Chapter 5: A Proposal for a Tool to Support Teachers in Divergent and Convergent Production of Test Items

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
5 A Proposal for a Tool to Support Teachers in Divergent and Convergent Production of Test Items

5.1 Abstract

In this paper, we present a proposal for a computer tool to support teachers in the task of designing selected response test items for achievement tests. The tool is based on a perspective on ill-defined problem solving in which both divergent and convergent production of ideas are central. Based on that perspective, multiple forms of support are blended into one computer tool resulting in a specific layout, specific interactions, and specific functions. We argue that the tool is fit for purpose with arguments derived from findings from educational measurement research, creativity research, research into problem solving, and research into computer support tools for problem solving. The tool we present balances the conflicting requirements of preventing underuse of information for support purposes and avoiding an obtrusive level of information that hinders the use of the tool.

Keywords: creative problem solving, test item writing, selected response test items, educational measurement, creativity support tools
5.2 Introduction

In this paper, we present the design of a computer tool to support teachers in the task of designing selected response test items for achievement tests. The goal of this tool is to improve the process by which those items are designed. The tool is based on a perspective for test item design in which ill-defined creative problem solving, divergent production to create novel content and ideas for test items, and convergent production to produce test items that are clear and concise are all central to the mechanism of the support offered.

We begin by describing the background and conceptual framework for the computer tool, followed by an outline of the tool’s design. Third, we provide the rationale for the design, including the arguments for the design’s appropriateness to support teachers optimally. The paper concludes with a discussion of the tool and its implications for theory and practice.

5.3 Background and conceptual framework

Many teachers in higher education have to develop selected response test items such as multiple-choice or true-false test items (Anderson, 1987; Parkes et al., 2003) for end-of-course achievement tests or, increasingly, for formative assessment purposes such as online quizzes (Gibbs, 2010) or activating learning materials (Aegerter-Wilmsen, Coppens, Janssen, Hartog, & Bisseling, 2005).

It is important for higher education that teachers be able to generate a range of test items that are appropriate for the increasing variety of goals for which they are needed; teachers should not only be able to design test items for rote learning purposes, but also ideally can query higher-order learning goals such as application of knowledge, critical thinking, and problem solving (Downing & Haladyna, 2006) with a variety of selected response test item formats such as ordering, matching, drag-and-drop, text match, or numeric. However, designing test items can be a difficult task that takes time to accomplish (Case, Holtzman, & Ripkey, 2001; Downing & Haladyna, 2006; Mayenga, 2009); it has been reported that in practice teachers design test items that tend to consist of decontextualized, verbatim factual recognition or recall test items (Buckles & Siegfried, 2006) or that contain flaws
(Hansen & Dexter, 1997; Jozefowicz et al., 2002). Improving teachers’ performance in the higher education design task can be expected to have a positive impact on the quality of teaching, learning, and assessment.

In this section, a cognitive process model for test item design that was developed (see Chapter 2) is used as the basis for designing a tool to support teachers in that task. In Chapter 2 the case is made that designing test items is effectively solving ill-defined problems for which new and creative solutions are necessary, and that experienced designers are often solution-focused and try to develop new ideas in an iterative manner and working with conjectures and prototypes (Tripp & Bichelmeyer, 1990; Visscher-Voerman & Gustafson, 2004).

Figure 1 shows the cognitive process model for the design of a limited set of test items that Draaijer et al. developed, making clear that the main phases in the process of designing test items revolve around divergent and convergent thinking (Brophy, 1998). Divergent thinking, which consists of conceiving original ideas, is a necessary step. The figure makes clear that initially generated ideas go through a selection phase of only the most promising examples, after which they are developed further, eliminating flaws in test item construction, improving clarity, or adding new elements to increase originality or refine technical quality. The perspective on the problem and the solution co-evolve in an iterative manner (Maher & Poon, 1996). The rationale for the diamond-shaped figure arises out of models for creative problem solving by Osborn (1953) and Isaksen and Treffinger (2004). Figure 1 shows an example of the number and quality of developed test items that might be encountered in practice. It makes clear that the perceptions of both problem and solution change during the process. The problem becomes increasingly clear, representing the fact that the perception of the problem that the teacher is trying to solve evolves with the process of thinking up ideas and refining and strengthening them.
A Proposal for a Tool to Support Teachers in Divergent and Convergent Production of Test Items

Figure 1. The process model of diverging and converging in designing test items for a specific topic or test. Dotted lines represent reflection of the conjectured ideas for test items in relation to the problem (co-evolution), other test items, and their internal characteristics and features. The wider the cylinders, the better originality of the idea; the higher the cylinder, the more clear and precise the idea is. The perception and clarity of the problem develop along with the development of the test items.

Based on Figure 1, a prescriptive model can be advanced that guides the teacher in conducting the steps of divergence and convergence in the design process in order to improve its outcome (Isaksen & Treffinger, 2004). Figure 2 presents this prescriptive model. In the diverging and converging phases, the designer is advised to use specific supporting information and techniques. The central insight of Figure 2 is that teachers are guided to diverge and converge intentionally and use support that is appropriate for whichever phase is at hand.
A cognitive process model as presented above does not result deductively in the design of practical effective computer tools to support teachers. For that goal, a design approach must be chosen in which the characteristics of the model are translated into functions, content, and interface elements that accord with this model. For that reason, the following research question is put forward in this chapter.

What are the main requirements for a computer support tool to support divergent and convergent production of selected response test items, and how can these requirements be met by a computer application?

To answer this question, we begin by describing the design process that leads to the proposal for the tool. Next, we describe the main features of the test item design tool as developed, after which we present the rationale underlying the tool’s main
features. The rationale provides the arguments regarding the requirements for the tool and the expected adequacy of the proposal.

5.4 Design process

There is no single, best way to design software or a software tool. Multiple perspectives exist regarding software design principles and design methods (Davis, 1995; Freeman & Hart, 2004; Ralph, 2010). To be practical for the design of the computer tool in this paper, a four-step approach was adopted as a form of “prototyping study” (Becker & Berkemeyer, 2002; Buxton, 2010; Ozenc, Kim, Zimmerman, Oney, & Myers, 2010; Tripp & Bichelmeyer, 1990). In a prototyping study one or more conceptual designs are prepared; examples include guiding principles, conceptual descriptions, drawings, mockups, global dimensions, wireframes, prototypes, etc. These conceptual designs are then studied to define features of a product which are used to elicit impressions and appraisals regarding the expected usefulness of that product and offer an inspection of the adequacy of the rationale underlying the design.² After an overview of the four steps below, the prototype and rationale for the adequacy of the tool are presented in sections 5.6 and 5.7.

In the first step, Hewett’s (2005) concept of a virtual workbench was chosen a guiding principle for the design of the support tool. This concept is a powerful but simple metaphor in which a number of recommendations and requirements for the design of computer tools for creative problem solving can be expressed. As a second step, as tools are central requirements for the virtual workbench of test item design, functions and resources that might serve as tools were inventoried. The main point at this stage was to develop an overview of support possibilities that might eventually be incorporated into the virtual workbench.

² The term “design study” in this chapter therefore does not refer to “design-based research” or “educational design research” (Barab & Squire, 2004; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006), as it often does in educational research.
Third, an important common requirement for a test item support tool stemming from the literature on support for test item design on the one hand and creative design problem solving on the other was the presence of an *archive* with developed ideas to be used for reference and as a working stock for the production of ideas. Based on that common requirement it was decided that as a starting point, a classic item banking structure would be a core function in the design and interface of the system.

In step four, a synthesis was made of the concept of the archive, the virtual workbench, and the functions and information resources inventoried, resulting in a single coherent interface for the user. In particular, conflicting user interaction requirements for a support tool were addressed. During step four, provisional prototypes of the system were made, tested, and improved with the aid of teachers working in higher education. The main goal at this stage was to ensure that the core functions were technically sound and that navigation and accessing support information was easy to understand. Finally, the system was tested in an experimental study that saw a number of students working with the system to design test items. At that stage, preliminary findings regarding the appropriateness of the tool were elicited.

### 5.5 Description of the test item design support tool

The tool’s first key characteristic is that it is built around an item banking functionality that can be accessed on the Internet, as shown in Figure 3. The interface has three main panels A, B, and C. Panel C contains the general item banking functionality for entering, storing, editing, retrieving, and listing test items that consist of stimuli, answers, distractors, metadata, comments, etc. (see Figure 4). The item banking system is subsequently augmented with support for the teacher.
The second key characteristic of the tool is the three phases of test item development presented prominently in panel (A): Orientation, Diverge, and Converge. Orientation refers to representing and understanding the problem, diverge to the diverging phase, and converge to the converging phase. In addition to indicating the phase, the selected test item format determines what support information is presented to the user in Panel (B). Panel (A) enables the user to call up this support information for test item design at any time; the panel is in essence a list of links to information or functions; the links are also grouped into the categories orientation, diverge, and converge. Finally, Panel (C) contains the item banking functionality for listing and entering test items and their elements, such as stimuli, answering options, metadata, etc.
The user can set the design phase in order to focus on the problem-understanding phase, the diverging phase, or the converging phase of test item design, as seen in ① in Figure 4. The information in Panel ⑧ and the tools in Panel ⑨ reflect the phase chosen; depending on the phase and the type of test item—true-false, multiple-choice, matching, etc.—support information relating to that phase is presented in panel ⑧ alongside the test item entry form panel ⑨. For example, when the phase is set to orientation, it shows general information regarding multiple-choice items. Conversely, when the phase is set to optimization, the panel displays guidelines to ensure that test items are formulated unambiguously.

When the diverging phase is selected, the test item entry form also displays additional options to scroll through and select Start Sentences for test items (see ② in Figure 4). Start Sentences is based on the concept of item shells as design patterns and can inspire the teacher to design test items with certain positively regarded characteristics (Haladyna & Shindoll, 1989); they were also reported as valuable inspirational guidelines by Draaijer et al. (2007, see Chapter 3). When
selecting Start Sentences, the appropriate sentence is automatically put into the stimulus box of the test item to begin test item design.

When the converging phase is selected, the test entry form can show a warning text in a pink box (see 3 in Figure 4) when a common construction error is detected, like omitting a question mark in the stimulus of the test item, using a negation in the test item, or using the phrase “all of the above” as a final distractor for a multiple-choice test item.

The support information in panel 8 consists of written information, self-paced presentations, example test items, worked examples of improved test items, links to Internet videos about question design, etc. Additionally, a link to an online freeform note-taking environment, including options to draw and diagram, can be activated for each test item or for the system as a whole.

Finally, all three panels are accessible at all times and cannot be closed. Each panel takes up one third of the screen width by default, though users are able to adjust the width of the panels to their individual needs.

5.6 Rationale for features of the support tool

Multiple perspectives and guiding principles exist for the design of creativity support tools (Greene, 2002; Hewett, 2005; Hewett & DePaul, 2000; Shneiderman, 2002; Shneiderman et al., 2006). However, it was decided to use the concept of Hewett’s virtual workbench as the guiding principle of the test item design support tool (2005). Hewett concludes on the basis of current research that the metaphor of a virtual workbench is one of the most powerful means to think about, describe, and design creative problem-solving environments. Using a virtual workbench to solve problems also reflects Davis’s (1995) guiding software design principle that the design structure should resemble the structure of the problem and how to solve it.

Hewlett’s requirements for a virtual workbench are that it be an open and adaptable environment that can be used for a number of purposes by offering a set of both general and specific tools to be used in nearly any sequence or
combination. A virtual workbench for test item design needs to be equipped with tools that enable crafting of instructional materials and learning objectives into test items at the disposal of users at any time. Table 1 compares a physical workbench and a virtual online workbench for test item design.

Table 1

*Comparison of a physical and virtual workbench for test item design.*

<table>
<thead>
<tr>
<th>Workbench</th>
<th>Element</th>
<th>Virtual Workbench</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tools and functions</td>
<td>Bench</td>
<td>Database, notebook</td>
</tr>
<tr>
<td></td>
<td>Clamp</td>
<td>Notebook, editing fields, drawing programs</td>
</tr>
<tr>
<td></td>
<td>Chisel, hammer, files, mill</td>
<td>Editing bars</td>
</tr>
<tr>
<td></td>
<td>Nails, glue, tape</td>
<td>Sets of editing fields constituting a test item, links to original data sources</td>
</tr>
<tr>
<td>Electric machines</td>
<td></td>
<td>Automated functions for editing, finding, retrieving, replacing</td>
</tr>
<tr>
<td>Materials</td>
<td>Wood, iron, plastic</td>
<td>Instructional materials, materials from external resources</td>
</tr>
<tr>
<td>Information and</td>
<td>Manuals, videos on optimal use</td>
<td>Online support information for optimal use of workbench materials and tools</td>
</tr>
<tr>
<td>knowledge</td>
<td>of materials and tools</td>
<td></td>
</tr>
<tr>
<td>Coaching, training,</td>
<td>Coaching and training for</td>
<td>An online context-sensitive system presenting just-in-time prompted support information relating to the task that the user executes.</td>
</tr>
<tr>
<td>guidance</td>
<td>working with tools and materials</td>
<td></td>
</tr>
</tbody>
</table>

Ideally, the requirements for the virtual workbench should coincide with the requirements for test item writing found in that literature, because coinciding requirements offer a common foundation for designing a support tool. An important common requirement in literature regarding creativity support tools and proposals for computer tools for test item design from the item writing domain (Glas, 1997; Millman & Westman, 1989; Rikers, 1988) is that users should have
access to an archive with intermediate and finished ideas that can easily be used for reference, comparison, and as working stock. This is the main reason to choose an item banking functionality as a core element in the designed support tool.

An important difference between proposals for test item design tools (Glas, 1997; Millman & Westman, 1989; Rikers, 1988) and a virtual workbench as proposed in the current study is the requirement that, in a virtual workbench, users are not forced into a linear, strictly structured workflow. Instead, Hewett emphasizes that users must be in control of when to work on which problem in whatever sequence they prefer. Given the understanding of test item design as solving an ill-defined creative design problem-solving task, the tool’s design should emphasize a flexible approach to generating test items in which the user is free to follow any path desired.

Further, in line with Hewett and Fischer and Giaccardi (2006), the test item design support tool is – within reason – adaptable by the user. In particular, the information elements in Panel A allow users to define their own sections and add content to them, such as links to online resources regarding test item design or links to relevant archives of comparable test items for reference and inspiration.

Table 1 makes clear that it is not only tools and materials but also manuals and guidance that are core parts of the workbench; they help the user work with the tools and materials more effectively. Manuals and guidance serve as support for acquiring knowledge and skill, metacognition, and self-regulation, all of which are crucial for enhancing performance in creative ill-defined problem solving (Azevedo, 2005; Azevedo, Moos, Greene, Winters, & Cromley, 2008). Table 2 (see Appendix to this chapter) lists a number of supporting functions (tools), materials, and supporting information for test item design, organized according to the steps involved in the problem-solving process.

5.7 Solving the main design challenge for user interaction

Given the concept of a virtual workbench and the features demanded, the problem arises of how to synthesize these features into a test item design support tool that is usable and effective in terms of human-computer interaction and user interface
design (Dix, Finlay, & Abowd, 2004; Jameson, 2009). Obviously, such a system must be user-friendly and easy to navigate in order not to disorient or overload the user (Kirschner, 2002; Kirsh, 2000). However, ensuring that support information in such a system is deployed effectively poses a major design challenge (Fischer, 2001; Jameson, 2009). The challenge stems from the conflicting requirements of preventing underuse of support information and of preventing obtrusiveness of support information. The design of the user interface needs to navigate carefully between these two challenges.

**Requirement 1: Prevent information underuse**

The support tool’s core purpose is to provide support information and functions to the user that are available at all times and sensitive to whatever design phase is active. Users can decide for themselves when they access this resource. This approach is in line with recommendations by Hewett (2005) and Dix et al. (2004), who report that support information should be available and flexible. In the conception of the tool, availability is ensured by preventing information panels \(^{A}\) and \(^{B}\) from being closed. Meanwhile, flexibility is ensured by enabling users to add their own support resources, for example. Regardless of source, calling up or processing this information is strictly *user-initiated*; until a user calls up information, it resides unobtrusively in the system.

However, research indicates that having information or functions merely reside silently in a system and relying on users to call on them does not lead to an effective system (Fischer, 2001; Jameson, 2009; Narciss, Prosko, & Koerndle, 2007); residence in a system is a necessary but hardly sufficient condition for effectiveness. First, Carroll & Rosson (1987), Furman & Spyridakis (1992), Foster (2006), and Narciss et al. (2007) have shown that there is a tradeoff between learning and working known as the “production paradox,” which describes the way in which legitimate short-term concerns of individuals to produce deliverables by a specific date inhibit those users’ acquiring new knowledge and skills that could yield more effective methods for future production occasions. Second, novice users who are not aware of the existence of additional or even required knowledge and skill will
not look for it because they fail to recognize their own deficiencies (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Kruger & Dunning, 1999).

Therefore, in conceiving the support tool, user-initiated support was combined with system-initiated support, which involves the computer actively suggesting or presenting tips or tentative solutions to the user as an external meta-cognitive learning aid (Azevedo et al., 2008). By employing relatively simple means and algorithms, it is possible to realize effective system-initiated support interventions. A hypothetical example of system-initiated support follows: a teacher starts to edit a test item, and the computer presents a popup window that suggests that the teacher switch to the optimize phase or spend some more time in the generate phase to improve the test item. If greater depth were desirable, the popup window could suggest that the user study the information in panel or present that information in a new window. Additionally, the tool has three other, more content-specific support features to which the popup windows could refer.

In the first content-specific form, common test item construction flaws as described by Hansen and Dexter (1997), Jozefowicz et al. (2002), and Tarrant et al. (2006) are employed for system-initiated user support. The most frequently occurring test item construction flaws that these scholars have identified can be detected by simple text string recognition. When such strings are detected in a test item, the advice statement is flagged and presented directly to the user in the form of a pink box appearing in the test item entry form (see in Figure 4). Column 1 in Table 3 (see Appendix to this chapter) presents solutions to these common item construction flaws in the form of short advice statements, while column 2 describes the algorithm that detects the strings. For example, it is preferable that test items do not contain negations or negative prefixes. Identifying this construction flaw is identified instantly by detecting text strings such as “not,” “un,” “dis,” or “im.” Though this detection may appear straightforward or even simple, care is taken regarding the way in which recommendations are communicated. Users are made aware that the system’s advice is not necessarily relevant (Jameson, 2009), as negations sometimes cannot be avoided for a certain test item or flagging may simply be inappropriate because small text strings can be part of regular words.
With regard to the second form, the computer support tool contains *processing* and *questioning* prompts, which are worked examples of test item design problems that users must study and elaborate on to learn about the example, general problems, and recommendations that these problems illustrate. Example 1 shows a worked example.

**Excess material in the stimulus**

Excess material in the stimulus that is not essential to answering the problem increases the reading burden and adds to student confusion over what is being asked.

<table>
<thead>
<tr>
<th>Poor Example</th>
<th>Better Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suppose you are a mathematics professor who wants to determine whether your</td>
<td>When analyzing your students’ pre-test and post-test scores to determine if</td>
</tr>
<tr>
<td>teaching of the unit on probability has had a significant effect on your</td>
<td>your teaching has had a significant effect, an appropriate statistic to use is</td>
</tr>
<tr>
<td>students. You decide to analyze their scores from a test they took before</td>
<td>the t-test for:</td>
</tr>
<tr>
<td>the unit and their scores from another exam taken after it. Which of the</td>
<td>*a. Dependent samples.</td>
</tr>
<tr>
<td>following t-tests is appropriate to use in this situation?</td>
<td>b. Heterogeneous samples.</td>
</tr>
<tr>
<td></td>
<td>c. Homogeneous samples.</td>
</tr>
<tr>
<td></td>
<td>d. Independent samples.</td>
</tr>
</tbody>
</table>

*a. Dependent samples.*

b. Heterogeneous samples.

c. Homogeneous samples.

d. Independent samples.

What is wrong with or can be improved in this test item?

**Comment**

The stem of the poor example above is excessively long for the problem it presents. The stem of the better example has been reworded to exclude most of the irrelevant material, and is less than half as long.

**Research**

Several studies have indicated that including irrelevant material in the item stem increases difficulty artificially and needlessly, and decreases both the reliability and the validity of the resulting test scores.

*Example 1. Worked example of a questioning prompt adapted from Burton (1991) as embedded in the test item design support tool. This questioning prompt presents a comparison of a poorly designed test item and an improved version of that test item.*
Gerjets, Scheiter, and Schuh (2007), Mao and Leung (2003), and Papadopoulos et al. (2009) all identify beneficial results for the problem-solving capabilities of respondents, primarily because of prolonged and intensified processing of the worked examples. Providing processing and questioning prompts was especially beneficial for novice learners. In the concept of the test item design support tool, such system-initiated prompts take the form of onscreen reminders, presented at several intervals, that encourage the learner to process the given worked-out examples.

Though providing users with question prompts can be beneficial, research has shown that because of the effort involved in this type of self-assessment, its actual employment by users can be low (Jameson, 2009; Tsandilas & Schraefel, 2004). In line with these authors’ recommendations, responses to the prompts are optional and the purpose of the prompts is made very clear.

The third type of support prompt embedded in the tool consists of presenting users item shells, as proposed by Haladyna and Shindoll (1989). Item shells are a form of design patterns, hollow syntactic structures of test items with a relatively high probability of resulting in good test item characteristics. These empty structures must be filled with information – facts, concepts, procedures, principles, examples, non-examples, problems – on the topic for which the test item is intended. An example of an item shell is the sentence “Given…, what is the most important reason that …” to start a query for problem-solving skills. Item shells support the divergent production of test items because they provide teachers with new ways to conceive ideas for test items. In the tool, these item shells are named Start Sentences to reinforce that they can be used to start the diverging test item design process. Table 4 in the Appendix provides more item shell examples.

In conclusion, the concern that information residing in the test item design tool is underused during task execution is allayed in the test item design tool by ensuring that the panels of the system are sufficiently large, cannot be closed by the user, and that system-initiated support information is provided to the user. The information is context-dependent with respect to two relatively simple conditions: the phase of design (orientation, divergence, or convergence) and the test item format.
**Requirement 2: Prevent obtrusiveness**

In order to prevent the underuse of information, it is tempting to trigger system-initiated prompts often and to dedicate a large proportion of the interface to that information. However, this approach is not appropriate because there is also a requirement that support systems prevent obtrusiveness (Dix et al., 2004; Jameson, 2009), which is the second core design problem that we address. The requirement to prevent obtrusiveness implies that a good support system should not interfere with the actual task that has to be performed by causing annoyance or hindering that task’s accomplishment. The task must always receive primary attention in such a system; the help function must never overwhelm it.

Support information should in that respect be displayed as minimally as possible or not at all; however, a move too far in that direction could in turn hamper the intended support function, because the information will not be visible. To balance these two essential but conflicting demands, the tool always shows context-sensitive support information on the side of the task environment in Panel \( \text{Panel B} \) (see Figure 3). Providing support in this way has proven effective in practice (Mao & Leung, 2003).

To prevent obtrusiveness and thus annoyance, system-initiated prompts are administered in such a way that they can be adapted in frequency or type. For example, when the user has made a test item construction flaw repeatedly and had it detected by text string recognition, the teacher should be able to disable that prompt. This approach is comparable to the features of the “Tip of the Day” (Horvitz, Jacobs, & Hovel, 1999) prompt that many computer programs offer when starting an application and which can be turned off. With regular frequency however, the “Tip of the Day” should be reactivated if a user has switched it off.

**5.8 Discussion and conclusion**

In this chapter, we have presented a solution to the problem that teachers often work in isolation when designing test items and do not consult experts, item-writing guides, or handbooks. This solution consisted of providing teachers with an online support tool that should provide support for test item design task so as to
improve their execution of it. The more specific question addressed in the chapter was identifying the key features of a computer tool that supports divergent and convergent production of selected response test items and how these characteristics could be integrated into one program that was likely to lead to improved outcomes of the test item design process and increased skill among system users. We have described the core features of this tool and justified its appropriateness with a number of arguments, based in particular on the guiding principle of building a virtual workbench with tools to support test item generation.

Further, we assert that the conflicting requirements for preventing underuse of information and preventing obtrusiveness were resolved in the design of the interface and the tool’s functions. The tool is novel in its approach to supporting the test item design task and draws from a wide variety of research and sources in its design. The design of the tool explicitly builds on the problem-solving framework and cognitive model developed described in Chapter 2 and incorporates guidelines stemming from human-computer interaction and computer tools for creative problem solving.

Evidence for the usability of the tool was derived by preliminary testing and improving prototypes of the tool with higher education teachers to ensure that the main technical functions and navigation of the tool were clear. Evidence remains to be gathered, whether experimentally or in practice, regarding whether the tool’s support functions are usable and effective in terms of improving the design process. In the next chapter, the tool is employed in an experimental setting to design test items and thus provide some preliminary evidence for its functionality and the extent of its effectiveness.

Previous concepts of test item design support tools were developed by Rikers (1988), Millman and Westman (1989), and Cheung (2006). A central limitation of their work was the failure to provide a foundation for the problem-solving characteristics and cognitive process involved in designing test items. Rikers’s work was further limited because his design for a computer support tool was restricted to the description of schemas and workflows; he focused on a wizard-like approach to test item design. Millman and Westman also simply described a tool rather than
proposing an example. Their conceptual tool adopted a wizard-like approach as well, though their descriptions of the user interface and interactions were more elaborate than Rikers’s. Finally, Cheung developed an actual tool to support test item designers, but limited his project to supporting only chemistry teachers; furthermore, Cheung did not use an underlying framework regarding the cognitive process involved in test item design. The implicit view of all these researchers’ attempts to develop a test item design support tool was that constructing test items is considered primarily solving a psychometric measurement problem and a rational process. Given that viewpoint, their emphases were on a linear process of test item design, focusing mainly on the convergent aspect of that process, but as made clear above, the aspect of diverging demands at least as much attention for successful test item design. That crucial element of the test item design task is explicitly addressed in the design of the support tool. The tool as presented in this chapter thus provides an important, new direction in the further development of support tools for test item design.
5.9 References


Buxton, B. (2010). *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann.


http://doi.org/10.1021/ed083p1399


A Proposal for a Tool to Support Teachers in Divergent and Convergent Production of Test Items

http://ascilite.org/archived-journals/e-jist/docs/vol10_no1/papers/full_papers/draaijer_hartog_hofstee.htm


5.10 Appendix

In Table 2 of this Appendix, findings from a literature search regarding specific functions and features that can support metacognition and regulation in creative problem solving and that are appropriate to support the test item design tasks are listed. These functions are grouped according to problem-solving steps as proposed by Jonassen (1997) and Isaksen and Treffinger (2004) in which solution generation is also incorporated. Column 1 presents these steps for problem solving. For each step, the relevant guidelines to support the creative process from literature regarding creative problem-solving environments are presented in column 2. Column 3 shows possible information sources relevant to test and test item design related to those guidelines. Column 4 presents features of the concept tool to present or prompt the information to enhance the rate of use or processing of information.

An example of such a feature follows. In the problem orientation phase (column 1, the problem-solving step) it can be effective to provide stories (Jonassen & Hernandez-Serrano, 2002) about how peers deal with the test item design task. Such stories enable the user of the test item design tool to relate to and identify more with the problem and find inspiration or a clearer mental image of the task or problem (column 2, guidelines). An example of such a story is provided by Russell A. Dewey (Dewey, 2007) who described on his personal website how he deals with his test item design task (column 3, relevant sources). The teacher using the test item design support tool has access to this story via Panel [A] and the stories can be enhanced with video (column 4, how to offer the information stemming from the guidelines and relevant sources to the user).
<table>
<thead>
<tr>
<th>Steps to solve problems</th>
<th>General guidelines</th>
<th>Possible information sources available through panel (A) and (B)</th>
<th>Possible functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: Understanding the problem: Identifying the goal or challenge, gather information, clarify the problem learners articulated problem space and contextual constraints, learners represent the problem.</td>
<td>Provide information about the goal of the problem-solving task (Mulet &amp; Vidal, 2008, pp. 104–105).</td>
<td>• Enable access to entry-level information about the goal of test and test item design, such as the website about test development by ETS (2010) • Enable access to online instruction about test and test item development such as websites about test development by ETS (2010), National CITO K12 test of the Netherlands (NPS, 2010), or resources from higher education</td>
<td>• Provide system-initiated prompts during first-time use through obligatory viewing of a linear introduction or instruction about test and test item design • Let users make notes about this information in a notebook • Let users make comments about test items in a comment field • Provide system-initiated prompts to encourage users to make notes in a notebook or note comments on test items in the item bank system</td>
</tr>
<tr>
<td></td>
<td>Provide structuring of the process of creative problem solving: emphasise problem representation, solution searching, emphasise diverging, and converging (Liu, Chakrabarti, &amp; Bligh, 2003).</td>
<td>• Enable access to information about the concept of diverging and converging • Enable access to information about the concept of solving ill-structured or ill-defined problems • Enable access to information about linear structuring of test and test item design</td>
<td>• Provide linear instruction in the form of a video or self-paced PowerPoint presentation • Provide system-initiated prompts to encourage users to view a linear instruction in the form of a video or self-paced PowerPoint presentation • Provide the possibility that developers can indicate whether they are in the stage of</td>
</tr>
<tr>
<td>Steps to solve problems</td>
<td>General guidelines</td>
<td>Possible information sources available through panel A and B</td>
<td>Possible functions</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------</td>
<td>----------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>processes (Haladyna, 2006)</td>
<td>divergence or convergence for each test item to be developed and have the support information shown adapted accordingly.</td>
</tr>
<tr>
<td></td>
<td>Provide system-initiated prompt to remind the user to engage actively the option of working work in divergence or convergence mode</td>
<td></td>
<td>Provide system-initiated prompt to encourage the user to study information about high-quality test items</td>
</tr>
<tr>
<td></td>
<td>Provide best practices and successful solved cases (Gerjets et al., 2007)</td>
<td></td>
<td>Provide system-initiated prompts to encourage the user to study worked examples of test items and use comparison possibilities</td>
</tr>
<tr>
<td></td>
<td>Encourage the user to study best practices and solved cases, using comparison possibilities (Gerjets et al., 2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enable access to information that shows what high-quality test items are</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enable access to worked examples that show how test items can be designed, evaluated, and improved. Examples can be taken from various sources, but the best examples will have similarities in subject matter to the test items that are to be designed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enable access to worked examples with comparison possibilities for test items</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enable access to methods in which test items are derived from learning objectives, learning goals, and taxonomies.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provide a mechanism to let the user link from within the environment to other external test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Let the developer incorporate previous developed test items or test items from other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steps to solve problems</td>
<td>General guidelines</td>
<td>Possible information sources available through panel A and B</td>
<td>Possible functions</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td>sources into the working environment.</td>
<td>• Enable access to the <em>story of item development</em> by Russell E. Dewey (Dewey, 2007).&lt;br&gt;• Enable access to the story of essential characteristics of item writers (Scoring Office, 2009, Chapter 2)&lt;br&gt;• Enable access to the video depicting the steps that ETS undertakes to develop tests (ETS, 2010).&lt;br&gt;• Enable access to the story of the National CITO K12 test of the Netherlands (NPS, 2010).&lt;br&gt;• Enable access to information about time typically required to develop high-quality test items (e.g., Hartog (2008), Draaijer et al. (2006), Mayenga (2009) or Case et al. (2001))</td>
<td>• Use international interoperability standards in a system to facilitate exchange of data, such as IMS QTI&lt;br&gt;• Use texts, personal photographs, animation, and video to present and increase story impact.&lt;br&gt;• Provide a possibility of letting the user make notes about problems and solutions during the test item design process for later use&lt;br&gt;• Provide system-initiated prompts to encourage the user to study this information&lt;br&gt;• Provide access to online spreadsheets from within the production environment</td>
</tr>
<tr>
<td></td>
<td>Provide <em>stories</em> that show how problems can be solved, evaluated, and improved (Jonassen &amp; Hernandez-Serrano, 2002).</td>
<td>• Provide the possibility of monitoring development time (Hewett, 2005).</td>
<td></td>
</tr>
<tr>
<td>Steps to solve problems</td>
<td>General guidelines</td>
<td>Possible information sources available through panel A and B</td>
<td>Possible functions</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------</td>
<td>---------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td>Provide functionality to paraphrase, diagram, draw, map, and tabulate content (Mulet &amp; Vidal, 2008, pp. 104-105; Shneiderman et al., 2006, p. 2)</td>
<td>• Enable access to a calculating module with benchmark data for calculating the total time needed to develop an item bank of a certain size.</td>
<td>• Enable (automatic) logging of the time spent per test item</td>
</tr>
<tr>
<td>Step 2: Identify and clarify alternative opinions, positions, and perspectives of stakeholders</td>
<td>Provide information sharing and collaboration opportunities with experts (Hewett, 2005; Shneiderman et al., 2006). Support pain-free exploration and experimentation, such as a sandbox mode in a notebook (Greene, 2002)</td>
<td>• Involve experts such as instructional technologists, psychometricians, multimedia specialists, but also students (Haladyna, 2004, 2006) • Subject provisional test items to a review process by other teachers or students to elicit information on the possible quality and performance of a test item (Haladyna, 2004, 2006)</td>
<td>• Provide online information and test item sharing and collaboration options with experts and students, perhaps on the basis of shared notebooks • Provide system-initiated prompts to encourage the user to make use of a shared notebook • Provide system-initiated prompts to share intermediate test items with reviewers</td>
</tr>
<tr>
<td>Step 3: Generate ideas and possible problem solutions,</td>
<td>Provide general information about various idea-generating techniques (Gerjets et al.,</td>
<td>• Enable access to online creativity technique resources, for example Mindtools (MindTools, n.d.)</td>
<td>• Provide access to online creativity technique resources • Provide system-initiated prompts to motivate</td>
</tr>
<tr>
<td>Steps to solve problems</td>
<td>General guidelines</td>
<td>Possible information sources available through panel A and B</td>
<td>Possible functions</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------------</td>
<td>------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>search for solutions</td>
<td>2007; Shneiderman, 2002)</td>
<td>Provide information about specific domain-related strategies for idea development (Hewett, 2005; Shneiderman et al., 2006)</td>
<td>the user to use a creativity generating technique</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Provide stimuli to begin crafting test items based on the intended cognitive process, such as compact inspirational guidelines (Draaijer, Hartog, &amp; Hofstee, 2007), action verbs (Gronlund, 1998, p. 36), generic item shells (Carriveau, 2010; Haladyna &amp; Shindoll, 1989), and design patterns (Draaijer &amp; Hartog, 2007)</td>
<td>• Enable users to collect and integrate their own information sources for creativity stimulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Enable access to information about systematic item-generating procedures (Roid &amp; Haladyna, 1982; Wilbrink, 1983; Williams &amp; Haladyna, 1982)</td>
<td>• Provide a list by which users can select inspirational guidelines and rework content in fields accompanying a test item record</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Provide access to example test items</td>
<td>• Provide an online list by which the user can select an action verb which is then put in the stimulus of a new test item</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Stimulate the problem solver to develop preliminary ideas further instead of generating new ideas</td>
<td>• Provide an online list by which the user can select a generic item shell which is then put in the stimulus of a new test item</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Provide system-initiated prompts to motivate the user to develop preliminary test item ideas further instead of generating new ones</td>
<td>• Provide an option to let the user link from within the environment to other external test items and item banks, directly or nearly relevant to the intended domain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Provide an option for users to develop a list of item shells for their own domains</td>
<td>• Provide an option for users to develop a list of item shells for their own domains</td>
</tr>
<tr>
<td>Steps to solve problems</td>
<td>General guidelines</td>
<td>Possible information sources available through panel A and B</td>
<td>Possible functions</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Step 4: Assess the viability of alternative solutions by constructing arguments and articulating personal beliefs</td>
<td>• Remind users about the goal of the problem-solving task (Mulet &amp; Vidal, 2008, pp. 104–105) • Remind users of the requirements for appropriate solutions (Treffinger, Isaksen, &amp; Stead-Dorval, 2005)</td>
<td>• Enable access to information relating to the goal of the test item development process: What is a good test? What is a test item’s purpose? • Enable access to information regarding both the need and the option to rework content, to use creativity techniques, or to employ systematic test item development strategies to devise and refine test items • Enable access to information regarding converging guidelines for test item design, such as a checklist of Do’s and Don’ts on the basis of Haladyna et al. (2002), Donahue (2009), or Malamed (2013)</td>
<td>• Provide system-initiated prompts to encourage the user to study the goals of test item development, to rework content, to use prompts to devise and refine test items, and to study lists of Do’s and Don’ts • Provide system-initiated prompts when common item construction flaws are detected • Enable a system to monitor planning and use of resources, particularly the time spent on the development of test items (Hewett, 2005) • Provide system-initiated prompts to motivate the user to make notes of successful design attempts and what underlying strategy that was followed in the notebook • Provide access to psychometric data of previously administered tests • Enable import of psychometric analyses data in order to relate the results to developing new test items (Guzmán, Conejo, &amp; García-Hervás, 2005;</td>
</tr>
<tr>
<td>Steps to solve problems</td>
<td>General guidelines</td>
<td>Possible information sources available through panel A and B</td>
<td>Possible functions</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------</td>
<td>------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
</tbody>
</table>

- Provide an interface in which a test item developer can see combined information on test items and their psychometric characteristics (Guzmán et al., 2005)
- Provide system-initiated prompts to encourage the user to process information concerning the meaning of psychometric data
Table 3

*Overview of built in prompts to point the user to common test item construction flaws.*

<table>
<thead>
<tr>
<th>Advice from item-writing literature which is displayed to the user</th>
<th>Detection mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is advised <em>not</em> to start every alternative of a multiple-choice question starts with the same wording.</td>
<td>Compare first string of text of each alternative of a multiple-choice questions and provide a prompt when these strings are identical or nearly so.</td>
</tr>
<tr>
<td>It is advised <em>not</em> to repeat specific words of the stimulus of the test item in the alternatives.</td>
<td>Compare strings of text in the stimulus with strings of text in the answers and distractors. Provide a prompt when comparable specific wording is encountered.</td>
</tr>
<tr>
<td>It is advised to keep alternatives of a test item of approximately the same length. Novice test item writers often make the correct answer of a multiple-choice test items longer in order to ensure correctness. When students are aware of this, they select the longest answer more quickly.</td>
<td>Compare the length of the correct answers and distractors and provide a prompt when excessively unequal lengths are detected.</td>
</tr>
<tr>
<td>It is advised not to use negations or negative text constructions in a test item. Negations can result more often in misinterpretations and sometimes make cognitive processing harder.</td>
<td>Search for text strings “not”, “dis*”, “none”, “im*”, “un*” etc. and provide a prompt when these strings are detected.</td>
</tr>
<tr>
<td>It is advised not to use unspecified determiners in a stimulus or answers; especially for true-false test items, such determiners can cause ambiguity.</td>
<td>Search for text strings “some”, “few”, “a number”, “often”, “seldom”, “regularly”, “sometimes” etc. and provide a prompt when these strings are detected.</td>
</tr>
<tr>
<td>It is advised not to use absolute determiners in test items; especially in true-false test items, exceptions almost always exist.</td>
<td>Search for text strings “always”, “never”, etc. and provide a prompt when these strings are detected.</td>
</tr>
</tbody>
</table>
**Advice from item-writing literature which is displayed to the user**

<table>
<thead>
<tr>
<th>Advice</th>
<th>Detection mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is advised not to have very long stimuli or alternatives when they are not necessary.</td>
<td>Search for a threshold length of stimulus and answering options and provide a prompt when it is exceeded.</td>
</tr>
<tr>
<td>It is advised to present a problem to the student in statements such as: “which of the following statements is true?” or “what is correct?”.</td>
<td>Search for a threshold of minimal length of the stimulus and provide a prompt when such a very small stimulus is encountered.</td>
</tr>
<tr>
<td>It is advised not to use “all of the above” as the final distractor.</td>
<td>Search for text strings “all of the above” “both A, B, and C”, etc., in the last answering option and provide a prompt when such a text string is encountered.</td>
</tr>
<tr>
<td>It is advised not to use “none of the above” as the final distractor.</td>
<td>Search for text strings “none of the above” or “not A, B, or C” in the last answering option and provide a prompt when such a text string is encountered.</td>
</tr>
<tr>
<td>Do not use double proposition type questions. They make test items artificially more difficult.</td>
<td>Search for text strings such as “both statements are correct”, “both propositions are correct”, “both A and B are correct”, “both I and II are correct”, etc..</td>
</tr>
</tbody>
</table>
Table 4

Examples of generic item shells as embedded in the concept of the tool (based on Haladyna & Shindoll, 1989).

<table>
<thead>
<tr>
<th>Intended learning outcome</th>
<th>Shell</th>
</tr>
</thead>
</table>
| Knowing and understanding: check if the student can recall information regarding a specific subject | What is the best definition for ...?  
What is (not) characteristic of ...?  
What are the parts of the concept or principle?  
What is the history of the fact, concept, or principle?  
What are different categories within the concept or problem? |
| Critical thinking (evaluation): check if the student can evaluate characteristics of facts, procedures, principles, or theories. * | What is most effective (suitable) for ....?  
What is better (worse) ...?  
What is the most critical step in a process or procedure?  
If you know that X is true, what is then true about Y?  
What is (not) necessary in a procedure?  
What is the importance of a problem? Does it matter? |
| Critical thinking (prediction): check if the student can predict consequences of specific combinations of facts, procedures, principles or theories.  
Problem solving (given a scenario): check if a student can solve a problem on the basis of a specific situation and can select or implement optimal solutions. | What would happen if ...?  
If this happens, what would X do?  
On the basis of ...., what would X do?  
Given ...., what is the most important reason that ...?  
What is the nature of the problem?  
What do you need to solve this problem?  
What is a possible solution?  
What is the most effective (efficient) solution?  
Why is ... the most effective (efficient) solution? |

* Of course, a teacher should find novel examples, effects, causes and problems to query for understanding or application of knowledge.

For the interested reader: a great number start sentences are developed and listed by Carriveau (2010).