Chapter 2: Developing Test Items as Solving an Ill-defined Design Problem

Silvester Draaijer\textsuperscript{a}, Judith Schoonenboom\textsuperscript{b}, Jos Beishuizen\textsuperscript{a}, Lambert Schuwirth\textsuperscript{c}

\textsuperscript{a}Department of Research and Theory in Education, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, The Netherlands. \textsuperscript{b}Department of Education, Faculty of Philosophy and Education, University of Vienna, Austria. \textsuperscript{c}Prideaux Research Centre, School of Medicine, Flinders University, Adelaide, Australia and Department of Educational Development and Research, Maastricht University, The Netherlands.

Manuscript submitted for publication
2 Developing Test Items as Solving an Ill-defined Design Problem

2.1 Abstract

In higher education, creating test items is a task that teachers have to perform regularly, and there have been reports of the need for support for teachers to improve their performance in this area. A cognitive process model for test item design has been developed to describe and explain the test item design process and explore opportunities for improving performance on the test item design task. The model is based on the assumption that test item design is an ill-defined creative design problem-solving task that requires extensive divergent and convergent production of ideas, in which both solutions to and perspectives on problems evolve together. The developed model is a general model that allows for incorporating other methods of problem solving, especially methods for generating test items that involve both divergent and convergent production of those items.

Keywords: test item writing; problem solving; higher education; assessment
2.2 Introduction

In this chapter, a cognitive process model describing the design process of selected response test items is presented. In the model, the concept of divergent and convergent production of ideas (Acar & Runco, 2012; Guilford, 1967) is central. Divergent production refers to generating new and original ideas, while convergent production refers to selecting and optimizing ideas. This model aims to facilitate the understanding of the process of test item design, which should in turn lead to improved methods that raise the quality of outcomes of test item design tasks.

Studying the test item design process is important in many ways, but of particular importance in higher education because test items are used throughout that milieu. Test items determine to a large extent what students perceive as being important for their area of study and future profession and consequently what they learn (Gerritsen-van Leeuwenkamp, 2012; Gielen, Dochy, & Dierick, 2003; Snyder, 1971). In general, teachers are responsible for producing these test items and therefore these items are typically called in-house developed items. At the same time, the task of high-quality test item design is reported to be difficult (Case, Holtzman, & Ripkey, 2001; Haladyna, 2004; Mayenga, 2009), while the quality of many test items found in higher education is reported to be low (Hansen & Dexter, 1997; Jozefowicz et al., 2002; Tarrant, Knierim, Hayes, & Ware, 2006). Quality suffers particularly when teachers attempt to design test items that move beyond querying for decontextualized verbatim information from instructional materials. However, teachers should be able to design test items for a variety of purposes that are suited to particular goals, ranging from test items for rote learning purposes to test items that query higher-order learning goals such as application of knowledge, critical thinking, and problem solving. Being able to design such a wide range of test items well requires specific knowledge and skills. As it is further reported that teachers are often left to their own devices in designing test items (Osterlind, 1998), better support for teachers is clearly needed.

However, despite much research regarding test item design, it is not at all clear what designing test items actually entails. The design of test items by teachers is largely a black box process. While in specific fields involving cognitive modelling
some progress has been reported (Gierl, Zhou, & Alves, 2008), there is an overall lack of research and theory building regarding test item development (Downing & Haladyna, 2006; Hattie, Jaeger, & Bond, 1998; Welch, 2006). What it particularly entails from the perspective of the cognitive processes exhibited by teachers when generating such test items is unclear. According to Haladyna, “the cognition that is used in writing the item is still a mystery” (personal communication, May 15, 2013). However, if we want to be more effective in supporting teachers’ efforts to design good test items, we are likely to design more effective support for teachers if we understand the processes involved in designing test items more clearly and thoroughly. This paper therefore seeks to study, describe, and model this cognitive process so that it can be used to develop support for the design of more effective and valuable test items. Subsequently, selected support materials and interventions are developed and studied for their effectiveness.

2.3 Test item development as an ill-defined design problem

The starting point for our model is to view the design of test items as the process of solving an ill-defined design problem. To unravel this view, the test item creation task is considered first from the perspective of its ill-definedness and then from the perspective of design. These two perspectives are presented below.

Though many views and definitions exist regarding the characteristics of ill-defined problems (Lynch, Ashley, Pinkwart, & Aleven, 2009; Newell & Simon, 1972; Simon, 1973), we use Reitman’s (1964, 1965) three classic key characteristics of ill-defined problems to argue that the task of designing test items is ill-defined. These three key characteristics are the degree to which the start state, goal state, and transformation function are unspecified in relation to the problem to solve. The start state (also called the initial problem state) refers to the starting point for solving a problem. The goal state refers to the endpoint of the problem-solving process, in which a solution is established that complies with the specifications and constraints that make it an appropriate solution to the problem. The transformation function refers to the constraints regarding the methods, rules, and instruments used to solve the problem. If any of these three characteristics is not well delineated, contains ambiguous information or rules, or allows for excessive
freedom in specification, a problem is ill-defined. This phenomenon is articulated below for the problem of test item design.

First, the start state of test item development in higher education is not well specified. As an initial example, consider content validity. Teachers can only strive for a sufficient degree of content validity, but the definition of sufficient is not clear. Learning objectives, topics, and accompanying test blueprints in which the number and cognitive level of test items to include are preordained (Downing & Haladyna, 2006) leave unspecified room for how many or what specific topics are to be included in a test blueprint or the cognitive level (Bloom, 1956). The process is thus open to individual or collective interpretation and negotiation and is sometimes intentionally abstractly or quite broadly defined to enable a solution space for test items to be developed. Validity concepts in general are not well specified and add to the ill-defined nature of the start state. Even more importantly, there is developing scientific theory and meaningful debate concerning the very concept of validity (Borsboom, Mellenbergh, & Van Heerden, 2004; Cronbach & Meehl, 1955; Kane, 1992; Messick, 1988) and whether it should, for example, include considerations regarding the nature of the consequences of test results for specific tests. As scholars are already debating the concept, such conceptions and debates are likely beyond easy comprehension for teachers in higher education charged with designing achievement tests, and cannot provide clear starting points for test item design. Bloom’s (1956) taxonomy of cognitive behavior serves as a second example; it is often proposed to guide the test item design process to link subject matter with students’ cognitive behavior such as remembering, understanding, or applying facts and concepts within the subject matter. However, because Bloom’s taxonomy is quite general, it has been reported that test item designers often experience ambiguity in the definitions and descriptions of the taxonomy when trying to develop test items for their particular domains and contexts. Multiple articles have pointed out problems and limitations (Crooks, 1988; Seddon, 1978) or a lack of agreement between teachers or judges as to the category in which a test item belongs (Fairbrother, 1975; Seddon, 1978). Using Bloom’s taxonomy therefore hardly leads automatically or univocally to test items.
of a specific cognitive category and does not alleviate the ill-defined character of the start state.

This brings us directly to Reitman’s second key characteristic, the goal state. The goal state in test item design is open-ended. First, tests as a whole are considered psychometrically to be a measurement of latent rather than directly observable traits (Spearman, 1904). In such a measurement, single items in a test provide only limited information regarding the trait and inherently contain error; this error can take many forms, largely due to the interactive nature of test items. This interactive nature means that test item designers and respondents have different perspectives on the meaning of information and even on specific wordings (Schober, 1999). For example, less competent students experience the nature of test items differently than more competent students (Beullens, Struyf, & Van Damme, 2005; Schuwirth, Van der Vleuten, & Donkers, 1996). Further, students can apply a number of different cognitive strategies to answer test items (Hibbison, 1991; Schuwirth et al., 1996), and it impossible to know in the practice of higher education which students use what answering strategies and whether they fit with the teachers’ intended answering strategies. Schober (1999) suggests in that respect the need for an interactional approach to question design. As an additional example of this interactional nature, studies of Angoff standard-setting procedures for tests (Impara & Plake, 1998; Verheggen, Muijtjens, Van Os, & Schuwirth, 2008; Verhoeven, Verwijnen, Muijtjens, Scherpbier, & Van der Vleuten, 2002) can be put forward. These studies show that in those procedures teachers have to estimate the ability for the “barely proficient” student to answer test items correctly, and there is often limited agreement among teachers. This is evidence that teachers have different, and possibly idiosyncratic, perceptions about the nature, level of difficulty, and ability of their target group of students in relation to characteristics of test items. A final example of ill-definedness relates to the extent of adherence to editorial guidelines for test item design that may involve a number of personal or institutional convictions regarding the use of specific item formats (Rodriguez, 2003) or terminology. These decisions are often negotiable and more often than not reside in the “hidden” curriculum (Snyder, 1971). This interactive nature of test items makes the goal state for test items inherently ill-defined. A final facet of the
ill-defined nature of the goal state lies in the fact that multiple test items may be appropriate for a specific topic or blueprint cell of a test. Overall, it is a serious challenge to determine when a test item is “finished” or of “sufficient” quality, particularly before gathering data on the basis of administered tests, which renders the goal state of the test item development process unquestionably poorly specified.

Finally, the transformation function can also be regarded an ill-defined feature of test item design, because many different processes and routes are available for creating test items. Conceiving test items is not governed by clear rules or algorithms. While some prescriptive and detailed methods and heuristics may exist (Roid & Haladyna, 1980), teachers in higher education are free to choose any approach they see fit, all of which may lead to good test items.

The fact that multiple solutions are possible and that exploring alternatives is needed to solve a problem is characteristic of a design task (Goel & Pirolli, 1992; Ulrich, 2011). To characterize design and designers, Simon (1996, p. 111) states that “the intellectual activity producing physical artefacts is not different fundamentally from the one [ … ] that devises a new sales plan for a company or a social welfare policy for a state.” Rowland (1993) presents instructional design as a form of problem solving and design and Thomas and Carrol (1979) explicitly called letter writing, which resembles test item design to a certain degree, a mini-design activity.

The arguments above make it clear that the test item design task is indeed an ill-defined design problem. This provides us with a fundamentally new perspective on the test item design task that enables a new description of that task, culminating in what we call a cognitive process model. The model builds on existing theory and findings regarding solving ill-defined design problems to underpin the model.

2.4 Findings and relevant process models from design research and problem-solving research

To set the stage for describing the design process, we turn first to research regarding processes employed by experienced designers, before a later discussion of the implications for novices. The research on design suggests that experienced
designers work according to a knowledge-building cycle (Jonassen, 2004; Robinson, 1986) in which the designer generates problem descriptions and perceptions, proposes hypotheses in the form of possible solutions, tests them, and develops arguments to support or reject them (Jonassen, 2004; Robinson, 1986), thus strengthening the perception of the problem. Experienced designers rely on rules of thumb (Rowland, 1993), heuristics (Mulet & Vidal, 2008), conjecture (Lawson, 2006; Tripp, 1994), design patterns (Alexander, 1979), and rapid prototyping (Tripp & Bichelmeyer, 1990; Visscher-Voerman & Gustafson, 2004) to find the best solutions. These approaches help designers keep the problem and solution spaces within manageable limits. Experienced designers use their previously acquired expertise in a flexible way (Rowland, 1993) in which the problem space (the perception, representation, and formulation of the problem to solve) and the solution space (the already generated and still possible ideas for solutions) evolve simultaneously during problem solving (Dorst & Cross, 2001; Goel & Pirolli, 1992). This process is called co-evolution and is considered a central part of the nature of design (Maher & Poon, 1996). More particularly, co-evolution is a process during which newly developed solutions instantiate new perspectives on the fundamentals and details of the problem and thus lead to novel ideas. With co-evolution, the problem space and solution space both evolve during an iterative development process, as shown in Figure 1.

![Figure 1](image.png)

Figure 1. Representation of the co-evolution of problem space and solution space based on Maher and Poon (1996).

During the iterations, the designer exhibits a cognitive process of diverging and converging operations (Acar & Runco, 2012; Pretz, Naples, & Sternberg, 2003). Guilford (1967) introduced the concepts of divergent and convergent thinking and
From a cognitive psychology perspective. Divergent production refers to the originality (uniqueness), fluency (number of ideas), flexibility (number of categories), and elaboration (amount of detail) of the ideas that an individual typically produces. By contrast, convergent thinking and production refer to optimization, detecting and repairing errors, logic, recognizing the familiar, technical detailing, and using standard solution strategies for problem solving. Convergent thinking bears on narrowing possibilities and zeroing in on solutions to find the single best solution.

2.5 How co-evolution and divergent and convergent production are manifest in test item design

Viewing design as an ill-defined creative problem-solving task as described above enables us to craft a more detailed description of the way in which test items are created. There are many different routes and processes available, as the start state, goal state, and transformation function are all nonspecific, so what follows is a general characterization of how the process is likely to advance.

As a first step, a test item writer starts with divergent production of ideas. The goal is to generate initial test item ideas that contain original information, such as making conjectures for a test item based on a fact, concept, principle, or procedure in the instructional text, course objective, or blueprint cell (Williams & Haladyna, 1982), or on the basis of already existing comparable test items. Ideas that come about by chance can also be included (Olsen & Bunderson, 2004). A test item writer often creates paraphrased versions of the instructional material to avoid calling on simple recognition and recall of information. A test item writer can perform this process purely mentally, can write down ideas as they arise, or perhaps employ concept maps or other representations. Test item writers are aware that misconceptions can serve as starting points for test items (Crouch & Mazur, 2001) and could try to develop relevant, novel, and professional problems or examples and non-examples of facts, concepts, principles, or procedures (Anderson, 1972). The test item writer could also try to imagine how different conceptions of these elements can be combined, as an intermediate action in divergent thinking, in a test item to create items that are more demanding, require higher-order thinking, or
have more appealing problems or examples. For example, the test item writer could try to set up an item in such a way that the student needs to apply multiple sources of knowledge and skill or lines of reasoning or procedures to find a correct answer. A test item writer can start by working on one idea at a time or on several ideas concurrently.

Following divergent production, the test item writer will enter a phase of convergent production, the goal of which is to ensure the technical quality of the test items, notably making sure that they are clear and unambiguous to the test taker and remain in keeping with the subject matter. In line with Olsen and Bunderson (2004), a competent test item writer starts critiquing these initial ideas from a variety of perspectives. Test item writers ask themselves whether the question and answers do indeed query the students about their knowledge or understanding of the relevant fact, concept, or principle, and determine whether the level of difficulty is too high or too low, and whether students with lower proficiency would choose a distractor instead of the correct answer.

In addition, the test item writer will evaluate, as a fast alternation of divergent and convergent thinking, whether the chosen wording or presentation of the stimulus makes it more difficult or clearer than alternative wording or presentation. Furthermore, a test item writer will try consciously or, even unconsciously and automatically at the expert level, to apply convergent requirements to ensure clear, concise test items. The test item writer might well employ lists of dos and don’ts as Haladyna, Downing, and Rodriguez (2002) supply for that purpose.

An additional difficult aspect in the test item design process is the need to balance two opposing requirements; the demand to present novel material and information while also presenting material that is conceptually close to the original instructional materials to remain in keeping with the learning goal or topic that is to be addressed and, not least, to prevent presenting distractors (in the case of multiple-choice test items) that are obviously not plausible. Ideas that are too far off topic conceptually are unfit. Teachers have to walk a thin line between producing something novel but also within bounds, however ill-defined they may be.
The results of such cognitive processes can be a revision of the initial idea, addition of information, rewording of information, re-establishing the validity of information, or finding germs of ideas for new test items. This is the process of co-evolution that demands serious divergent and convergent thinking and reassessment by the test item writer of the content of instructional material and estimation of student behavior. Dewey (2007) provides a revealing example of the reassessing process, reporting that he prefers to write test items with the textbook closed and then construct multiple-choice test items. At that point:

*With the book closed, I must rely on my own memory of the material. I figure if I cannot remember something myself, it is not reasonable to ask students to remember it. This means I have to double-check later to make sure my own memory of the material was correct (2007, p. 1).*

In almost all cases, moreover, test item writers find themselves in a situation where time is always an issue and it is simply not possible to collect more information. In particular, teachers remain in the dark regarding the fit of their achieved level of originality, technical quality, difficulty, and discrimination power of their test items with the abilities of the students and how students will perceive and interact with the items. At that point, teachers have to decide what to do. Phenomena known from naturalistic decision making (Lipshitz, Klein, Orasanu, & Salas, 2001; Zsambok & Klein, 1997), in which decision makers need to cope with uncertainty, come into play. In naturalistic decision making, expertise plays a central role in the quality of the decisions made.

In short, test item design is an active process of constantly going back and forth, diverging and converging, judging the solution and evaluating the problem, all in a co-evolutionary manner.

### 2.6 Models for solving ill-defined design problems

In the previous sections, it was argued that the task of designing test items is a matter of ill-defined design problem solving, divergent production, convergent production, and co-evolution. That characterization offers a fruitful perspective for describing the test item design task. The question now becomes how such a process
can be represented in a model in which the design problem outlook is acknowledged. Second, however, it is also important that a model is developed that provides cues for deriving support for teachers to design test items. Third, to present a clear theory, the model should be parsimonious and simple which also, fourthly, can enhance effectiveness in practice for teachers in higher education. Simplicity is deemed important because in other creative problem-solving domains it has been observed that complex models hinder clear communication, appreciation, and utility for more novice designers (Buijs, 2003). In this section, relevant models from literature for solving ill-defined design tasks are presented, resulting in the presentation of creative problem solving (Isaksen & Treffinger, 2004) as a foundation on which to develop the final model.

For the generation of design ideas, especially in engineering design, various representations of divergent and convergent cognitive production have been developed and explored by Pugh (1991), Cross (1994), and Liu, Chakrabarti, and Bligh (2003). Cross and Pugh outline visual models that represent the path that a designer follows in multiple stages of divergence and convergence to reach the chosen solution, as shown in Figures 2 and 3. These models assume that a suitable solution is in principle always available within a universe of possible solutions and that the designer goes through several stages to find and develop this solution. These models show the probable cognitive route that designers follow during design in a visual way and demonstrate the dimensions of time and breadth of scope of the number and originality of both generated and discarded ideas.
The models capture visually the iterative, diverging, and converging nature of design and show that the process must eventually result in a single solution. These models therefore reflect the type of model appropriate for a cognitive process model for test item design, but they do not show how either the problem or the
perspective on the problem changes during development, as shown in Figure 1. Therefore, it is necessary to adopt a more general problem-solving model in which the conception of ideas is central.

Such a general model to capture the iterative diverging and converging aspects of design is the creative problem-solving (CPS) framework developed by Osborn (1953) and examined and refined in several longer-term studies (Brophy, 1998; Isaksen & Treffinger, 2004). The CPS framework presents a structured but flexible prescriptive process approach consisting of several steps for solving ill-defined problems, of which designing solutions is only one step. Typical steps in the CPS framework are arranged in three groups: understanding the problem, with steps to identify the goal or challenge, gather information, and clarify the problem; generating ideas; and planning for action, with steps of first strengthening and then implementing the solution. These steps display significant similarity with the general problem-solving steps that Jonassen (2004) identifies. The CPS steps guide the creative process and lead to the production of one or more creative solutions, yet remain within strict requirements so that the solutions are acceptable to all stakeholders (problem owners, users of the solution, producers of the solution, regulators, etc.). CPS steps can be applied not only to more complex problems that take substantial time to develop, such as constructing a curriculum or an assessment program, or even an entire test, but also to smaller problems. Each step contains recursive and iterative loops that emphasize the supple nature of the process. Research has shown that providing CPS training to individuals or groups resulted in improved attitudes to and execution of divergent thinking and ideation (Puccio, Firestien, Coyle, & Masucci, 2006). In the fourth version of CPS, these steps are presented as successive diamonds that characterize the diverging and converging nature of problem solving (Figure 4). In the figure, following Pugh (1991), the outward-bound arrows indicate divergence that produces more and more original ideas and the inward-bound arrows indicate convergence that selects and optimizes ideas. Within these steps, recursive loops and iterations of diverging and converging can take place. The diamonds in the models reflect how the attention of the designer shifts between diverging and converging.
2.7 **Towards a process model for test item design**

Section 2.5 described designers’ activities and section 2.6 described contiguous models capturing the design process. These sections have made the case for characterizing and describing the task of designing test items in terms of ill-defined design problem solving, divergent production of test item ideas, convergent production of test items, and co-evolution of the perception of both problem and solution. Based on that characterization, general models for solving ill-defined design problems are presented, resulting in the presentation of CPS as a foundation on which to develop the final model.

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
<th>Step 5</th>
<th>Step 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding the Problem</td>
<td>Generate ideas</td>
<td>Planning for Action</td>
<td>Monitor the problem space and solution options.</td>
<td>Implement and monitor the solution</td>
<td></td>
</tr>
<tr>
<td>Understand the problem.</td>
<td>Gather information, clarify the problem</td>
<td>Identify and clarify alternative opinions, positions, and perspectives of stakeholders.</td>
<td>Generate ideas and possible problem solutions, search for solutions.</td>
<td>Assess the viability of alternative solution by constructing arguments and articulating personal beliefs</td>
<td>Strengthen the solution.</td>
</tr>
<tr>
<td>Identify the goal or challenge.</td>
<td>Articulate the problem space and contextual constraints that represent the problem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 4.* The creative problem-solving (CPS) framework (Brophy, 1998; Isaksen & Treffinger, 2004; Osborn, 1953) is shown as six diverging and converging steps with divergent and convergent production within each step. The steps are based on Jonassen (2004; 2005) and Isaksen and Treffinger (2004).
Figure 5, which is constructed on that basis, presents a cognitive process model for a particular test item design task that could be encountered in practice. The process is used to reach a final test item or set of test items. Cylinder height represents the technical quality (the degree to which test items are clear, concise, unambiguous, discriminating, etc.), while cylinder width represents the degree of originality (novel content, novel examples, measure of paraphrasing, authenticity of content, content demanding higher cognitive processing, etc.). White cylinders represent newly-conceived ideas, gray cylinders represent ideas during the development process, and black cylinders represent final ideas. Gray cylinders that are connected with a solid line to a white cylinder represent the same or a further-progressed test item idea when they are connected with an arrow. The same protocol applies to black cylinders if they are connected with a solid arrow to a grey cylinder. There are fewer black cylinders than white cylinders. In the convergent phase, ideas are intended to be stretched both in terms of originality and technical quality. For illustrative purposes, nine ideas are initially conceived in the figure; one new idea leads directly to one additional new idea. All ideas are evaluated during the process. From those intermediate ideas, three ideas are discarded, while two others merge into one new idea. One idea dissolves in an intermediate idea, and three intermediate ideas are improved. Ultimately, four ideas reach the final stage.

Dotted lines with two arrows indicate that each idea, intermediate test item, or nearly-final test item is evaluated in view of the current understanding of the problem. This evaluation can be either spontaneous or a deliberate reflection, and diverging and converging thoughts alternate (Brophy, 1998). Based on the developed test items (not depicted), the problem perception can change and, lead to the test items’ being modified. The dotted lines thus represent the co-evolutionary process.

It is clear that for more test items to be designed, the process model can be repeated a number of times. Depending on expertise and personal preference (Hunt, 1991; Newell & Simon, 1972), the test item writer can choose between three approaches: (1) a breadth-first approach that starts by generating many initial solutions and then refines each one before subsequently choosing a limited number of ultimate
solutions; (2) a depth-first approach that generates only one or a very limited number of initial solutions and refines them fully before entering a new cycle of generating and refining; or (3) different combinations or sequences of the breadth and depth approaches. Expert designers generally follow a more narrow path through this process than novices and spend more time on optimizing ideas and eliminating problems (Cross, 2004; Fricke, 1996). On the basis of their proficiency, experts’ initial ideas are more likely to be good ideas and they will be able to optimize those ideas more effectively (Chi, 2006; Regehr & Norman, 1996). Novices, by contrast, generally use unstructured generate-test cycles to solve problems.

The process model shows an admittedly complex process, but the overall flow of the design process that first moves outward (diverging) and then inward (converging) is clearly recognizable. To make the model even clearer, a simpler representation is provided in Figure 6, showing the successive stages of increasingly clear perceptions of the problem that follows the process of developing ideas and culminates in the final test items.
Figure 5. Final graphical representation of the process model of diverging and converging in designing test items for a specific topic or test. The steps follow from section 2.5. Solid lines represent the progression of ideas into chosen solutions. Dotted lines represent reflection of the conjectured and progressed ideas for the test item in relation to the problem, other test items, and their internal characteristics and features.
Figure 6. Simplified final graphical representation of the process model of diverging and converging in designing test items for a specific topic or test.
We thus have a new understanding and description of the test item design process. As noted in the introduction, this more nuanced appreciation of the process should be of help in designing instruction and support techniques that will improve the outcomes of the test item design process for teachers.

The main point of the process model, which distinguishes it from the classic approach to test item design, is that it emphasizes co-evolution and stimulation of the divergent and convergent production of test items. Stimulating divergent production is an especially distinctive feature. The process model demarcates the design process phases in which teachers can be supported with specific interventions (aids, techniques, etc.) to improve the divergent and convergent production of test items. First, interventions such as inspirational guidelines to spark imagination (see Chapter 3) or design patterns (see Chapter 4) could constitute such support. Second, general techniques and guidelines stemming from the CPS framework, like separating the divergent and convergent phases in time, postponing criticism, appraising positive characteristics of intermediate solutions, combining solutions, and employing a range of general problem-solving techniques (Michalko, 2006) could form such support. Concept maps (Novak, 1998) or associative thinking techniques (Waks, 1997) are especially promising tools (see Chapters 6 and 7, respectively). Finally, specific techniques stemming from item-writing literature can be employed. An overview of existing methods and techniques from item-writing literature is listed in the Appendix of this chapter for the interested reader.

An important feature of the process model developed in this chapter is that it can be used to explain and illustrate the simultaneously divergent and convergent, co-evolutionary character of the test item design process to teachers, so as to strengthen their self-regulatory processes with respect to generating ideas and then optimizing them. This metacognitive ability to alternate deliberately between cognitive strategies, according to the current need during the design process, can improve results in creative tasks (Brophy, 1998) in terms of both the number and quality of test items produced.
2.8 Conclusion and discussion

In this chapter, a number of arguments have been made to support building a process model for test item design. First, it was argued that test item design can well be characterized as solving ill-defined problems based on Reitman’s three characteristics of ill-defined problem solving. Second, it was argues that test item design is as a form of design problem solving. Third, the process of test item design was explained to be concerned primarily with divergent and convergent production and thinking and co-evolution of the problem and solution spaces. In particular, divergent production of test item ideas was argued to be a crucial part of test item design, for without that element test items would never come into existence. Based on this evidence, process models for solving ill-defined design problems were presented and a process model specifically for test items was developed and explicited on that basis. The implications of this model for describing test item design and establishing opportunities to improve the test item design process were enumerated.

The focus on the divergent production of test items in this paper by no means undervalues, let alone dismisses, the current state of research into item-writing guidelines that focus primarily on the convergent production of test items. Furthermore, sophisticated formal procedures to generate test items, especially in large numbers, are unquestionably an urgent need (see for example Gierl & Haladyna, 2012; Karamanis et al., 2006; Mitkov, Ha, & Karamanis, 2006).

However, the current study added an additional and vital perspective by emphasizing the ill-defined and inherent design problem-solving character of test item design to the existing scholarship on the topic. Making teachers more aware of the nuances of the test item design task can only aid in communicating and enhancing the skills needed to design those items. As noted above, the model can be supplemented with new, specific techniques to stimulate divergent thinking and production of ideas, such as those familiar from the classic literature regarding test item design can (see the Appendix). When a teacher has gained sufficient basic expertise and skills in test item design, the routes to applying more sophisticated and formal methods lie open for deeper exploration when resources permit.
Finally, the simpler means-end or generate-test approach always remains open for teachers to employ.

The process model is also suitable for both novice and experienced teachers. The model shows novices that test item design actually has a simple overall structure. Simplicity is of importance for novices because they need to be guided and work along well-trodden paths in order to be scaffolded appropriately (Kirschner, Sweller, & Clark, 2006). These paths should not be too detailed, too abstract, or too circular because that would cause novice designers to find them too complex or difficult to use (Buijs, 2003; Jonassen, 2004). Therefore, the process must be presented initially as a simple linear route that can be further developed, as familiarity grows, with more elaborate problem-solving options, heuristics, and design patterns, before shifting to a more complex and cyclic-concentric and chaotic process representation (Buijs, 2003; Isaksen & Treffinger, 2004; Maher & Poon, 1996). The developed process model allows for such growth and complements.

Second, the model also has something important to offer to experts. Making the cognitive process of designing test items explicit is highly likely to enhance experts’ efficacy and raise their skills to even higher levels. Experts who may know and apply the process unconsciously will most likely follow a more focused path through the design process and expend more effort to overcome any problems encountered than novices might (Chi, Glaser, & Farr, 1988; Cross, 2007; Regehr & Norman, 1996). Experts, however, are also prone to inhibition or design fixation (Chi, 2006; Jansson & Smith, 1991; Luchins, 1942) that causes them to rely too much on their expertise and idiosyncratic knowledge and to avoid new strategies. The process model can reinforce experts’ appreciation of others’ approaches and result in their experimenting with novel methods to overcome expert inhibition.

Furthermore, the model provides a basis for the design of computer environments to support the test item design process, which should improve the application of creative techniques for the divergent production of test items. Additionally, computer environments should support the use of more formal prescriptive design
methods or offer and apply converging guidelines to improve the clarity of test items.

Finally, the developed model is based on theories and findings regarding problem solving and design from a number of domains, notably engineering and product design (Cross, 2007), but also from cognition, creativity, expertise and design as in psychology, sociology, architecture, business, and instructional design. In that sense it is an interdisciplinary model. These domains are of course not identical or always directly and mechanically applicable to the problem of test item design. However, building a cognitive model on this foundation strengthens the validity of and ensures a firm basis for the model and for further research into the problem of test item design and conceiving effective support for that task.
2.9 References


Gerritsen-van Leeuwenkamp, K. (2012). *Het relatieve belang van vijftig kwaliteitskenmerken van toetsing voor studententevredenheid in het hoger beroepsonderwijs [The relative importance of fifty quality indicators for measurement of student satisfaction in higher education]*. Open University, NL. Retrieved from http://hdl.handle.net/1820/4295


http://doi.org/10.1023/A:1022534310151


## 2.10 Appendix

Table 1

*Overview of methods and techniques for test item development found in the literature and possible usefulness for teachers in higher education.*

<table>
<thead>
<tr>
<th>N°</th>
<th>Year of Origin (approx.)</th>
<th>Method/Technique</th>
<th>References</th>
<th>Short description</th>
<th>Remarks</th>
<th>Possible usefulness</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1967</td>
<td>Guttman’s Facet Theory</td>
<td>(Berk, 1978; Foa, 1965; Guttman, 1944; Guttman &amp; Schlesinger, 1967)</td>
<td>The complete instructional space is mapped out and distractors are derived from that space.</td>
<td>Very formal method; laborious to execute which limits its usefulness (Haladyna, 1990). There is a strong focus on primary and secondary education.</td>
<td>No</td>
</tr>
<tr>
<td>02</td>
<td>1970</td>
<td>Sentence mapping approach, the ‘Wh’ technique</td>
<td>(Bormuth, 1970)</td>
<td>The syntactic structure of paragraphs, sentences, and contents of instructional text is dissected minutely and transformed into test items. Each possible information unit for a test is transformed into a question involving one or more of “Who, Which, When, Why, and What” regarding that unit.</td>
<td>Very formal; laborious to execute which limits their usefulness (Haladyna, 1990). There is a strong focus on primary and secondary education.</td>
<td>Possible</td>
</tr>
<tr>
<td>N°</td>
<td>Year of Origin (approx.)</td>
<td>Method/Technique</td>
<td>References</td>
<td>Short description</td>
<td>Remarks</td>
<td>Possible usefulness</td>
</tr>
<tr>
<td>----</td>
<td>--------------------------</td>
<td>------------------</td>
<td>------------</td>
<td>-------------------</td>
<td>---------</td>
<td>---------------------</td>
</tr>
<tr>
<td>03</td>
<td>1972</td>
<td>Transformation of information</td>
<td>(Ebel, 1972)</td>
<td>Distractors are derived from this space.</td>
<td>Ebel (1972, pp. 172–175) presents ‘suggestions’ for the development of true-false test items such as ‘restate the essential idea in different words,’ ‘restate a part of the original idea,’ ‘relate the basic idea to some other idea,’ ‘develop implications of the basic idea,’ and ‘infer the effect of different (even impossible) circumstances.’</td>
<td>Yes</td>
</tr>
<tr>
<td>04</td>
<td>1974</td>
<td>Item Forms Amplified Objectives</td>
<td>(Hively, 1974; Popham, 1975; Schott, Neeb, &amp; Wieberg, 1984)</td>
<td>A single objective is defined by several Item Forms that elicited particular errors. Amplified objectives are objectives that are worked out in detail with prescriptions for resulting test items.</td>
<td>Very formal. Item Forms mainly used for arithmetic problems. Amplified objective is laborious to execute.</td>
<td>Doubtful</td>
</tr>
<tr>
<td>05</td>
<td>1978</td>
<td>Six Steps Approach</td>
<td>(Miller, Williams, &amp;</td>
<td>Miller et al. describe how test items extending beyond recall should develop:</td>
<td>Loosely prescriptive, easy to understand procedure. A clear process for each test item to be</td>
<td>Yes</td>
</tr>
<tr>
<td>N°</td>
<td>Year of Origin (approx.)</td>
<td>Method/Technique</td>
<td>References</td>
<td>Short description</td>
<td>Remarks</td>
<td>Possible usefulness</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------</td>
<td>------------------</td>
<td>------------</td>
<td>-------------------</td>
<td>---------</td>
<td>---------------------</td>
</tr>
<tr>
<td>06</td>
<td>1956 and later</td>
<td>Action Verbs based on the taxonomy of Bloom</td>
<td>(Bloom, 1956; Gagné &amp; Briggs, 1974; Gronlund, 1998)</td>
<td>Using general taxonomic notions to develop test items and specific verbs in the test item are supposed to elicit a certain cognitive behaviour in the examinee. Often advocated or use in test item writing literature.</td>
<td>Research has shown that cognitive taxonomies have low validity (e.g. Seddon, 1978) and are primarily prescriptive. Not empirically studied if and how taxonomies results in better test items or aid developers in their development as a process.</td>
<td>Possible</td>
</tr>
<tr>
<td>07</td>
<td>1965 and further</td>
<td>A - Taxonomic approaches</td>
<td>(Biggs &amp; Collis, 1982; Ebel, 1965; Ebel &amp; Frisbie, 1991; Quellmalz &amp;</td>
<td>Suggestions are made concerning test item setup to design test items by adhering to a certain cognitive process taxonomy or task for information handling. In this literature these can be related in</td>
<td>Ebel and others provide ample examples. No empirical studies that provide evidence that they are effective to use by teachers. Some approaches are rather formal.</td>
<td>Possible</td>
</tr>
<tr>
<td>N°</td>
<td>Year of Origin (approx.)</td>
<td>Method/Technique</td>
<td>References</td>
<td>Short description</td>
<td>Remarks</td>
<td>Possible usefulness</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>------------------</td>
<td>------------</td>
<td>-------------------</td>
<td>---------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kozma, 2003; Williams &amp; Haladyna, 1982</td>
<td>general terms to facts, concepts, and principles.</td>
<td>Research has shown that cognitive taxonomies have low validity (e.g. Seddon, 1978) and are primarily prescriptive.</td>
<td>Possible</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Wilbrink, 1983)</td>
<td>Wilbrink presents methods to design test items, based on an inventory of instructional materials, for separate facts, concepts, and principles, for relations among facts, concepts, and principles and for texts.</td>
<td>Wilbrink has unfortunately published very little in English language scientific literature.</td>
<td>Possible</td>
</tr>
<tr>
<td>08</td>
<td>1951</td>
<td>Instruction Practice Feedback Reinforcement</td>
<td>(Downing &amp; Haladyna, 2006; Ebel, 1951)</td>
<td>Learning to develop test items is difficult and takes an intensive learning cycle to become proficient. Feedback by experts and prospective examinees is of major importance.</td>
<td>Without doubt, gaining expertise from deliberate practice (Ericsson, 2006) is of importance to become a skilled item writer. In higher education, lack of time other and resources limits this approach.</td>
<td>Possible</td>
</tr>
<tr>
<td>09</td>
<td>1983 and further</td>
<td>Item parts, Replacement Procedure</td>
<td>(Gierl et al., 2008; Glas &amp; Van der Linden, 2003; Millman &amp;</td>
<td>Test items have a static structure. Variables (table data or derived data) fill the structure to generate variants of test items.</td>
<td>The computer plays an important role in generating variants of test items (Gierl &amp; Haladyna, 2012; Millman &amp; Westman, 1989).</td>
<td>Possible</td>
</tr>
<tr>
<td>N°</td>
<td>Year of Origin (approx.)</td>
<td>Method/Technique</td>
<td>References</td>
<td>Short description</td>
<td>Remarks</td>
<td>Possible usefulness</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------</td>
<td>------------------</td>
<td>------------</td>
<td>-------------------</td>
<td>---------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Item cloning, Controlled Item Generation techniques</td>
<td>Westman, 1989; Wilbrink, 1983b</td>
<td>Approaches for test item development aimed at controlling difficulty levels and validity</td>
<td>The computer can only be applied when structures for test items have been generated in advance. Techniques for professional test development bodies, especially primary and secondary education, reading, writing, mathematics, intelligence, biology. Attempts have been made in the field of medicine in higher education (Gierl &amp; Lai, 2013).</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1989</td>
<td>Item shells</td>
<td>(Carriveau, 2010; Gronlund, 1998; Haladyna, 2004; Haladyna &amp; Shindoll, 1989)</td>
<td>Designing test items starts with selecting a 'proven' effective syntactic questioning structure which is filled with substance. Features of replacement procedure are incorporated. Features of taxonomic categorization are incorporated by categorizing groups of syntactic structures.</td>
<td>Not experimentally tested for its effectiveness, only narrative evidence reported that teachers were able to produce test items based on the shells within the domain of health sciences (Haladyna &amp; Shindoll, 1989).</td>
<td>Yes</td>
</tr>
<tr>
<td>N°</td>
<td>Year of Origin (approx.)</td>
<td>Method/Technique</td>
<td>References</td>
<td>Short description</td>
<td>Remarks</td>
<td>Possible usefulness</td>
</tr>
<tr>
<td>----</td>
<td>--------------------------</td>
<td>------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>11</td>
<td>1986</td>
<td>Item modelling</td>
<td>(Haladyna, 1991; LaDuca, Staples, Templeton, &amp; Holzman, 1986)</td>
<td>Using facets consisting of different contexts, tasks, and other occurrences, new variants of items can be constructed. In addition, scenario-based questions can be created with this technique.</td>
<td>Quite elaborate methods. Mainly useful in the domain of health sciences and mathematics with well-delimited domain specification.</td>
<td>Possible</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B – Key-feature questions</td>
<td>(Bordage &amp; Page, 1987)</td>
<td>Test items present a case or problem for which the next step is queried in order to solve the case.</td>
<td>Mainly used in the domains of health sciences.</td>
<td>Possible, for health sciences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C – Extended</td>
<td>(Case &amp; Swanson, 1994)</td>
<td>A number of test items make use of the same set of answering options that</td>
<td>Mainly used in the domains of health sciences.</td>
<td>Possible, for</td>
</tr>
<tr>
<td>N°</td>
<td>Year of Origin (approx.)</td>
<td>Method/Technique</td>
<td>References</td>
<td>Short description</td>
<td>Remarks</td>
<td>Possible usefulness</td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>15</td>
<td>2000</td>
<td>Matching questions</td>
<td>(Charlin, Roy, Brailovsky, Goulet, &amp; Van der Vleuten, 2000)</td>
<td>are correct-incorrect choices depending on the test item.</td>
<td>Mainly used in the domains of health sciences.</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>2000</td>
<td>D - Script Concordance Test</td>
<td></td>
<td>Examinees are presented authentic clinical situations in which they have to interpret data to make decisions.</td>
<td>Mainly used in the domains of health sciences.</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E - Comprehensive Integrative Puzzle</td>
<td>(Ber, 2003)</td>
<td>Its answer sheet is a grid comprising rows and columns. The left-hand column contains diagnoses or brief clinical vignettes. To complete the cells of the grid the student is required to match, stepwise, the various 'disciplinary investigations' to the diagnoses or clinical vignettes.</td>
<td>Mainly used in the domains of health sciences.</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>2000</td>
<td>Various digital innovative</td>
<td>(Boyle, 2005; Draaijer &amp; Hartog, 2007;</td>
<td>A growing collection of sample digital test items is developed and described. For specific domains,</td>
<td>Possible usefulness</td>
<td></td>
</tr>
<tr>
<td>N°</td>
<td>Year of Origin (approx.)</td>
<td>Method/Technique</td>
<td>References</td>
<td>Short description</td>
<td>Remarks</td>
<td>Possible usefulness</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------</td>
<td>------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>item formats and taxonomies</td>
<td>Freie Universität Berlin, 2003; Parshall, Davey, &amp; Pashley, 2000; Parshall &amp; Harmes, 2007, 2008; Scalise &amp; Gifford, 2006; R. C. Thomas et al., 2004; R. C. Thomas &amp; Milligan, 2003; WebAssign, 2003)</td>
<td>specific technical item types are becoming available, specifically for language testing, mathematics, chemistry.</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>14</td>
<td>2004</td>
<td>Computer generated</td>
<td>(Karamanis et al., 2006; Computers can be fed instructional materials and other resources and are generated, but manual checking is</td>
<td>Appropriate test items can be generated, but manual checking is</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>N°</td>
<td>Year of Origin (approx.)</td>
<td>Method/Technique</td>
<td>References</td>
<td>Short description</td>
<td>Remarks</td>
<td>Possible usefulness</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------</td>
<td>------------------</td>
<td>------------</td>
<td>-------------------</td>
<td>---------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>multiple choice items based on natural language processing techniques.</td>
<td>Mitkov et al., 2006</td>
<td>able to process it to generate multiple choice questions.</td>
<td>needed and it is difficult to generate higher order learning items.</td>
<td></td>
</tr>
</tbody>
</table>