

# Guidelines for the Design of Digital Closed Questions for Assessment and Learning in Higher Education

Centre for Educational Training, Assessment and Research, Vrije Universiteit Amsterdam  
[s.draaijer@ond.vu.nl](mailto:s.draaijer@ond.vu.nl)

R.J.M. Hartog  
Wageningen MultiMedia Research Centre, Wageningen University  
[rob.hartog@wur.nl](mailto:rob.hartog@wur.nl)

Stichting Cito Instituut voor Toetsontwikkeling, Arnhem  
[joke.hofstee@cito.nl](mailto:joke.hofstee@cito.nl)

## Abstract

Systems for computer based assessment as well as learning management systems offer a number of innovative closed question types, which are used more and more in higher education. These closed questions are used in computer based summative exams, in diagnostic tests, and in computer based activating learning material. Guidelines focusing on the design of closed questions were formulated. The use of these guidelines was evaluated in fifteen case studies in higher education. The conclusion is drawn that guidelines are useful, but should be applied in a broad approach that is best to be supported by educational technologists.

Keywords: assessment, ICT, computer based testing, technology, testing, methodology, digital learning material, closed question formats

## Introduction

During the last decade, a range of selected response format questions and other formats that allow for automatic scoring, have emerged in computer based testing software (Bull & McKenna, 2001; Mills, Potenza, Fremer, & Ward, 2002; Parshall, Spray, Kalohn, & Davey, 2002) and Learning Management Systems (LMS's) or Virtual Learning Environments (VLEs). Examples of such questions are 'multiple response', 'drag-and-drop', 'fill-in-the-blank', 'hot spot' and 'matching'. For reasons of readability, from now on the term 'closed question' will be used. In higher education such closed questions are used in summative tests (exams), in diagnostic tests but also in activating learning material (ALM). ALM forces the student to actively engage with the learning material by making selections and decisions (Aegerter-Wilmsen, Coppens, Janssen, Hartog, & Bisseling, 2005; Diederens, Gruppen, Hartog, Moerland, & Voragen, 2003).

As any design endeavour, the design of sets of closed questions is likely to benefit from a design methodology. The ALTB project (Hartog, 2005) aims to develop such a methodology for the design and development of closed questions for summative exams (SE) and activating

learning material (ALM) for engineering and life sciences in higher education. This methodology is expected to consist of design requirements, design guidelines, design patterns, components, and task structures.

The research question of the ALTB project is essentially: ‘How and under what conditions is it possible to support the design and development of digital closed questions in higher education?’ The answer should support the rationale for the methodology. This article focuses specifically on the development and evaluation of design *guidelines*.

## Limitations in current literature on design guidelines

Literature on the design of questions with a closed format is mainly restricted to the design of summative tests that consist of ‘traditional’ multiple-choice questions. This literature, for example Haladyna et al. (2002), usually presents a large set of design requirements i.e. constraints that must be satisfied by the questions that are output of the design process. An example of such a constraint is the rule that every choice in a multiple-choice question should be plausible. A constraint like this helps to eliminate a wrong or poorly constructed question, but it does not help to create a new question or better distractors. Only certain requirements can be regarded as direction giving requirements rather than as constraints, but many requirements are not useful for directing and inspiring question designers.

Nevertheless, in literature on the design and development of questions and tests, requirements are often denominated as ‘guidelines’. The use of the term ‘guideline’ for ‘requirements’ obscures the lack of real design guidelines i.e. rules that open up creative possibilities for question design and support the designer(s) during the design process.

Insofar literature does provide inspirational guidance for designers and developers of closed questions - as for example by Roid and Haladyna (1982), Haladyna (1997) or Scalise and Gifford (2006) - these sources are in the form of quite elaborate texts or research reports and more suited for secondary or vocational education. Given the limited time for training or study available to lecturers, guest lecturers and instructors (SME’s) in higher education, they do not use these sources and do not feel that they are appropriate.

For that reason, it is assumed that more ‘compact’ and easily accessible guidelines, preferably in the form of simple suggestions, can be more useful in practical situations in higher education. Based on that idea, a set of 10 categories and direction giving requirements was formulated and made available in the form of an overview table and brief explanations.

In practice in higher education, the same technology and the same question types are used for both summative exams as for activating learning material. Therefore, at the outset of the project, it was the intention to develop guidelines that were suitable for the summative role and the activating learning role.

## The Guidelines: dimensions of inspiration

In this section a set of guidelines for the design of closed questions and the rationale of these guidelines will be described. The guidelines should serve as an easy to use and effective support for SME’s and assistants for the design and development of questions and tests.

In order to arrive at a set of potentially useful guidelines, the ALTB project team formulated a set of guidelines. These guidelines were partly derived from literature and partly from experience of the project team members. Some guidelines are quite abstract, other guidelines are very specific, some guidelines refer to methods, others guidelines refer to yet another 'inspirational' category. The guidelines were grouped into specific categories each of which was intended to define a coherent set of guidelines. The list comprised ten categories: seven categories consisted of guidelines that tap into the use of experiences and available resources for question designers:

- A. Professional context
- B. Interactions and Media
- C. Design Patterns
- D. Sources
- E. Learning Objectives
- F. Students
- G. Sources

Three categories were essentially traditional requirements. However, those requirements give direction and inspiration to the design process

- H. Motivation
- I. Validity
- J. Equivalence

These categories were sub divided in more specific guidelines, resulting in a total of 60 guidelines. In the following sections, the guidelines are described in more detail.

## A: Professional context

This category of guidelines makes question designers focus on the idea that information is more meaningful when it is presented or embedded in real life professional situations (e.g. Merriënboer, Clark, & Croock, 2002). Based on that idea, the professional context of a graduated professional in a specific domain could be the basis of these questions. To cover multiple aspects of such cases, more than one question should be defined. An obvious source for such authentic situations can be the professional experience of the question designer himself.

In a more systematic way, question designers can use explicit techniques for constructing and describing cases, for example in the form of vignettes (Anderson & Krathwohl, 2001), as elaborate item shells and item sets (Haladyna, 2004; LaDuca, Staples, Templeton, & Holzman, 1986; Roossink, Bonnes, Diepen, & Moerkerke, 1992).

A second source that thrives on professional knowledge and experience is to tap into 'Eureka' experiences the professional has had in his own learning and professional development. More specifically these types of situations were worked out in tips and tricks, surprising experiences, counter-intuitive observations and natural laws, relevant orders of magnitude, typical problems and best first steps for tackling them.

Finally a guideline that often pops up in the practice of instructional design projects is the advise to collect all kinds of material (interviews, documentaries, descriptions, journal clippings, broadcast video and audio), which can be used to construct or illustrate cases.

<b>Professional context</b>	
A1	Develop cases with authentic professional context and multiple relevant questions.
A2	Develop vignettes using an item-modelling procedure: split up authentic cases in various components and develop new content for each component and combine them into questions.
A3	Investigate your own professional experience. Make lists of:
A3.1	Tips and tricks.
A3.2	Surprising experiences.
A3.3	Counter-intuitive observations and natural laws.
A3.4	Relevant orders of magnitude.
A3.5	Typical problems and the best first steps.
A4	Collect interviews, documentaries, descriptions (in text, audio or video) of relevant professional situations. Use these for question design.

## B: Interactions

The introduction of the computer in learning and assessment makes a new gamut of question types and interactions possible. The ALTB project team anticipated that when question designers play with assessment software and study the accompanying examples, they become inspired.

To guide question designers more specifically on the dimension of digital media inclusion, guidelines were formulated that take *specific* digital media types into mind which would lead to more appealing questions or that would measure the intended attribute of interest more directly: pictures and photos, video's, audio, graphs, diagrams, process diagrams.

<b>Interactions</b>	
B1	Play with available assessment software. There is a variety of assessment systems on the market. For inspiration on asking new questions and test set-ups: try out the interactions in the system that is used in one's own organization.
B2	Scan the IMS-QTI interaction types on usability.
B3	Collect material for media inclusion:
B3.1	Pictures / photos.
B3.2	Video clips.
B3.3	Sounds / audio fragments.
B3.4	Graphs.
B3.5	Diagrams.
B3.6	Process diagrams.

## C: Design patterns

The term 'design patterns' is introduced by Alexander (1979) in the seventies of the last century as a concept in architectural design. In design in general, reuse of components as well as reuse of patterns is beneficial because it usually is efficient but also because reuse of components and/or patterns increases the probability that errors or disadvantages will be revealed. An experienced designer is supposed to have many patterns in his mind. "It is only because a person has a pattern language in his mind, that he can be creative when he builds" (Alexander, 1979: p. 206).

Because design patterns for digital closed questions were not readily available, a simpler approach was taken, using types of directions that could be indicative for design patterns. A few guidelines were presented that could be viewed as preliminary versions of design patterns or families of design patterns.

The first pattern was taken from Haladyna (2004: p. 152). This pattern, presented as a guideline, advises question designers to use successful ‘starting sentences’ that can easily result in interesting and relevant questions. A similar guideline by Haladyna (Haladyna, 2004: p. 153) advises question designers to take successful items, strip the items of specific content, however leaving the systematic of the question unaltered, and then systematically design questions based on variations of content. This can be regarded as a generic advice to use design patterns. Another set of design patterns direct question designers toward questions that ask for completion of statements or calculations, to identify mistakes in reasoning or calculations, and to identify the best descriptions or key words for presented texts. The last guideline is based on ideas by Wilbrink (1983). Wilbrink suggests that – especially for designing True/False questions – it is a worthwhile technique to relate different (mis)concepts, to use (in)correct causes and (in)correct effects of concepts as a starting point for questions.

<b>Design Patterns</b>	
C1	Items shells I: Use a list of generic shells. Examples: <ul style="list-style-type: none"> <li>• Which is the definition of ...?</li> <li>• Which is the cause of ...?</li> <li>• Which is the consequence of ... ?</li> <li>• What is the difference between ... and ...?</li> </ul>
C2	Item shells II: Transform highly successful items into item shells.
C3	Collect chains of inference and calculations as a basis for a completion question. The completion question requests to fill in the missing rule in an inference chain or calculation
C4	Use design pattern “Localize the mistake”: introduce a mistake in a text (paragraphs), photo, diagram etc. and use this as the stem. (Collect texts, photo’s and so on.)
C5	Use design pattern “Select the (3) best key words” to a text. (Collect texts)
C6	Use design pattern “select a title” to a text. (Collect texts)
C7	Develop implications of statements.

## D: Textbooks

In many courses in higher education, the dominant instructional sources are publishers’ textbooks or the course syllabus. These books hold the core of the subject matter for a given course. For question design, a guideline is to use the content of these books not at random, but systematically. Whilst it was anticipated that a large number of question designers could feel that such a guideline was too ‘simplistic’, pointers that are more specific were added to guide question designers more precisely. The pointers were categorized into the use of media such as photos, graphs, and diagrams on the one hand and statements, contradictions, conclusions, exceptions, examples, abstract concepts, and course specific content emphasis made by the instructor on the other hand.

<b>Textbooks</b>	
D1	Walk systematically through the textbook (paragraph by paragraph) and look for:
D1.1	Photos.

D1.2	Diagrams.
D1.3	Graphs.
D1.4	Statements.
D1.5	Contradictions.
D1.6	Conclusions.
D1.7	Exceptions.
D1.8	Examples.
D1.9	Abstract concepts.
D1.10	What paragraphs and concepts hold key information and which do not.

## E: Learning Objectives

Course goals and learning objectives are essential ingredients in instructional design (Dick & Cary, 1990) and for the design and development of tests and questions. Clear learning objectives are the basis for establishing valid assessment and test objectives: what will be assessed in what way, at what level (often resulting in a test matrix). Detailed learning goals are not well specified in many design and development situations. In such situations, making questions without first specifying the detailed learning objectives is a realistic option.

Furthermore, a question designer could analyse and categorise the questions that are already available in previously designed assessment material thus raising the objective formulation to a higher level of abstraction. Based on the assumption that previous assessments reflect the knowledge and skills the instructor finds important for a course, this categorisation can be used to design new questions.

Categorisations as described above will often be formulated in terms of domain specific knowledge and skills that need to be acquired. Taking a top down approach however, questions designers are advised to start with using more abstract formulations of the types of knowledge and types of cognitive processes that need to be assessed with the support of a *taxonomy* or *competency descriptions*. There are more taxonomies available, but an often proposed taxonomy is Bloom's taxonomy (1956) or the taxonomy as proposed by Anderson and Krathwohl (2001).

Learning Objectives	
E1	Use an existing list of very specific and detailed formulated learning objectives.
E2	Make a list of very specific and detailed formulated learning objectives.
E3	Analyse educational objectives using a taxonomy of objectives.
E4	Use the competency description of a course as a starting point to design questions.

## F: Students

The students' mind set, experiences and drives should be – at least for learning materials – a source of inspiration for the question designer (Vygotsky, 1978). Four guidelines express this point of view.

The first guideline directs the question designer towards imagining prior knowledge of the student; specifically insofar this might be related to the subject matter or the learning objectives of the course. Thus, questions relating to for example food chemistry, should build on students experiences with their chemistry knowledge as acquired at secondary education.

The second guideline directs the question designer in thinking of the more daily experiences that students have. In the food chemistry case study, questions could start by using examples

of food that students typically consume. The third guideline asks question designers to use facts, events, or conclusions that can motivate and inspire students. Again, for food chemistry, students in certain target populations are motivated for example by questions that relate to toxic effects or environmental pollution.

Finally, it makes sense to use a common error or a common misconception as starting point for the design of a question. This method is elaborated in detail by Mazur (2001) with his ConcepTest approach.

<b>Students</b>	
F1	Imagine and use prior knowledge of the student.
F2	Imagine and use the experience of the student.
F3	Imagine and use the things that motivate and inspire students.
F4	Collect errors and misconceptions that students have.

## G: Sources

In a wider perspective than already proposed in A (Professional context), D (Textbooks), a set of guidelines was formulated to stimulate the systematic use of every possible information resource for inspiration. Five specific guidelines were formulated.

The first two guidelines call upon question designers to get informed by interviewing colleagues at the educational institution and professionals working in the field of the domain. A third guideline asks question designer to get informed by, or work with, Educational Technologists (ET's). They can inspire question designers not so much on content related aspects, but much more on the rules and techniques to design questions in general. A fourth guideline suggest that question designers should set up brainstorming or brain writing exercises and the like (Paulus & Brown, 2003). The goal of such a session is to come up with as much as possible questions and pointers towards possible questions without being restricted too much by all kinds of requirements, impracticalities, or even impossibilities. Restriction and convergence is dealt with in a later stadium. A fifth guideline proposes question designers to systemically collect as much as possible relevant information from sources outside their institution and outside their own social and professional network and in particular from sources that can be accessed over the internet.

<b>Sources</b>	
G1	Question colleague instructors of the faculty.
G2	Question professionals working in the field of the subject matter.
G3	Question educational technologists.
G4	Set up and execute brainstorm sessions.
G5	Collect information from various sources such as news papers, the internet, news broadcasts.

## H: Motivation

Attention is a bottleneck in learning (Simon, 1994) and motivation is essential for effective and efficient learning. Keller (1983) formulated four variables that are important for motivation. Based on the variables 'direction giving requirements' are formulated that could inspire question designers. These requirements are conform Keller's ARCS model (A: the question should captivate the Attention of the student, R: the question should be perceived as Relevant by the student, C: the question should raise the level of Confidence of the student and S: the question should raise the level of Satisfaction of the student).

So, motivation is regarded as a separate inspirational category. A question designer should try to design questions that meet the requirements given in this category. Only afterwards, it can be established whether a question meets the requirement.

<b>Motivation</b>	
H1	The question focuses the attention of the student for a sufficient amount of time.
H2	The question is experienced as relevant to the student.
H3	The question raises the level of confidence by the student.
H4	Answering a questions yields satisfaction by the student.

## I: Validity

Validity in assessment is an important requirement. Tests and questions should measure what they are intended to measure and operationalise the learning objectives (criterion referencing). Because of their relation with learning objectives, validity requirements also give direction to the design process. Three direction giving validity requirements were formulated.

The first guideline reflects the requirement that questions need to measure the intended knowledge or construct that should be learned. The second guideline advises question designers to think more in terms of *sets* of questions to measure knowledge and skill than solitaire questions. The third guideline is actually a requirement to the test as a whole: in a test, the weight of a learning objective should be proportional to the number of questions measuring the knowledge and skills involved in that objective.

The scope of the ALTB project was limited to question design and did not focus on the design of complete assessments. Nevertheless, some of the guidelines clearly apply to the design of complete assessments as well. Guidelines that tap into designing valid assessments and test are formulated in D (Textbooks) and E (Learning Objectives). These guidelines direct the question designer to layout the field of knowledge and skill to be questioned so that a good coverage of the learning material can be achieved.

<b>Validity</b>	
I1	The question is an adequate operationalisation of the learning objectives.
I2	The question itself is not an operationalisation of the learning objectives, but the set of questions is.
I3	Within a test, the weight of a learning objective is represented in the number of questions that operationalise that learning objective.

## J: Equivalence

In higher education in general, tests and questions for summative purposes cannot be used again when they have been deployed. The reason for this is that assessments and test questions in general cannot be secured sufficiently and that subsequent cohorts of student would be assessed non-equivalent if they already have been exposed to the questions. Consequently, instructors need to design equivalent assessment and test questions to ensure that every cohort of students is assessed fairly and comparably. Four equivalence requirements were expected to function as not only a filter on questions but also as beacons that could direct the design process. These were equivalence with respect to content (subject matter), interaction type, cognitive process and finally also to scoring rules.

Equivalence	
J4.1	Equivalent in relation to subject matter.
J4.2	Equivalent in relation to interaction type.
J4.3	Equivalent in relation to level of difficulty and cognitive processes.
J4.4	Equivalent in relation to scoring rules.

## Case studies to investigate the appropriateness of the developed guidelines

The use of the guidelines, has been observed in fifteen case studies. An overview of the case studies is presented in Appendix 1. Most case studies had a lead time of less than half a year. The case studies overlapped in time. Later case studies could make use of experience in earlier case studies. The numbering of the case studies is an indication of the point in time when the case studies were carried out. Column two represents the institution in which the case took place. Column three indicates the course level and column four the course subject. The fifth column depicts the role of the questions within the course: summative, (formative) diagnostic or (formative) activating. Column six lists the authoring software that was used en the last column lists the main actors within the development team.

The cases mostly consisted of design projects for university level courses in which SME's, their assistants and sometimes ET's, designed and developed digital closed questions to be used as summative exam material or activating learning material.

The question designers or teams of question designers (SME's, assistants, ET's) were introduced to the guidelines in an introductory workshop. The function of the guidelines (i.e. inspire the question designers) was emphasized during these introductions, the how and why of the categories was explained and the guidelines were briefly discussed and illustrated with some additional materials. In the first workshop, the teams exercised in question design using those guidelines. Later on, during the execution of the projects, an overview sheet of the guidelines was at the disposal of the SME's and assistants, any time they felt they wanted to use it.

The set of guidelines was formulated while the case studies WU1 and WU2 and the first part of TUD1 were running. The direction of the literature search for design guidelines was partly determined by projects on the design of digital learning materials that gave rise to the ALTB project and partly by these first three case studies.

Once the set of design guidelines was considered complete, all designer teams in the ALTB project were asked to start using the guidelines in all question design and development activities and to provide two reports.

For the first report the procedure was:

- Design and develop 30 closed questions as follows:
- For each question do:
  - For each design guideline/direction-giving-requirement do:
    - Record if it was useful;
    - Record if its use is recognizable in the resulting question.

It was expected that this procedure would demand considerable discipline from the designers. Therefore, the number of questions that would be subjected to this procedure was limited to 30. The second report would be a less formal record of the experience of working with the guidelines for the remaining questions. A short report was made of every case. For most cases, data were recorded on the execution of the process and use or non-use of guidelines. In the Appendix 2, the major findings per case are listed.

In case studies VU1, VU2, TUD2, WU9 and WU10 – partly based on preliminary versions of both reports – ET's tried to support the designer teams in using the guidelines and described their experience.

## Criteria for assessing the value of the guidelines

The research question of the ALTB project as stated in the introduction, can be mapped onto a research design consisting of multiple cases with multiple embedded units of analysis (Ma, 2004). A small set of units of analysis was identified. These units of analysis are: a set of design requirements, a set of design guidelines, a set of design patterns, a set of interaction types, a task structure, and resource allocation. As said, this article focuses on the development and evaluation of set of guidelines. What are the useful criteria to establish whether guidelines are a worthwhile component of a methodology?

First, within a methodology, guidelines form a worthwhile component if, for any given design team, the set of guidelines includes at least five guidelines the team can use. It is expected that the value of specific guidelines will depend on the specific domain, the competency of the question designers, and so forth and so on. However, a general finding that guidelines can support the design and development process must be answered positively.

Second, the ALTB team wanted to investigate how the development teams would and could work with the complete set of guidelines in practice. Is a team willing and capable of dealing with a fairly great number of guidelines and able to select the guidelines that are most useful for them?

Third, a methodology for the design and development of closed questions must in principle be as general applicable as possible. As closed questions are used in both summative tests and activating learning material, it is worthwhile to examine the assumption that one set of guidelines can be used equally well for both roles. Maybe however, given the intended role for question design, different sets should be offered upfront in a development project.

## Observations

### Execution of the method

One team of question designers declined to work with the set of design guidelines. This team was involved in a transition from learning objective oriented education to competency directed education. The goal for this team was to design and develop diagnostic assessments. The team argued that the guidelines had a too narrow focus on single questions instead of on clusters of questions. Furthermore, this team expected that the guidelines would prevent creativity instead of boosting creativity. This team proposed to start developing questions without any guideline and abstract later from their behaviour a set of guidelines. De facto, it turned out that this team focussed completely on guideline A1. The resulting questions

however did not to reflect their efforts in developing cases. Furthermore, the questions did not reflect the philosophy of competency based education. A number of questions had feedback that consisted of closed questions. No other guidelines came out of this case study.

All other teams were initially positive about performing the two tasks. However, it soon turned out that rigorous following the procedure was more difficult than expected.

Two teams (VU1 and VU2) tried to execute the procedure but got entangled in a discussion on the appropriateness of the guidelines. This caused them to lose track of the procedure. As a result no careful record was produced. However, these two teams did produce a number of closed questions on the basis of the guidelines. All the other teams produced a record of the thirty-question-procedure.

A final general observation is that budget estimations were too low for all cases. The design and development of questions took three to four times the amount of time that was budgeted based on previous reports.

## Use of the guidelines

The developed set of guidelines was actively used by all teams but one. Browsing through the guidelines and discussing them made SME's and assistants aware of multiple ways to start and execute the conception of closed questions. Within the set, there were always four to five guidelines available that in fact helped question designers to find new crystallization points for question design they had not thought of before.

In VU1, VU2, TUD2, WU9, WU10, SME's were of the opinion that categories B (Interactions) and C (Design Patterns) often resulted in questions that were new for the intended subject matter. Example questions, presented by the ET (often devised by the ET on the basis of preliminary information, textbooks or identified within other sources such as the internet), or questions stemming from previous developed tests, quickly invoked conceptual common ground between SME, assistant and ET. This common ground enabled the assistant to apply the core idea of the given example to questions within the intended domain. It was also noted that this effect was the strongest when the example questions were as closely as possible linked to the intended domain.

The guidelines to use digital media (B3x, D1.1, D1.2 and D1.3) in the form of photos, graphs, diagrams, and chemical structures and so on, turned out to be a worthwhile guideline for the majority of teams. Systematic focus in the design process to use such media was regarded as useful and led to new questions for the teams.

For the design and development of summative exams, category J (Equivalence) turned out to be a dominant guideline. This is due to the fact that for summative exams a representative coverage of a larger number of detailed learning objectives is necessary and that re-exams should be as equivalent as possible as long as the learning objectives do not change.

Given the observation that the guidelines in category J were not tangible enough, a new guideline for that role was formulated. This guideline advises question designers to aim directly at a cluster of five equivalent questions for each detailed learning objective, textbook paragraph or image by making variations on one question. This guideline is phrased as: *design and develop clusters of five equivalent questions*. Making slight variations on one question (paraphrasing, changing responses orders, splitting up multiple choice question in variations of 2, 3 or 4 alternative questions, using different examples, questioning other

aspects of the same concept, varying the opening sentences) will cost relatively little effort as compared to designing and developing a new question.

### General critique in the case study reports regarding the set of guidelines

Many question designers were of the opinion that the presentation of the complete set of design guidelines made them see the wood for the trees. SME's and assistant repeatedly called for "Give me only the guidelines that really can help me". Presenting the complete set resulted in a lower appreciation for the guidelines as a whole.

At the same time, a number of guidelines were regarded as 'too obvious' by SME's and assistants or were regarded as variations of the same guideline. This counts especially for guidelines Professional context (A), Textbooks (D), Learning Objectives (E), Validity (I) and Equivalence (J). Of course, the perceived usefulness of a guideline is in practice related to the extent to which a guideline is new for a designer/developer. However, declaring any guideline that is well known, as useless, is in our opinion not a valid reason to exclude it from the set of guidelines. However, this perception of the guidelines by SME's and assistants also results in a lower appreciation for the guidelines as a whole.

### Limitations regarding specific guidelines

Often the SME's and assistants could formulate why they had *not* used a *specific* guideline.

The first general reason for this was that it was unclear how a specific guideline operates. SME's and assistant simply did not always see *how* to use certain guidelines. For instance H1, the directional requirement to capture and hold the attention of the student, induced the designers to ask: "Yes but how?"

With respect to categories B (Interactions) and C (Design Patterns), the case studies supported the idea that common available question examples (stemming from secondary education) lead SME's and assistants too quickly come to conclude that "such questioning is not suitable for use in higher education". The content and perceived difficulty of such questions make it explicitly necessary to discriminate between the actual example and the concept underlying such examples to see their potential for use in higher education. This calls for extra mental effort and time, which often is not available in practice. Once new design patterns became available, the case studies in the last stages of the project revealed the value of design patterns: design patterns can have a greater impact on the conception of innovative digital questions than general guidelines and therefore should receive more attention in the methodology.

Secondly, certain guidelines were perceived as incurring additional costs, which were not balanced by the expectation of additional benefits. For instance, developing a case or a video and using it as the foundation for a question was said to involve too much effort in comparison to the expected benefits. This effect was increased by the fact that most project budgets were underestimated which sometimes was given as a reason to restrict design and development to the more simpler question formats (simple, text based MC questions) and not actively work on more elaborate design activities (such as A2, E3 or G), question types and media use. At the same time the formulation of distractors for traditional text based MC questions was in some case studies reported as being very time consuming in comparison to other design and development tasks and guidelines to avoid having to develop distractors were called for.

Thirdly, in a number of case studies, the SME's and assistants were of the opinion that a specific guideline was not relevant given the subject matter or that a certain guideline 'did not fit the purpose of the exam'. For example, physiologists stated that contradictions in their subject matter 'do not exist' (though of course they could design questions that use contradictions as foil answering options for example).

Fourth, in a number of case studies, the SME's and assistant were of the opinion that the role of the question (summative or activating) did not allow to use a specific guideline. In particular, for summative exams, Category B (Interactions) invoked, in a number of case studies, discussion on the scoring models of specific question types. How should questions involving multiple possible responses (such as Multiple Answer question, Matching questions, and Ordering questions) be scored? This uncertainty made SME's and assistant decide not to pursue the design of such questions.

Summarizing: *specific* guidelines were perceived to have different value depending on the subject matter, the role of the questions, time constraints and the competencies of the designers. Reasons not to use a specific guideline can be categorized under the following labels:

- Directions on how to use the guideline are lacking given the available team knowledge and skill.
- Cost-Benefit estimations of using the guideline were too high given the project conditions.
- The guideline is not relevant given the subject matter.
- The guideline is not relevant given the role of the questions. The guideline cannot be used until the question about transparent scoring is resolved.

## Intervention and input of the educational technologist

In case studies VU1, VU2, WU9, WU10 and TUD2, an ET helped the SME and assistants to gain more benefit of the guidelines by extra explication and demonstration and by selecting guidelines that could be most beneficial given the project constraints. Moreover, the ET could actually take successful part in the idea generation process when sufficient and adequate learning materials were available. In particular, the incorporation of various media in question design could be stimulated by the ET. When insufficient learning materials were available, it was very difficult for the ET to contribute to the design and development process. Thus, the actual involvement of the ET with the subject matter and the availability of learning materials is an important context variable for a successful contribution of an ET.

## Evaluation of the set of guidelines

As said, this article focuses on the development and evaluation of a set of guidelines for question design.

The case studies have confirmed that for the majority of teams, four to five guidelines are used and are perceived as worthwhile. Given the criterion that for a methodology, for any given team, a minimum of five guidelines must be useful, it is fair to conclude that the set of guidelines is a useful component within a methodology.

Second, the ALTB project wanted to investigate if question development teams can work with the complete set of guidelines in practice. From the case studies it becomes evident that

this is not the case. Simply presenting a set of guidelines had only very limited effect on the process. Offering some modest training and support increased the effect, but not substantially. It truly calls for a considerable effort by the team members for the guidelines to really have an impact on the quality of the design process and the quality of the questions that are developed. Most teams wanted a preselected set of three to five guidelines exactly targeted to their situation without having to select those themselves.

The third criterion that most of the guidelines would be applicable, irrespective of the intended role of the questions (summative or activating), is not met by the set of guidelines. Designing questions for the specific roles calls upfront for different sets of guidelines. A major discriminating factor for this is that for summative exams there is a lack of clear scoring rules for innovative question types and that emphasis is put on effective ways to develop multiple equivalent questions. For activating learning material, transparent scoring is less important and more emphasis must be put on engaging the learner more with the subject matter. In that respect, it is actually beneficial to use a wide variety of innovative closed question types.

## Conclusions

Literature provides little guidance for the initial stages of design and development of digital closed questions. This is an important reason to conduct research in these stages and develop specific tools to support the initial design process. One tool that is developed in the ALTB project is a set of guidelines focussing on the initial stages of design and development in order to boost creativity. This set of guidelines was presented to question design teams and used in 15 case studies. These case studies are described and summarized in this article.

### A set of guidelines is an inspirational source for question design but must be embedded in a broader approach

The developed set of guidelines offers inspiration to the majority of teams. There are always four or more guidelines available in the set that help question designers to find inspiration for question design. Within a broader methodology, the guidelines will certainly be appropriate.

From the case studies it is concluded that different set of guidelines should be compiled for the summative role or the activating role of questions. In the future, more and different guidelines will with no doubt emerge for the specific roles.

Furthermore, it has become clear that guidelines cannot function on their own. Design and development of digital closed questions requires specialized knowledge and skills. That can only be acquired through thorough study and practice. SME's and assistants need support to interpret and use the guidelines effectively. In particular SME's and assistants need help in selecting those guidelines which are most useful for them in their situation. Without such help, they loose focus and become frustrated.

### Design patterns have potential to be a powerful aid

The case studies revealed the value of design patterns: design patterns can have a great impact on the creative design of digital questions. They can be more effective than general guidelines or too general question examples. Draaijer and Hartog (2007) present – on the basis of the ALTB project – a detailed description of the concept of design patterns and a number of design patterns.

## A question design methodology must be geared towards educational technologists

Given the observed intricacy of question design and development, the conclusion is drawn in the ALTB project that a methodology must be geared specifically towards ET's. They must be able to use guidelines and design patterns in a variety of situations and domains to support SME's and assistants. A methodology should help an ET to select a few specific guidelines and a number of adequate design patterns in order to produce quick and effective results when working with SME's and assistants. The question of what procedures ET's can best act upon to perform that task is a matter for further research.

### Acknowledgements

The ALTB Project has been realized with support of SURF Foundation. SURF Foundation is the higher education and research partnership organisation for network services and information and communications technology (ICT) in the Netherlands. For more information about SURF Foundation: <http://www.surf.nl>.

## References

- Aegerter-Wilmsen, T., Coppens, M., Janssen, F. J. J. M., Hartog, R., & Bisseling, T. (2005). Digital learning material for student-directed model building in molecular biology. *Biochemistry and Molecular Biology Education*, 33, 325-329.
- Alexander, C. (1979). *The Timeless Way of Building*: Oxford Univ. Press.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives, the classification of educational goals – Handbook I: Cognitive Domain*. New York: McKay.
- Bull, J., & McKenna, C. (2001). *Blueprint for Computer-assisted Assessment*: RoutledgeFalmer.
- Dick, W., & Cary, L. (1990). *The Systematic Design of Instruction*. (Third Edition ed.): Harper Collins.
- Diederens, J., Gruppen, H., Hartog, R., Moerland, G., & Voragen, A. G. J. (2003). Design of activating digital learning material for food chemistry education. *Chemistry Education: Research and Practice*, 4, 353-371.
- Draaijer, S., & Hartog, R. (2007). Design Patterns for digital item types in Higher Education. *e-Journal of Instructional Science and Technology*, 10(1).
- Haladyna, T., M. (2004). *Developing and Validating Multiple-Choice Test Items* (Third Edition ed.). London: Lawrence Erlbaum Associates.
- Haladyna, T., M., (1997). *Writing Test Items to Evaluate Higher Order Thinking*. Needham Heights: Allyn & Bacon.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. In *Applied Measurement in Education* (Vol. 15, pp. 309-334): Lawrence Erlbaum Associates, Inc.

Hartog, R. (2005, december 2006). *Actief Leren Transparant Beoordelen*. SURF Foundation of the Netherlands, retrieved december 2006, from <http://fbt.wur.nl/altb>

Keller, J. M. (1983). Development and Use of the ARCS Model of Motivational Design (No. IR 014 039). Enschede: Twente University of Technology.

LaDuca, A., Staples, W. I., Templeton, B., & Holzman, G. B. (1986). Item modelling procedure for constructing content-equivalent multiple choice questions. *Medical Education*, 20(1), 53-56.

Ma, X. (2004). An investigation of alternative approaches to scoring multiple response items on a certification exam. University of Massachusetts Amherst, Massachusetts.

Mazur, E., & Crouch, C. H. (2001). Peer Instruction: Ten Years of Experience and Results. *American Journal of Physics*, 69(9), 970-977.

Merrïenboer, J. J. G., van, Clark, R. E., & Croock, M. B. M., de. (2002). Blueprints for complex learning: the 4C/ID- Model. *Educational Technology Research and Development*, 50(2), 39-64.

Mills, C. N., Potenza, M. T., Fremer, J. J., & Ward, W. C. (2002). *Computer-Based Testing, Building the Foundation for Future Assessments*. London: Lawrence Erlbaum Associates.

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer-Verlag.

Paulus, P. B., & Brown, V. R. (2003). *Enhancing ideational creativity in groups: Lessons from research on brainstorming*. Oxford: Oxford University Press.

Roid, G. H., & Haladyna, T. M. (1982). *A Technology for Test-Item Writing*. Orlando, Florida: Academic Press.

Roosink, H. J., Bonnes, H. J. G., Diepen, N. M., van, & Moerkerke, G. (1992). Een werkwijze om tentamenopgaven te maken en tentamens samen te stellen (No. 73): Universiteit Twente.

Scalise, K., & Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. *The Journal of Technology, Learning and Assessment*, 4(6).

Simon, H. A. (1994). The bottleneck of attention: connecting thought with Motivation. In W. D. Spaulding (Ed.), *Integrative views of motivation, cognition and emotion*. (Vol. 41, pp. 1-21). Lincoln: University of Nebraska Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wilbrink, B. (1983). *Toetsvragen schrijven* (Vol. 809). Utrecht/Antwerpen.

## Appendix 1: Overview of case studies

Case	Course Level	Course Subject	Role of the questions	Software	Development team	
1	WU1	Master	Food Safety (Toxicology/ Food Microbiology)	summative	QM	SME and assistant
2	WU2	Master	Food Safety Management	activating	Bb	SME and ET
3	VU1	2nd year	Heart and Blood flow (physiology, ECG measurement and clinical ECG interpretation)	diagnostic and summative	QM	SME and ET
4	VU2	3rd year	Special Senses (vision, smell, hearing, taste, equilibrium)	summative	QM	SME and ET
5	TUD1	3rd year	Drinking water treatment	activating	Bb	SME and assistant
6	WU3	Master	Epidemiology	summative (open book)		SME and assistant
7	TUD2	3rd year	Sanitary Engineering	activating	Bb	SME and assistant and ET
8	WU4	Master	Food Toxicology	summative	QM	SME and assistant
9	WU5	Master	Food Micro Biology	activating	Bb	assistant
10	WU6	Master	Advanced Food Micro Biology	activating	Bb	assistant
11	WU7	Master	Food Chemistry (general introduction module for candidate students)	diagnostic	QTI delivery	SME = ET
12	WU8	Master	Food Toxicology	diagnostic	QM	SME and assistant
13	WU9	Master	Sampling and Monitoring	diagnostic (self - )	Flash	SME and Assistant and ET and Flash programmer

14	WU10	Master	Food Safety Economics		summative (not open book)	Bb and on paper	SME assistant and ET
15	FO1	1st year	Curriculum: Sciences	General	Diagnostic-'pluss'	N@tschool	SME's question specialist and entry

(WU = Wageningen University, VU = Vrije Universiteit Amsterdam, TUD = University of Technology Delft, FO = Fontys University of Professional Education, QM = Questionmark Perception, Bb = Blackboard LMS, QTI = Question and Test Interoperability 2.0 format, N@tschool = N@tschool LMS, SME = Subject Matter Expert such as lecturer, professor, instructor, ET = Educational technologist, Assistant = recently graduated student or student-assistant)

## Appendix 2 Overview of cases and the use or non-use of guidelines

Case	Role	Development team	Initially available material	Which Guidelines Used And How	Summary of case report
WU1	summative	SME and assistant	<ul style="list-style-type: none"> <li>• Toxicology Part</li> <li>• Lecture notes</li> <li>• Handouts of Presentations</li> <li>• Detailed learning objectives in natural language</li> <li>•</li> <li>• Food Microbiology Part</li> <li>• Handouts of Presentations</li> <li>• Articles</li> </ul>	<ul style="list-style-type: none"> <li>• E1, C1</li> </ul>	<ul style="list-style-type: none"> <li>• Given the intended role and task of the designer the need of guidelines for design became very apparent.</li> <li>• A comprehensive overview of guidelines which are useful in the domains of the ALTB project at the level of higher education could not be found.</li> <li>• For summative testing the contour of a new guideline became visible : next to designing one question, design 4 equivalent questions using the guidelines for ‘parallel design and development’</li> <li>• Useful guidelines for ‘parallel design and development’ are                             <ul style="list-style-type: none"> <li>• E1 Use a list of detailed learning objectives</li> <li>• C1 Use a list of generic item shells</li> </ul> </li> </ul> <p>Remarks:</p> <ul style="list-style-type: none"> <li>• The guidelines E1 and C1 came available during the introductory workshop that the assistant attended.</li> </ul>
WU2	activating	SME and assistant	<ul style="list-style-type: none"> <li>• Documents and reports</li> <li>• Examples of Cases and questions in Blackboard</li> <li>• Experience in the team with guidelines for activating learning materials</li> <li>• Literature on guidelines for the</li> </ul>	<ul style="list-style-type: none"> <li>• No conscious use of guidelines</li> <li>• Implicit use of A1</li> </ul>	<ul style="list-style-type: none"> <li>• Designer/developer gave most attention to development of cases and to formulation of extended feedback.</li> <li>• The most pressing need felt by the designer/developer was not the need for design guidelines</li> <li>• The designer/developer needed more and better sources on more subject matter knowledge and input with respect to professional experience</li> </ul>

Case	Role	Development team	Initially available material	Which Guidelines Used And How	Summary of case report
VU1	diagnostic and summative	SME and ET	<p>design and development of activating learning materials</p> <ul style="list-style-type: none"> <li>Guidelines A1, C1, C2, (design patterns), G (scan sources)</li> <li>During the inspiration session, no material was available.</li> <li>Later on, material was available in the form of:                             <ul style="list-style-type: none"> <li>Previous Exams</li> <li>Physiology textbook</li> <li>Complete set of guidelines was available</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>The following guidelines were not used: A2, C3, C5, C6, F.</li> <li>All other guidelines were used.</li> </ul>	<ul style="list-style-type: none"> <li>The bare availability of guidelines is not sufficient to induce the use of guidelines.</li> <li>All guidelines were systematically discussed and ‘forced-fitted’ to use in two rounds of ‘inspiration sessions’ in which an ET guided a question design session.</li> <li>The subject matter and the learning objectives allow for the definition of authentic cases and authentic ‘what to do’ questions. Thus, the instructor was already used to apply guideline A1. Guideline A2 was evaluated as too labour intensive to execute and not appropriate for the course. The SME was of the opinion that guidelines A3.1 to A3.2 actually defined instructional content and should not define exam content. Guidelines A3.4 and A3.5 provided some inspiration. Guideline A4 could be used.</li> <li>Guidelines B1, B2, B3.1 really invoked enthusiasm. Example questions presented by ET resulted in ideas on new questions. However, problems with unclear scoring rules diminished enthusiasm.</li> <li>C1 was felt to be very useful too, but so straightforward that it was not used during the inspiration session. C2 looked promising but turned out to be difficult to handle. C3, C5 and C6 were not regarded as useful because it was felt to be difficult to develop univocal problems and answer sets. However, if the questions were intended for active learning, the SME was of the opinion that they were very useful. C4 offered opportunity for question generation. G (search for extra sources on the internet) was very worthwhile for the instructor, based on the extra source the Educational technologist retrieved for him). It resulted in a collection of pointers to useful cases, graphics and multimedia elements.</li> <li>Guidelines F (take mindset of student as starting point) were not used because the instructor was of the opinion that any assumption</li> </ul>

Case	Role	Development team	Initially available material	Which used And How	Guidelines	Summary of case report
VU2	summative	SME and ET	<ul style="list-style-type: none"> <li>• During the inspiration session, no material was available.</li> <li>• Later on, material was available in the form of:                             <ul style="list-style-type: none"> <li>• Previous Exams</li> <li>• A course website with digital materials and cases.</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• The following guidelines were not used: A2, C3, C5, C6, F4 and H.</li> <li>• All other guidelines were used.</li> </ul>	<ul style="list-style-type: none"> <li>• Directional requirements H were not used. They were considered relevant, but not helpful. ( “aim for attention – yes but how”)</li> <li>• Guidelines D (textbooks) was considered an ‘too obvious’ (“how else can you start developing questions”)</li> <li>• Directional requirements E (learning objectives), I (validity), J (equivalence) were felt to be ‘too obvious’ also. They were used all the time but were not considered to provide inspiration.</li> <li>• G3 and G4 were used in the form of the ‘inspiration session’.</li> <li>• The instructor preferred to be offered a much smaller dedicated selection of guidelines. Also the overlap between guidelines should be avoided.</li> </ul> <p>Bottom line:</p> <ul style="list-style-type: none"> <li>• Offering guidelines to question designer in an intensive inspiration session results in questions of types that are new for the course and for the SME</li> <li>• Especially discussing example questions is considered worthwhile.</li> <li>• The ET is an enabler for a greater divergence of questions conceived</li> </ul>	<p>about the mindset of the students would apply to a very limited part of the student population and would introduce bias.</p> <ul style="list-style-type: none"> <li>• All guidelines were systematically discussed and ‘forced-fitted’ to use in two rounds of ‘inspiration sessions’ in which an ET guided a question design session (see also case VU1)</li> <li>• Guidelines result in new types of questions as in case VU1.</li> <li>• Comments about the use of authentic cases as in case VU1. This SME normally develops cases as follows: medical specialists deliver questions; the SME edits them and combines them in such a</li> </ul>

Case	Role	Development team	Initially available material	Which used And How	Guidelines	Summary of case report
TUD 1	activating	SME and assistant	<ul style="list-style-type: none"> <li>• Textbook with many photos, graphs, diagrams, examples, explicit calculations, exam questions with answers</li> <li>• Hand-outs of Presentations</li> <li>• Hand-outs of Lecture Notes</li> <li>• The complete set of guidelines was available</li> </ul>	<ul style="list-style-type: none"> <li>• The complete set of guidelines was available.</li> </ul>	<ul style="list-style-type: none"> <li>• H1, H2, C4 and D1.8 B2 and D1.2 were used most by the <b>assistant</b>.</li> <li>• A1, B3.5, D1.10, and I3 were used most by the <b>SME</b>.</li> </ul>	<p>way that a case is the result.</p> <ul style="list-style-type: none"> <li>• B1, B2, B3 were felt useful, but would not be used by the instructor unless she could rely on the sustained support and input of the ET.</li> <li>• The assessment of the guidelines C, G, D, E, G and I and J was similar to that of case VU1.</li> <li>• With respect to F (students' mind set): The instructor was already used to design questions that relate to students daily life and experiences</li> <li>• The instructor felt that requirement H (motivation) was not really necessary, though in practice she actually used it to 'spice up' the final exam (and that is guideline F).</li> <li>• Guidelines A* were not used by the student assistant because she did not have sufficient professional experience and because the SME could 'take' the tasks that are related to these guidelines.</li> <li>• Guidelines A1* on cases were not used because the SME wanted to cover all subject matter</li> <li>• The main determinants for the use of specific guidelines were               <ul style="list-style-type: none"> <li>• the role of the questions,</li> <li>• the extent of professional experience,</li> <li>• the characteristics of the subject matter.</li> </ul> </li> </ul> <p>The use of a number of guidelines can be recognized but the case study did not provide positive evidence about any added value of presenting a set of guidelines to the designers/developers.</p> <p>Bottom line:</p> <ul style="list-style-type: none"> <li>• Many guidelines were considered 'too obvious'</li> </ul>

Case	Role	Development team	Initially available material	Which Guidelines Used And How	Summary of case report
WU3	summative (open book)	SME and assistant	<ul style="list-style-type: none"> <li>Textbook</li> <li>Hand-outs of Presentations</li> <li>A large set of MC questions, mostly based on 2 propositions</li> <li>The complete set of guidelines including initial experience with the guidelines</li> </ul>	<ul style="list-style-type: none"> <li>J 4.1</li> <li>C1, C2, C3, C4, C7, D1i, G5</li> </ul>	<ul style="list-style-type: none"> <li>For almost every guideline that was not used there was a good reason not to use that guideline.</li> <li>Guidelines that cannot be used in a specific design and development project for a good reason should not be offered in that project.</li> <li>Systematically scanning inspirational dimensions did not work</li> <li>The directional requirement to design a set of equivalent questions for each detailed learning goal was considered to be crucial.</li> <li>Textbook (guidelines D) and other sources like internet and journals (guideline G5) were scanned for inspiration.</li> <li>Guidelines C3 and C4 were relatively useful for design and development of questions of a different format.</li> <li>Guideline I was used unconsciously whenever the questions were discussed with the SME.</li> </ul> <p>Main conclusion:</p> <ul style="list-style-type: none"> <li>The guidelines do hardly result in new question types for the course/instructor</li> <li>The guidelines do hardly result in quicker or more efficient design of questions</li> <li>Remark: The summative test is an open book exam, which made it more difficult to design questions. Developing questions which are directly based on text of the book is not an option; questions needed to be formulated in a different way or should test application.</li> </ul>
TUD 2	activating	SME and assistant	<ul style="list-style-type: none"> <li>Textbook with many examples, graphs, open questions.</li> <li>Exam questions, answers to questions</li> </ul>	<ul style="list-style-type: none"> <li>C2, C3, C4, and Ci were i denotes any new design pattern that was</li> </ul>	<ul style="list-style-type: none"> <li>Focus on design patterns results in new questions and more use of question types other than True/False and MC</li> <li>Guidelines A1 and A2 are not considered because cases are supposed to direct too much attention of the student to a small part</li> </ul>

Case	Role	Development team	Initially available material	Which used And How	Guidelines	Summary of case report
		and ET	<ul style="list-style-type: none"> <li>The textbook was authored by the chair group sanitary engineering.</li> <li>Also the pictures in the textbook where available electronically</li> <li>Additional handouts of presentations</li> <li>Lecture notes</li> <li>Relevant Websites</li> <li>The complete set of guidelines was available.</li> </ul>	<ul style="list-style-type: none"> <li>not yet listed</li> <li>Di where i denotes any of the textbook components or questions inspired by textbook components</li> <li>E was used implicitly as the textbook covered E.</li> <li>G3 (ET)</li> </ul>		<p>of the subject matter that has to be covered according to the definition of the course.</p> <ul style="list-style-type: none"> <li>As it was agreed that the consultant would take the lead also A3 did not get much attention</li> <li>B1 had already been done in the previous project</li> <li>Once more scanning B2 was not inspiring</li> <li>B3.2 (sound) and B3.4 (video) were not considered because of capacity constraints</li> <li>A number of new design patterns were used. These patterns will be presented in a publication on design patterns.</li> <li>D9 (abstract concepts) and D10 (what to remember) were not considered</li> <li>F (prior knowledge of student as starting point) was not considered useful by both the lecturer and the question designer</li> <li>G1,2,4,5 were not used because of time constraints</li> <li>H was not considered useful by the lecturer and the question designer</li> <li>I was used implicitly whenever a suggestion of the consultant had to be discussed. I was also implicit in the textbook</li> <li>J is not relevant for activating learning material</li> <li>Presenting design patterns and focussing on design patterns was much more effective in generating a variety of innovative questions than presenting guidelines or inspirational dimensions.</li> <li>The design patterns sometimes ‘use’ one guideline but often ‘use’ more guidelines</li> </ul>
WU4	summative	SME and	<ul style="list-style-type: none"> <li>Lecture notes</li> </ul>	<ul style="list-style-type: none"> <li>C1-3, D1i, J4.1</li> </ul>		<ul style="list-style-type: none"> <li>C1, C2 and J4.1 were felt to be useful to create equivalent exams.</li> <li>The guidelines D were used in the sense that the learning material</li> </ul>

Case	Role	Development team	Initially available material	Which used And How	Guidelines	Summary of case report
		assistant	<ul style="list-style-type: none"> <li>Hand-outs of presentations</li> <li>Articles</li> </ul>			<p>is scanned for inspiration.</p> <ul style="list-style-type: none"> <li>Directional requirements F (students), H (motivating) and I (validity) are used but are not considered to provide inspiration.</li> </ul>
WU5	activating	assistant	<ul style="list-style-type: none"> <li>Textbook</li> <li>Handouts of presentations</li> </ul>	<ul style="list-style-type: none"> <li>C1 and C3, D, E</li> </ul>		<p>Remark: The exam was to be digital. Technical and organisational aspects required much attention of Question Designer as well</p> <ul style="list-style-type: none"> <li>The guidelines D were used in the sense that the learning material is scanned for inspiration.</li> <li>Guidelines concerning the interaction types (B) were used unconsciously as already a lot of experience had been gained by developing other questions.</li> <li>The guidelines F (students), H (motivating) and I (validity) are seen as important issues that require attention but that are not concerned to provide inspiration. (“Yes but HOW”)</li> <li>J is not relevant for activating learning material.</li> </ul>
WU6	activating	assistant	<ul style="list-style-type: none"> <li>Handouts of presentations</li> <li>Articles</li> </ul>	<ul style="list-style-type: none"> <li>C1 and C3, E, G1 and G5</li> </ul>		<ul style="list-style-type: none"> <li>Guidelines concerning the interaction types (B) were used unconsciously as already a lot of experience had been gained by developing other questions.</li> <li>The guidelines F (students), H (motivating) and I (validity) are seen as important issues that require attention but that are not concerned to provide inspiration.</li> <li>J is not relevant for activating learning material.</li> <li>As there was no textbook guidelines D were not really helpful, but instead guidelines G1 and G5 were.</li> </ul>
WU7	diagnostic	SME = ET	<ul style="list-style-type: none"> <li>Textbook</li> <li>many examples of closed questions for Food Chemistry in FLASH though often not specifically for</li> </ul>	<ul style="list-style-type: none"> <li>Guidelines that were mainly used : B1, B2, B3, C3, C4, C7, D1.i except D1.7, E1,</li> </ul>		<ul style="list-style-type: none"> <li>The SME/ET could clearly explain why she did not use the following guidelines:</li> <li>A1 (cases) was difficult to match with the test matrix</li> </ul>

Case	Role	Development team	Initially available material	Which used And How	Guidelines	Summary of case report
			exactly the same subject matter	E2, E3, F.i, , G1, G3, G5, H2, H4, I1 and I3		<ul style="list-style-type: none"> <li>• A2.i (LaDuca) did not match the purpose of the diagnostic test</li> <li>• A3.1 (tips, tricks) did not match the purpose of the diagnostic test</li> <li>• A3.2 (surprise in profession) incidentally provided inspiration</li> <li>• C1 and C2 did not match the purpose of the diagnostic test C1 and C2 are actually not very useful unless one wants to develop a set of exams</li> <li>• C5 and C6 (for designing and developing text based questions) did not match very well the subject matter</li> <li>• D1.7 (exceptions) did not help at all. In the related courses it is not usual to pay attention to exceptions</li> <li>• E4 (target competencies) was not yet useful because the target competencies are only defined at curriculum level and articulating them at the course level is considered to be a task that does not fit within the scope of the project.</li> <li>• Fi (students) were all used but F1 and F2 more than F3 and F4</li> <li>• G2 (ask content experts) and G4 (brainstorm sessions) were not used because not within the budget.</li> <li>• H1 (gain attention) and H3 (aim for confidence) did not strongly match with the purpose of the questions</li> <li>• I2 was not used</li> </ul> <p>Bottom-line</p> <ul style="list-style-type: none"> <li>• A very experienced designer can use about two thirds of the guidelines and can give a clear explanation of any reasons not to use a specific guideline.</li> </ul>
WU8	diagnostic	SME and	<ul style="list-style-type: none"> <li>• Detailed list of learning objectives</li> </ul>	<ul style="list-style-type: none"> <li>• New guideline “cluster of five”,</li> </ul>		<ul style="list-style-type: none"> <li>• Content Expert already had gained some experience in case WU1</li> <li>• Quickly decided to focus on MC, MA, ordering, match and fill-in-</li> </ul>

Case	Role	Development team	Initially available material	Which used And How	Guidelines	Summary of case report
		assistant	<ul style="list-style-type: none"> <li>Lecture Notes</li> <li>Handouts</li> </ul>		E2, I1	<p>the-blank and not to use any diagrams or pictures. Subject matter does not require such diagrams</p> <ul style="list-style-type: none"> <li>Quickly decided to use new guideline (“design and develop cluster of five equivalent questions approach”)</li> <li>Questions were designed in MS Word, later formulated by technical assistant in QTI 2.0</li> <li>Most design guidelines were not used</li> </ul>
WU9	diagnostic (self - )	Assistant and ET	<ul style="list-style-type: none"> <li>Scientific articles</li> <li>Learning Material that was designed and developed in parallel with the design of closed questions</li> </ul>			<ul style="list-style-type: none"> <li>Initial confrontation with the complete initial set of guidelines resulted in very limited use</li> <li>On basis of that it was agreed to focus on the following subset : B2 interaction types – B 3.4 graphs – B 3.5 diagrams– B 3.6 process diagrams– C 3 completion - C 4 introduce error – D systematically scan learning material (self developed) – G2 ask food safety experts – G5 other sources – H1 capture attention E use detailed learning objectives</li> <li>Together with an educational technologist, new design patterns were developed</li> <li>The educational technologist presentation that covered most of the subject matter and this presentation contained a wealth of diagrams and figures to be used as foundation for closed questions</li> <li>New design pattern: match symbols in a given equation with data in a given problem description. Thus understanding of operational semantics of an equation can be separated from the ability to execute a calculation</li> <li>Technical implementation was delegated to a FLASH programmer.</li> <li>Questions developed in MS Word and MS PowerPoint</li> </ul>
WU10	summative	SME and assist	<ul style="list-style-type: none"> <li>Lecture Notes</li> <li>Articles</li> </ul>		<ul style="list-style-type: none"> <li>Guidelines that were mainly used: A1 and A4,</li> </ul>	<ul style="list-style-type: none"> <li>Focus by ET on design patterns (guidelines C) that imply the use of pictures</li> </ul>

Case	Role	Development team	Initially available material	Which used And How	Guidelines	Summary of case report
	(not open book)	ant and ET	<ul style="list-style-type: none"> <li>• Handouts of Presentations</li> <li>• The handouts include many diagrams and graphs and other pictorial information</li> <li>• The handouts include many procedures and computations</li> <li>• Computer Practical instructions</li> </ul>	C3, D		<ul style="list-style-type: none"> <li>• Not limited to the few design patterns that were initially available. Result: Many more design patterns were conceived.</li> </ul> <p>Preliminary conclusion:</p> <ul style="list-style-type: none"> <li>• The combination of:                             <ul style="list-style-type: none"> <li>• availability of many digitized diagrams, graphs and other pictures</li> <li>• many computations and corresponding chains of inference</li> <li>• many questions</li> <li>• high degree of involvement of the content expert/instructor</li> </ul> </li> <li>• is in keeping with the hypothesis that - the more conditions are satisfied the more guidelines are useful and the better a condition is satisfied the more one tends to focus on the guidelines that match this condition</li> <li>• In this case study many PowerPoint slides formed an obvious basis for a question.</li> <li>• In particular application of guidelines D in combination with C and some new design patterns was effective.</li> <li>• Guidelines A1 and A4 were followed to develop cases. A2 and A3 were not useful as the question designer did not have practical experience.</li> <li>• I was used unconsciously whenever the questions were discussed with the content expert</li> </ul>
FO1	Diagnostic	SME's	<ul style="list-style-type: none"> <li>• Textbook(s)</li> </ul>		A1	<ul style="list-style-type: none"> <li>• Only guideline A1: (develop cases) was used</li> <li>• When the initial set of guidelines was presented representatives of the team indicated that they would not adopt these guidelines</li> <li>• Fundamental critique was</li> <li>• that the presented guidelines suggested too much focus on</li> </ul>

Case	Role	Development team	Initially available material	Which used And How	Guidelines	Summary of case report
						<p>individual questions instead of sets of questions</p> <ul style="list-style-type: none"> <li>• that the set of guidelines killed creativity</li> <li>• It was agreed to develop 30 questions and record what alternative guidelines were actually used. The team however did not succeed in formulating any alternative guideline.</li> </ul>