment that they save and upload. Sometimes the software can connect to external applications such as Excel or another application. This feature is mainly used with practical exams. Not all assessment software, however, includes this functionality.

Questions that are marked automatically:

3. **Multiple-choice questions**
   One answer has to be selected from a set of alternatives.

4. **Multiple-select questions**
   Several answers can be selected from a set of alternatives.

5. **Hotspot questions (singular)**
An answer can be entered by clicking on a location in a figure or image. An arrow (marker) is put in the respective place. The answer can be subsequently changed by dragging the arrow (marker) to a new location with a mouse. Attention: this is not a multiple-choice question, unless the figure or image contains predefined, visible hotspots. With modern techniques, this is, however, not strictly required and you can define randomly occurring hotspots in the figure or image.

6. **Hotspot questions (multiple)**
In this case, several answers can be entered by clicking on the appropriate locations in a figure or image. The answers can be subsequently changed, if necessary, by dragging the arrow (marker) to a new location with a mouse.

7. **Video hotspot questions**
This type of question contains a video that has to be clicked at the right moment. It can, for example, be used if the aim of a question is to test whether a candidate can correctly indicate a specific moment in a conversation. The answer can subsequently be changed by clicking on the original cross and then on a new point in the film.

8. **Classification (ranking) questions**
The answer options (text or pictures) have to be placed in the correct order.

9. **Drag and drop questions**
Objects (text or pictures) have to be dragged to and dropped in the right places in a figure or image.

10. **Matching/matrix questions**
    This type of question consists of a number or rows and columns. Candidates have to place crosses in the right row/column. This type of question is suitable for testing knowledge that involves establishing relationships between two variables. For example, which shrubs and flowers belong to which type of vegetation?

The following types of questions have a pre-defined correct answer and can be marked automatically:

11. **Numerical questions**
    An answer can be entered in the form of a number. Sometimes it is also possible to clone answers by using variables. The sum (in the question) has different representations, but the same parameters.

12. **Fill-in questions (singular)**
    An answer can be entered in the input field in the form of a short text or a single word. In the underlying correction model for singular fill-in questions, several answers can be correct. In this way, common spelling mistakes can still be marked as correct automatically.

13. **Fill-in questions (multiple)**
    Multiple answers - text and numerical - can be entered in the provided input fields.

With modern techniques, questions can include forms, videos and audio files.

Tip: Media that you use for teaching can be reused in an exam. You must, however, remember to issue headphones to candidates when the exam is conducted.

Tip: Think about copyrights when using media. If you are unsure whether you can use media, ask the creator for permission.

4.4 Grading different types of questions

After choosing which questions are going to be used in an exam, you have to decide how to grade them. In other words, how are you going to rate the answers? With multiple-choice questions, grading is fairly simple: the answers are right or wrong. Multiple-choice questions are also called dichotomous. With polytomous grading, the answer to a question can, however, be partly right. With a multiple-select question, a candidate might only choose one of the two correct alternatives. Do you award the answer full points, half a point or no points?

Grading different types of questions:

- **Dichotomous**: a candidate is only given the (maximum number of) points if they completely answer the question correctly. The candidate is not awarded any points for a wrong answer.
- **Polytomous**: a candidate is awarded a proportion of the total number of points per correct answer to the whole question. Selecting a diversionary (or wrong) answer can lead to point deduction. The lower limit of the question is zero points.

The grading is calculated as follows:

> (Percentage of correct answers chosen – percentage of diversionary answers chosen) * maximum points to be awarded for the question

A common form of polytomous grading is the product rule. The product rule takes account of the possibility that a candidate ticks all the alternative answers and obtains some of the points. What makes the product rule unique is that a candidate's point score for a question decreases if they erroneously tick a wrong answer as being 'right'.

The quotient rule is usually used for grading classification and drag and drop questions. The point score is calculated according to a built-in formula based on the number of possible ways of ordering a given number of objects:

> (If total number of objects is 4 - A, B, C and D, they can be arranged in 6 ways: A before B, A before C, A before D, B before C, B before D, C before D).

Candidates have to get two thirds of the answers right to obtain half the number of points.

In digital assessment, the applicable grading method is usually selected before designing exams. The most important decision is whether points may be awarded for different parts of a question; in other words, whether the grading should be polytomous. This requires some explanation for the stakeholders concerned, but it is a fair assessment method because candidates receive (some) points for all correct answers or aspects thereof. Digital assessment is sometimes obstructed because the way digital exams are marked is allegedly unfair - answers can only be

right or wrong. From the diversity of the above grading rules, you can see that this is unjustified. In fact, digital assessment makes grading exams more objective because it compels item authors to think carefully about using the correct grading rule for a particular question (Examinations, No. 3, 2010). Attention: Not all assessment software is features extensive grading rules.

5

The choice of assessment software

- Choosing assessment software is difficult. There is a large number of software packages on the market with widely ranging prices and predicting the future requirements it has to meet is not always easy.

choic

The choice of assessment software is a long-term decision. You therefore also have to consider the specific requirements that the package has to meet at the present time as well as in the future. The number of candidates or the number of users of exam locations may increase as a result of internal reorganisation. For example, in a large internationally operating organisation, exams are currently organised to take place in a number of locations at the same time. In the future, the organisation intends to make exams independent of time and location, which imposes requirements on the assessment software. This is just one example, but there are often many uncertainties.

To help you consider them, a number of focus points are listed below:

1. Technology
2. Security and privacy (GDPR)
3. Roles and rights
4. Modular structure
5. Registration and fees
6. User-friendliness
7. Open as well as or only closed questions?
8. Availability of statistical data
9. Multi-media applications
10. Links to other software
11. Options for taking exams
12. Customisation

1. Technology

Will multiple people spread over different locations use the assessment software? Remember to include exam builders and question revisors. This calls for a web-based system with questions stored

in a central and properly secure d location. In addition, asking questions about the security of the system is important. Is it properly arranged? Security is a critical issue especially with summative assessment. The security for systems that are only suitable for formative assessment is not always optimal. Lastly, it is important to study how the software behaves when a large number of candidates take an exam simultaneously. Are the servers suitable for the number of candidates?

2. Security and privacy (GDPR)

To what extent is the system eq uipped to meet the requirements of the GDPR? The security, storage and use of personal data and the possibility of deleting it are important.

3. Roles and rights

Setting up exam processes usually involves assigning different responsibilities and authorities. To what extent can roles and rights be incorporated in your digital assessment software. Who is ultimately responsible for (or the owner of) the question bank? Who is allowed to enter, check and revise questions? Who is allowed to edit the templates the system uses and adjust the difficulty level of an entire exam, etc.? When looking for suitable assessment software, it is important to know what your requirements are and to find out the extent to which the system meets them.

4. Modular structure

How is the packet structured and is it also suitable for modular use? For example, does the package include modules for holding exams on location, for supervision, for enrolment and registration, for analysing data and for generating reports. Is it possible to use a 'light' version of the software, for example, for formative assessments?

5. Registration and fees

How easy is it for candidates to book a place for an exam, change their registration details and to pay the fee? Are there possibilities for creating a link between the supplier's website and the respective assessment software? And how is communication with candidates arranged? Is there a mail module and how easy can it be customised?

As a user, you will benefit from a clearly organised structure.

6. User-friendliness and quality

How easy is it to enter questions? Can it be done by relatively untrained people? It can be particularly

relevant to make to assessment software accessible exam builders and question revisors, and to assign various user groups different data input, editing and sign-off rights. In addition, it can also be important to make the status of questions (signed off, under review, etc.) visible in the system.

Another important point of attention is to look at the possibilities the system has for analysing results. Does it already contain formulas for calculating the p and rit values, also for non-multiple-choice questions? Some systems require user input in this area, while others have this option built in as standard.

Lastly, as a user, you will benefit from a clearly organised structure. How are the questions in the database arranged? Can you immediately see the structure, or does it look more like a jigsaw puzzle?

7. Open as well as or only closed questions?

What is the best way of examining the subject matter - with open or closed questions? Open questions are, for example, needed if you have to examine whether a candidate can formulate their own answers or when examining a 'higher-order' competence, such as analysis. If this is at all pertinent, you have to look for a system that offers this functionality. You should also look into the amount of leeway in the marking method: can it be used by multiple markers? And are there various possibilities for calculating marks and grades?

8. Availability of statistical data

Assessment software systems do not all have the same possibilities for extracting and analysing statistical data; for example, the p, rit and a values, but also data on how markers rate questions and a representation of the chance of guessing the answer to each question. Sometimes you will also need information about which candidates have already had a specific question in an exam. This is, for example, the case if you use flexible-moment exams or if you want to exclude one or more questions from an exam after it has been taken.

9. Multimedia applications

With some types of questions, it can be useful to employ visual material in the form of illustrations and videos. How easy is it to do this with the software? What are the respective image-format and file-size requirements?

10. Links to other software

Is it to be expected that a link to other software will be required? This may be the case if other programs are to be used during the exam, such as Excel or another bookkeeping program. Some systems offer this possibility.

11. Options for taking exams

Written exams have completely different possibilities to digital exams. We have already discussed some of them in this booklet. To what extent is it necessary to hold exams at a specific time and or location? In this respect, you have to consider measures taken to make the system future proof. What do you do if candidates are spread all over the country or live abroad? In that case, it would be wise to take this into account from the start. When you look at the options for holding exams, also check the possibilities for dyslexics, the visually impaired, etc. And the possibilities for taking exams offline; there are bound to be situations where there is no internet connection.

12. Customisation

How pleasant is to increase the recognisability for users and candidates by creating an exam in your own style and with your own logo? Various software packages offer this possibility. Can the system be customised? Is the supplier willing to make adjustments to the system to meet specific requirements? Are they open to innovative ideas and product development suggestions?

Choosing the right software for your organisation is a complex process and some of the aspects are critically important. They are briefly outlined above, but each system has to be reviewed on its own merits. The above aspects will help you ask assessment software suppliers the right questions.

Innovation in digital assessment

- People are increasingly and more frequently demanding to be able to use digital tools. This also applies to education. A large amount of course material is presented in a digital format; some schools already work with laptops and/or tablets. As a result, people are becoming more and more accustomed to using digital tools, which is also an argument for presenting exams digitally as much as possible.

Innova

It is a good idea to point out that, at the present time, exams with open questions are still virtual ly always marked by real people. Closed questions are marked by computers. Human marking has a greater margin of error than automated marking. Different types are prejudices can also play a role in human marking. In the future, it is not inconceivable that AI (artificial intelligence) will be used for marking open questions.

Several other innovations in digital assessment are discussed below:

- Skills testing
- Online proctoring
- Fraud prevention

Skills testing

With respect to digital applications in the field of skills learning and testing, relevant developments, such as serious gaming and virtual reality, have taken off in recent years. Serious gaming is a form of education and testing that goes beyond playing a game for fun. Research has already demonstrated the learning effect of serious gaming. You can learn more from serious gaming than you can from e-learning (source: itpedia.nl). Virtual reality is also relevant in education, especially where the perception of students and exam candidates is involved. Because virtual reality gives you the feeling that you are part of a situation, it is not only easier to learn but also easier to give the right answers. Training more complex skills with serious gaming is especially booming.

Serious gaming and virtual reality are also used for making predictive assessments for businesses. The way in which air traffic controllers are examined is a good example in this respect. Serious gaming is used to assess the suitability of candidates for a job as an air traffic controller. Another example is the use of virtual reality in predictive and summative assessments of a candidate's leadership skills. Assessment based on serious gaming and virtual reality generally has a high face (or perceived) validity. This means that the relevance to professional practice of this form of assessment is perceived by candidates as high.

What happens on the candidate's screen can be monitored remotely.

Online proctoring

One of the methods for taking exams that is developing rapidly and significantly increasing flexibility for candidates is online proctoring. This option is briefly mentioned in paragraph 2.1. It is discussed below in more detail to provide insight into how the online proctoring process precisely works.

A candidate makes a recording of their surroundings and identify themselves with a webcam and a mobile telephone. The proctor can see what is happening on the candidate's screen and, via the webcam, can also check the surroundings

for any irregularities or unpermitted aids. There are two variants: one with which the proctor is in direct contact with the candidate and one with which the proctor reviews the surroundings for irregularities after an exam has been taken. The second variant gives candidates more freedom regarding when and where they take an exam, but no action can be taken while the exam is in progress, for example if a camera does not point in the right direction. In both cases, it is important to take measures to protect candidates' personal data and to ask permission to use online proctoring. This is particularly true when collecting, recording and storing visual material.

Fraud prevention

Finally, fraud prevention methods are being developed to detect unexpected answer patterns and unusual amounts of time spent on answering different questions. Possibly fraudulent practices can be tracked down on the basis of data analysis. This can, for example, be the case if a candidate needs almost no time at all to answer difficult questions correctly and takes a relatively long time to answer easy questions. This technique is still under development and the findings to date cannot be blindly assumed to be 'true'. The technique cannot be used independently from other methods, such as direct observation at the exam location. If a candidate, for example, is believed to have cheated in an exam, subsequent data analysis can be used to verify or dispel suspicions.

In short: advances in digital assessment are being made on many fronts.

7
Literature

Literature

Eggen, T.J.H.M., & Sanders, P. F. (Eds.) (1993), *Psychometrie in de praktijk (Psychometry in practice)*, Arnhem: Cito.

Hattie, J., & Timperley, H. (2007). *The power of feedback*. Review of Educational Research, 77(1), 81-112.

Jordan, Sally and Mitchell, Tom (2009). *e-Assessment for learning? The potential of short-answer freetext questions with tailored feedback*. British Journal of Educational Technology, 40(2), pp. 371-385.

Kaap, A. van der (2018), *Formatieve evaluatie, werken aan groei (Formative evolution, working towards growth)*, Enschede: SLO.

Karpicke, J. D., & Blunt, J. R. (2011). *Retrieval practice produces more learning than elaborative studying with concept mapping.* Science, 331, 772-775.

Kleij, F.M. van der (2017), *Leeropbrengst van feedback in computergestuurde toetsen; wat is effectief? (Learning outcomes from feedback in computer-controlled testing; what is effective?)*, Arnhem: CITO.

McKenna, C., & Bull, J. (1999). *Designing effective objective test questions: an introductory workshop*, CAA Centre, Loughborough University.

Sluijsmans, D., Peeters, A., Jakobs, L., Weijzen, S. (2012). *De kwaliteit van toetsing onder de loep (Assessment quality under a microscope).*

Sluiter, C., Hemker, B., en Eggen, T. (2018). *Beoordelen van de kwaliteit van toetsen en examens (examens (Assessing the quality of tests and exams).* Arnhem: Cito.

Stobart, G. (2008). Testing times: *The uses and abuses of assessment*. Londen: Routledge.

Veldhuijzen, N.H. , Goldebeld, P., & Sanders, P.F., (1993). *Klassieke testtheorie en generaliseerbaarheidstheorie (Classic test theory and generalisability theory).* In T.J.H.M. Eggen & P. F. Sanders (Eds.) , Psychometrie in de praktijk (Psychometry in practice) (pp33-82), Arnhem: Cito.

Vrieze, de J., *Test uzelf en haal een hoger cijfer (cijfer (Test yourself and get a higher grade)* , NRC, 25-1-2011.

Wilbrink, B. (2005). *Toetsvragen Schrijven (Writing exam questions)* , Het Spectrum.

General (digital) assessment

http://nl.wikibooks.org/wiki/Onderwijstechnologie/Klastechnologie/Digitaal_toetsen

https://wij-leren.nl/taxonomie-van-bloom.php

https://www.utwente.nl/en/examination/

Question types

http://www.open.ac.uk/blogs/SallyJordan/?p=215

Learning and examining with serious gaming

https://www.itpedia.nl/2017/09/01/serious-games-zijn-effectiever-dan-e-learning/

https://www.ou.nl/welten-serious-gaming

https://www.brainsfirst.com/in-het-nieuws/

Appendix 1

Checklist for the transition to digital assessment

Notes marked with an asterisk (*) indicate specific differences for fixed-version exams.

Setting up question banks

Checklist	Explanatory notes
1. Are at random or fixed-version exams to be used	The advantages of at random exams are the efficiency, overview and ease of the development process, exam preparation and maintenance, and the limited size of the database. ***** *The advantage of fixed-version exams is that the structure of the question bank does not have to be worked out in detail.*
2. Does each exam have the same number of points to be awarded/number of questions or is variation allowed? **3.** How many questions are there in one exam?	This Is important with at random and fixed- version exams. If this is not kept the same in the question bank, exams may be unfair.
4. How many questions must be developed in the first instance? How many in the second instance?	This depends on the applicable standards. For example, the possibility of drawing an exam three times in the first instance. ***** *A minimum of four rounds of questions should be selected for fixed-version exams to allow a safety margin*

5. What is the maximum number of questions that can be active in the database?	ith relatively small numbers of candidates (for example 150 per year), it can take a long time before the same questions are have been asked often enough to enable you to carry out meaningful analyses. In this type of situation, limiting the maximum number of active questions in the database to, for example, twice the size of the number needed for the exam can be a good idea.
6. What is the maximum attainable score per question? **7.** How many questions are needed per subject?	The attainable score per question and the number of questions per subject have to correspond with the respective subject weighting in the exam matrix. With randomised exams, this only has to be set once. ***** *You always have to check whether fixed-version exams meet the weighting requirements by hand.*
8. What subjects are so important that they have to be included in every exam?	Make these subjects a fixed component in the que stion tree. ***** *Or make a note to include them in fixed-version exams.*
9. Which exam requirements or subjects are nearly always tested together? **10.** Do any exam requirements have to be met by answering follow-up questions with the correct level of depth?	Both aspects justify cre ating associated instances or cases in the question tree and question bank. ***** *With fixed-version exams, questions can be selected and used more freely*

11. What is the ratio between knowledge, insight and application questions (Bloom, 1956)?	Are the required levels of knowledge, insight and application incorporated in the exam matrix and, if yes, are they leading? You can deviate from this, but the reasons for doing so must be clear in the question tree. As a rule, questions on the same branch on the question tree should be equally weighted.
12. At which proficiency level (knowledge, understanding and/or application (see Bloom 1956) is what assessed?	** With fixed-version exams, you can handle levels of knowledge, insight and application more freely, but you still have to check whether the exam as a whole has the correct ratio (between knowledge, understanding and application).*
13. Which exam requirements are to be assessed with open questions or with closed questions?	You look at whether an open or closed question is best for assessing each exam requirement. With randomised exams, you do this to ensure that the requirements are assessed on an equal footing. If the answer to a closed question is quickly given away, it would be better to opt for an open question, even if candidates have to formulate their own answers.
	* With fixed versions, you look at whether the exam as a whole has the same ratio of open/closed questions as exams that have already been developed. You have to do this by hand.

14. What guidelines do we use for using different forms of questions?	***** *This is important for randomised and fixed-version exams.*
	How do we ensure that the questions in the database have the same parity? This means that guidelines have to be drawn up for the scope of the questions. What is the maximum number of alternatives for each form of question?
15. Do we award part or only whole marks for correct answers to closed questions?	
	This also depends on whether part marks may be given and the formulas used for calculating them. Giving part marks leads to more balanced and fairer overall scoring, but the underlying reasons have to be explained.
16. What is the pass mark?	Is guessing the correct answer taken into account?
	Is there a reason for deviating from the 'standard' norm?
	Should we use step marking to allow candidates to become 'accustomed' to an exam?
17. How much time is needed per question?	Although it may not be necessary to indicate a time allowance in your system, it is still a good idea to make an estimate. See also checklist question 24.
18. What is the allocation of tasks?	Who enters the questions? Who checks the input?

The 'look and feel' of exams

Checklist	Explanatory notes
19. How does the enrolment process work?	Can candidates enrol and pay for exams by themselves?
20. What emails are sent to candidates when then enrol?	
21. Are logos needed in an exam?	
22. What do the start and end screens for the exam look like?	
23. How much time is available for taking an exam?	
24. Wat is de beschikbare tijd	
25. How much extra time may be allowed if it is requested?	
26. What resources can be made available for special target groups, such as dyslectics and the visually impaired people?	

27. Can a hard copy exam be printed in the event of a system malfunction?	

Marking, results and analysis

Checklist	Explanatory notes
28. How many markers are needed per hosted exam?	When is a second and/or third marker called in? Within what range of marks?
29. How is the mark structured with open questions	Is the mark for each question the average given by two markers? Or are deviating marks referred to an adjudicator? Or does the second marker award marks?
30. How many days are markers given to do their work?	Is there any difference between the first, second or, if applicable, third marker?
31. How long do candidates have to wait for their results?	
32. Can this period vary in length?	Having markers readily available means that results can be quickly announced.

33. Can the results be announced immediately after closed question exams have been taken?	Candidates usually want to know the results immediately after taking an exam, but telling them their result straight away can cause unrest
34. How is feedback given?	When the results are announced or before? And on what level is feedback given – per subject or per question?
35. Are candidates given the opportunity to see their exam after it has been marked?	Is additional feedback visible on correct and/or incorrect answers? Are all answers to questions visible or only the incorrect ones?
36. Can any situation arise in which the results have to be individually or collectively adjusted?	What would prompt results adjustment?
37. What data do we use for analysing the questions?	From the analysis results, you can see which questions may be too difficult or improperly formulated (P, A, RIT values).
38. What do we do with questions that are virtually always answered incorrectly?	Is periodic maintenance planned? Or is the first analysis performed only after sufficient data has been accumulated for statistic relevance? (i.e. when the questions have been asked sufficiently frequently)
39. Who checks the quality of the questions following complaints?	

40. Who checks the settings in assessment software if complaints are made?	
41. What other data do we use?	For example: • Marks per instructor; • Differences in the way markers interpret questions and answers; • The amount of time needed per exam; • Pass mark percentages over a given period
42. Do we ask candidates to fill out a brief questionnaire immediately after taking an exam? Or do they have the opportunity to give immediate feedback?	Example questions • Whether they think the exam ties in with the course; • Whether they think the exam is too easy/difficult; • Or they think the exam content is recognisable.

This checklist does not contain any information about organisational processes, dealing with (freelance) content experts and communication with candidates and other stakeholders. This is because these things are extremely context-specific and fall outside the scope of this document. Most assessment software suppliers will be able to advise you so don't be afraid to ask them your questions.

If you have any questions about digital assessment,

please contact Optimum Assessment:
T: +31 (0)13 528 63 63
E: info@optimumassessment.com

optimum
assessment