Chapter 7: Fostering the Process of Designing Original Test Items for Teachers in Higher Education

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
7 Fostering the Process of Designing Original Test Items for Teachers in Higher Education

7.1 Abstract

An important objective in test item development in higher education is that alongside test items querying for recall of information from instructional materials or lectures, test items that query higher-order skills are needed. Teachers in higher education therefore must develop test items in which instructional texts are rephrased and presented from different perspectives and which present novel problems and examples. Test items of this type are called original test items in this study. Generating such test items calls for creative effort. An experiment was conducted in which teachers were provided with a creativity technique and exerted extended effort in order to increase the originality of test items. The challenge inherent in such a technique is that it must lead to improved originality without extensive training or time to execute in order to be acceptable in the practice of higher education. The rationale for the creativity technique in the form of an idea leaflet is provided. The experiment showed that providing participants with such an idea leaflet improved the originality of test items if it was administered after initial exhaustion of ideas. Implications for studying creativity techniques, test item design, and training item writers are discussed.

Keywords: test item writing, question design, creativity, selected response questions, scamper
7.2 Background

In this paper, a study is presented into the effects of providing higher education teachers with a creativity technique intended to increase the quality of the selected response test items they develop for in-house developed tests. Specifically, the creativity technique aims to increase the degree of originality of the items developed.

Selected response test items for in-house developed tests make up the bulk of many quizzes and exams in higher education, especially in domains with large enrolments such as nursing, medicine, law, business, management, economics, social sciences, and psychology. Selected response test items keep teachers’ grading workloads at manageable levels while providing opportunities to query students for a broad set of knowledge and skills. The quality of individual test items determines to a major extent the quality of tests and assessment (Downing & Haladyna, 2006), and students direct their attention and effort regarding their learning in line with formal and informal assessment programming, form, and content (Snyder, 1971). Furthermore, as in higher education new test items need to be designed for each successive exam to prevent the negative effects of item exposure, teachers are under constant pressure to design new content for test items.

It is therefore crucial to successful higher education that teachers be able to generate a range of test items tailored closely to the various goals that test items help fulfil. Teachers should be able to design test items for rote learning purposes, but also for higher-order learning goals such as application of knowledge, critical thinking, and problem solving. Designing these test items demands that teachers be able to conceive items that consist of rephrased instructional material that is presented in a different manner (Anderson, 1972; Haladyna, 2004), employs synonyms, antonyms, and homonyms (Haladyna, 1997), uses misconceptions (Anderson, 1972) or related facts, concepts, and principles (Williams & Haladyna, 1982), presents novel situations, professional tasks, and novel problems (Anderson, 1987; Gerritsen-van Leeuwenkamp, 2012), and aims to engage students (Keller, 1983). Furthermore, distractors for multiple-choice test items must offer plausible, attractive, but incorrect or only partially correct information for the incompetent
or partially competent examinee (Nitko & Brookhart, 2007; Rodriguez, 2005). For the current study, we regard test items with one or more of these characteristics to be of higher quality and more original.

Simply stating what original test items are does not necessarily help in generating such items. Authors of test item development handbooks regard developing high-quality items as difficult (Carter, 1984; Haladyna, 1997; Osterlind, 1998) and finding plausible, attractive distractors especially challenging (Mayenga, 2009). According to some authors, creativity is essential to developing such high-quality items (Osterlind, 1998; Rikers, 1988; Vale, 1995). Besides content knowledge, training, and experience in test item design (Downing & Haladyna, 2006), creativity is necessary to achieve a high level of performance in the crucial task of test item design.

However, in the practice of higher education, it is difficult for teachers to develop original test items (Conole & Warburton, 2005; Downing, 2002; Jozefowicz et al., 2002; Popham, 2003; Van der Vleuten et al., 2004). For example, many teachers in higher education have little time, perform this task with limited or no formal training in test development, do not use an item writing guide, have scant access to expert help, or do not consult such supporting resources as are available (Carter, 1984; Downing, 2002, 2005; Jozefowicz et al., 2002; Osterlind, 1998). Further, as test items in general cannot be reused after exposure to students, the task of developing selected response test items is a recurring task for higher education many teachers (Anderson, 1987; Parkes, Fix, & Harris, 2003), which leads to increased difficulties in finding new ideas for test items.

As a remedy for these problems, tools and heuristics that teachers can access easily and learn without extensive training could well improve the test item design process. At least with respect to the originality of item content, an improved divergent production process of ideas (Acar & Runco, 2012) may be a fruitful avenue in test item design. Draaijer et al. (2016) as presented in Chapter 6, have designed an intervention and found that priming participants prior to test item design by instructing them explicitly to diverge and develop a concept map and by providing start sentences led to the generation of more original test items. That study
concluded that other creativity techniques could prove valuable, notably those that require as little effort as possible in terms of time training and time, as teachers in higher education have very full schedules. In this study, a specific, brief creativity technique is therefore outlined and studied for its effectiveness.

7.3 Theoretical framework

The current study focuses on a specific brief intervention that could be used as a cognitive strategy to improve the divergent production of test items. This intervention fits in with the cognitive model as presented in Chapter 2, in which divergent and convergent production of test items are central. In particular, the identification of the importance of the divergent phase in test item design is an addition to existing views on test item design.

The intervention studied in the current chapter consists of providing teachers with a job aid (Rosset & Schafer, 2006) in the form of an idea leaflet that contains a set of stimuli (cues, hints, and prompts) derived from test item design literature and from general creativity techniques. The aim of the stimuli is to encourage the teacher to generate different perspectives on the instructional material and their own cognitive schemata of instructional information. Using the stimuli aims to mitigate the effect of design fixation (Jansson & Smith, 1991; Linsey et al., 2010), which is the phenomenon by which individuals in a creative design task adhere unconsciously to a set of ideas and concepts presented earlier or already generated. This adherence limits their output of new categories of ideas and hence the number of original ideas. As a result of generating different perspectives, an increase in the possibility of finding novel and original ideas for test items could emerge. Creativity research in organization science, management science, and industrial and creative design science suggests in general that it is likely that such creativity techniques lead to an increase in generating more original ideas (R. C. Anderson & Pichert, 1978; Chulvi, Mulet, Chakrabarti, López-Mesa, & González-Cruz, 2011; Daly, Yilmaz, Christian, Seifert, & Gonzalez, 2012; De Bono, 1993; De la Harpe, 2006; Eberle, 1972; López-Mesa, Mulet, Vidal, & Thompson, 2011; Ma, 2006; Scott, Leritz, & Mumford, 2004; Waks, 1997).
The extent of achieving positive effects by applying a creativity technique as a support tool for a particular task is however dependent on a number of factors. In the current study, three factors are of particular importance: the fit for purpose of the creativity technique chosen, the quality of implementation of that technique, and the amount and type of extended effort when applying the technique.

First, some techniques simply yield no effects *per se* (Ma, 2006) or yield no effects if the technique is not sufficiently tailored to the actual domain for which it is needed (Kilgour & Koslow, 2009; Scott et al., 2004). This means that not every technique is suitable for each situation and that the effectiveness of techniques is likely to increase if the key characteristics of the domain for which problems are to be solved are incorporated into the technique or its training. For example, some techniques lend themselves well to engineering problems, while others are better suited to tasks in the arts, business, or management fields.

Second, the quality of implementation such as training intensity, repeated practice, and instruction details are also influencing factors. When these factors are more appropriate or present to a higher degree, it is more likely that positive effects will occur. This follows from studies in which techniques did not show results when training was limited in duration and effort (Basadur & Thompson, 1986; Parnes, 1961; Scott et al., 2004) or when the technique was carried out with insufficient quality of implementation (Paulus, Nakui, Putman, & Brown, 2006), such as a lack of clarity of the procedure or instructions.

Third, the duration of an intervention is of particular interest because ideas generated by individuals tend to become more creative over time, especially when *extended* effort can be realized (Beaty & Silvia, 2012; Basadur & Thompson, 1986; Parnes, 1961). Extended effort refers to a prolonged or renewed endeavor to generate new ideas after exhaustion of initial ideas. What extended effort means in practice depends heavily on the individual, task, and context. In the context of the current study, initial effort is expected to range from minutes to an hour. Research has shown that extended effort generally leads to the generation of more creative ideas (Brophy, 1998; Dodds, Smith, & Ward, 2002; Dodds, Ward, & Smith, 2003; Sio & Ormerod, 2009). However, if the duration of a creative problem-solving task
is too short and no extended effort can be exerted, it is likely that the generation of more original ideas will be limited.

To elaborate, it is a well-known phenomenon that the first sets of ideas generated by individuals, such as the Alternate Uses Test (Christensen, Guilford, Merrifield, & Wilson, 1960), contain mainly obvious ideas; only after more time do more creative ideas emerge. However, after even more time, exhaustion of ideas does occur and the emergence of new and original ideas declines (Basadur & Thompson, 1986; Beaty & Silvia, 2012). Yet, after that period and when some incubation has taken place and renewed effort can occur as a result of regained motivation or deliberate interventions to extend effort, new and even more original ideas could emerge (Brophy, 1998; Dodds et al., 2002, 2003; Sio & Ormerod, 2009).

It might be expected that providing a creativity technique at the beginning of a creative problem-solving task would lead to an increase of the production of original ideas. To our knowledge, however, neither the extent nor conditions of applying a creativity technique for test item design at different stages of the idea generation process has been studied systematically. At the start of a task, an individual might simply be busy with getting out the obvious ideas, learning the task, and grasping the problem to solve. Using a creativity technique from the start might lead to cognitive overloading (Kirsh, 2000), causing an intervention to be ineffective. The phenomenon of expertise reversal could also occur (Kalyuga, Ayres, Chandler, & Sweller, 2003), by which experts perform even worse than novices. However, when the problem of the task is better understood by the individual, supporting interventions could be more effective.

The factors identified above present a challenge regarding the design of an effective creativity technique, as a process intervention for test item design for higher education teachers; the challenge lies in balancing the need to design a technique that is concise and general with the requirement of achieving a meaningful effect. A technique should be brief because higher education teachers cannot be forced to apply such a technique; they must choose it voluntarily because it works for them. Teachers will not use elaborate or demanding design techniques, given their chronic lack of time (Carter, 1984; Downing, 2002, 2005; Jozefowicz et al., 2002;
A brief intervention could however result in limited or absent effects. Second, in order to obtain an intervention that is applicable to more than one domain, the intervention must contain cognitive support that is generic and not specific to one domain. As domain-specific techniques have a higher chance to result in positive effects, a generic intervention limits the possible effectiveness. Both factors therefore limit the effectiveness of any proposed technique. Given that the goal of any creativity intervention is to generate an effect, these contradictory requirements need to be managed carefully in designing a creativity technique.

Further, in order to be most attractive to apply it in practice, a creativity technique must demonstrate a positive impact as soon as possible in the process of generating test item ideas. However, the effect of applying a technique at different stages of this process is unknown. Can an intervention be effective when a teacher is still in the first stages of idea generation? Is a teacher in the first phases not too busy working out the actual test item design problem to tackle a creativity technique and risk becoming cognitively overloaded (Kirsh, 2000)?

Therefore, in studying a brief creativity technique, two situations in which it is applied should be distinguished: when a teacher starts to design test items for a specific topic for the very first time and when a teacher has already put in some effort and has depleted the initial idea reservoir and is entering the extended effort stage.

### 7.4 The idea leaflet

The challenge of designing an attractive, brief, and effective intervention was answered in this study by assembling a one-page document as a job aid (Rossett & Schafer, 2006) that contains stimuli as effective as possible to support the process of perspective shifting and putting in effort during the task to design test items (see Figure 1).
Instruction (5 mins)

**Break free from the text**

In order to increase the number of possible good test items for an instructional text which – along with factual recall test items – query students on the application of knowledge, it is important to **break free** from that original instructional text. It is important to **stimulate the flow of thoughts** and ideas and to **incorporate these ideas in already generated test items and develop them further**. This increases the divergence of ideas; more and more multiform test items come into existence.

**Method**

The method below describes how to stimulate this process for developing such test items.

1. Identify keywords and primary concepts in the text; write them down.
2. Study each of the bullets in the box below. Each bullet is a suggestion on perceiving the content of an instructional text in a different way. Try to employ each suggestion with the instructional text provided, working that out in catchphrases and ideas. Devote plenty of time to this step.

Given the instructional text, develop ideas on the basis of the suggestions below:

- Synonyms – Antonyms – Analogies – Homonyms – Equivalents
- Examples – Counterexamples
- Misconceptions
- Related concepts – Sub-concepts
- Contexts
- Causes, Consequences
- Reversal of reasoning
- Enlarge, Reduce
- What if you use the concept in a different way?
- What would happen if you changed part of the concept?

3. Think up test items and distractors for those items using the bullets. Additionally, try to move the bullets further along.

Think about the steps and bullets for a while. Continue to the task itself when the experiment leader signals to do so.

*Figure 1. The idea leaflet (originally written in Dutch)*
It was decided that part ① of the idea leaflet should consist of a statement regarding the goal that the teacher was to pursue during test item design. The rationale was to direct the teacher to the concept of and need for divergence in test item design and encourage the teacher to approach the task with a positive attitude and full effort. This is a form of priming; it was anticipated that teachers that were primed for and aware of the process that they were entering could perform better, as has been found in other studies (Baumeister, Schmeichel, DeWall, & Vohs, 2007; Hacker, Dunlosky, & Graesser, 2009).

For content area ②, various heuristics from item writing and general creativity literature to support creativity were selected for their ease of application without specific training or instructions, brevity, general applicability, and a minimally anecdotal track record of effectiveness as reported in the literature.

Based on Miller, Williams, and Haladyna (1978; 1982), content area ② contains statements to condense information, rework original material, and use similes, metaphors, analogies, homonyms, and antonyms. Further, heuristics based on Bormuth’s “Wh” technique (1970) were selected. With the “Wh” technique, each possible information unit for a test was transformed into a test item beginning with one of “Who,” “Which,” “When,” “Why,” or “What” regarding that unit.

Additionally, stimuli originating from test item design literature which suggested specifically that participants think about examples and counterexamples, misconceptions, related concepts, sub-concepts, contexts, causes, and consequences (Anderson, 1972; Crouch & Mazur, 2001; Ebel, 1979) were included in the idea leaflet.

Based on general creativity technique literature, content area ② also contains stimuli to direct the teacher to think of more general associations with the instructional material. First, association questions were taken from the scamper technique (Eberle, 1996; Meador, 1997; Michalko, 2006; Osborn, 1979). Scamper is an acronym for the verbs “substitute, combine, adapt, modify/magnify/minify, put to other uses, eliminate or reverse/rearrange” and is a well-known creativity technique in the area of design and engineering. These verbs direct the individual to change perspective on a problem or information using those verbs. The scamper
technique resembles the suggestions for test item design by Ebel (1972), who describes several options for the development of true-false test items aimed at improving distractor development. Ebel’s proposals involve restating 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 or even impossible circumstances.

7.5 Research questions

As it is the aim of this study to investigate the extent of effectiveness of a brief creativity technique (the idea leaflet), one control condition and two experimental conditions were created. For the control condition, participants were instructed to develop test items without any support at all. In the first experimental condition, the participants received the idea leaflet before beginning the task, which enabled assessing whether the intervention would have an effect in the first phase of divergence, as well as in a later phase. In the second experimental condition, the participants received the leaflet at a later point in time. The effect of exhaustion of ideas for test items was operationalized in the experiment by setting a fixed amount of time for all participants for the first phase of item development time. The time for the first phase was set at 20 minutes development time in line with estimations found in the literature (Case, Holtzman, & Ripkey, 2001; Mayenga, 2009) and on the basis of informal pilot efforts. 20 minutes was sufficient for participants to exhaust their initial sources and effort for ideas. Entering the second phase, it was assumed that extended effort was exerted by the participants.
The following two research questions were formulated to assess the extent in which the idea leaflet affected the degree of originality of teacher developed test items.

To what extent does the presentation and use of an idea leaflet result in test items that are more original as compared to not using an idea leaflet,

a. when the idea leaflet is presented prior to a test item design task and used in the first phase of idea generation?

b. when the idea leaflet is presented and used after initial exhaustion of ideas?

It must be noted that the number of developed test items was not incorporated in the research questions, as the previous study presented in Chapter 6, had shown no relation between the number of developed test items and mean degree of originality was found.

7.6 Method

Participants

The participants were recruited from individuals in a higher education teacher training program at one university. The individuals were both teacher trainers and educational advisors; all had at least a Master’s degree. This group of participants was chosen as it was expected that they had some experience in test item design and would be representative of teachers in higher education. The participants took part voluntarily without financial reward. A total of 13 males and 28 females participated, with a mean age of 37 years.

The participants were assigned to one of the three conditions of no support (they did not receive the leaflet), prior support (they obtained the leaflet before beginning), or midway support (they received the leaflet partway through the task). The participants were randomly assigned to the conditions and were not informed to what condition they were assigned. Seven participants were assigned to no support, 13 participants to prior support, and 21 participants to midway support.
Procedure

Participants in the experiment were assigned the task of developing the best possible selected response test items for the concept of mercantilism, as shown in Figure 2.

Main Task
Generate as many good selected response test items as possible (true-false, multiple-choice, matching) on the basis of the text presented in the box. You have a total of 40 minutes development time.

The concept of mercantilism
The economic philosophy of mercantilism is intended to serve the interest of a nation, if necessary at the cost of other nations. This goal should be achieved by adopting governmental strictures that favor exports and discourage imports.

Enter your questions in the online word processor at: <URL>

Figure 2. Test item design task instruction for the experiment (translated from Dutch).

The concept of mercantilism was chosen for two reasons. First, the concept of mercantilism is multifaceted but reasonably easy to grasp using general knowledge and thus allows for the generation of ideas by both novices and domain experts. Second, the availability of example true-false test items as developed by Ebel about mercantilism (1972) served as a point of reference for possibilities for the developed test items in the experiment. An example of the use of the scamper/Ebel technique with regard to mercantilism is shown in Table 1 of Appendix A of this chapter. The participants did not receive additional instruction on the goals or use of the leaflet beyond what it contained. In all conditions the participants received two verbal statements to “keep on going as well as you can.”

The task instruction and information provided about mercantilism is very brief. The task required the participants to construct meaning actively on the basis of this short paragraph using their prior knowledge on the logic, history, cultural,
economic and political aspects, causes, consequences, examples, counterexamples, wordings, presentation and personal beliefs of the central concept discussed. This short paragraph left wide room for interpretation and construction of meaning and hence originality.

In the no support condition, participants were not given the idea leaflet. In the two experimental conditions, the participants were instructed to read and apply the guidelines from the idea leaflet. In the first experimental condition, the intervention occurred prior to executing the task to develop test items. In the other experimental condition, the intervention was positioned midway during the execution of the task. Figure 3 summarizes the experimental setup and timing.

Participants in the no support condition were allotted 40 minutes to perform the task; after the first 20 minutes, they were encouraged to continue with one verbal statement of the researcher: “keep on going as well as you can.” If they indicated that they could not think up more ideas, they were motivated a second time, but received no further instructions or encouragement. Participants in the prior support condition were instructed before the execution of the test item design task to read the idea leaflet and consider the support it could offer for the question design task in general and successively being assigned the task of applying the ideas in the idea leaflet to the mercantilism paragraph. Following the application of the idea leaflet content to the concept of mercantilism, they were instructed to develop test items for 40 minutes. Again, after 20 minutes they were motivated to continue and if they indicated that they were genuinely unable to conceive any new ideas, they were encouraged only one more time. The participants in the midway support condition began the process exactly as those in the no support condition. However, they were provided with the idea leaflet after 20 minutes of developing test items and were told to use it and continue to develop test items for another 20 minutes. We recorded the developed test items in phase one (the first 20 minutes of development time) and in phase two (the second 20 minutes or actual time elapsed).
### Time

<table>
<thead>
<tr>
<th>Condition</th>
<th>3-5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
<th>5 mins</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 mins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.** Overview of the conditions and timing of the experimental settings.
7.7  Instruments

Originality of test items

The rating of the items was carried out by the lead author based on a scoring rubric (Table 1). To reinforce the validity of the scoring scale, the second author of this study also scored test items after receiving training in the scoring task. A correlation of 0.8 was achieved, which is acceptable in the view of research regarding originality scoring of creative artifacts (Amabile, 1983; Caroff & Besançon, 2008; Runco & Charles, 1993); the scores of the lead author were used for analysis purposes.

Participants’ expertise

As it is likely that experts are able to attain a higher degree of achievement in a domain more easily than novices (Baumeister, Masicampo, & Vohs, 2011; Beaty & Silvia, 2012; Chi, 2006; Ericsson, 1999; Regehr & Norman, 1996), expertise was included as a covariate in the design. The degree of the participants’ expertise regarding the concept of mercantilism was determined on the basis of the teachers’ subject areas. Closely-related domains were defined as business, economics, governance, philosophy, law, and history. Participants in those domains were assigned a value of 2, while participants in fields that had a more distant connection to mercantilism, such as theology and sociology, were assigned a value of 1. Finally, participants working in totally unrelated domains such as chemistry, medicine, arts, or general education were assigned a value of 0. Table 2 provides a summary of the distribution of these scores.
Table 1

*Scoring criteria for originality measure of the developed test items.*

<table>
<thead>
<tr>
<th>Score</th>
<th>Scoring criteria</th>
<th>Example test items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The test item text remains very close to the passage text, barely paraphrasing, no new information added.</td>
<td>Mercantilism is aimed at serving other nations, even at the expense of the national interest. (True or False)\n</td>
</tr>
<tr>
<td>2</td>
<td>The text of the test item is paraphrased from the original passage. Some implications of mercantilism are developed or are rephrased.</td>
<td>Mercantilism is against government interference in the economy. (True or False)\n</td>
</tr>
<tr>
<td>3</td>
<td>Some new ideas are added to the test item. Examples include involving political parties, the origin of the concept, the concept’s philosophical implications, etc.</td>
<td>Mercantilism is an attack on the autonomy of the individual. (True or False)\n</td>
</tr>
<tr>
<td>4</td>
<td>The test item adopts a new approach to the problem, challenging the original proposition.</td>
<td>Mercantilism relates to government measures as ...\nA) Freud relates to psychoanalysis\nB) journalism relates to newspapers\nC) this exam relates to the university\nD) the cello relates to the piano\n</td>
</tr>
</tbody>
</table>
Table 2

Summary of domains, with scores regarding domain relatedness awarded and number of participants in each condition.

<table>
<thead>
<tr>
<th>Score</th>
<th>Domains</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>17</td>
<td>24</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>32</td>
<td>41</td>
</tr>
</tbody>
</table>

The validation, by means of an independent judgement of this classification, was carried out by a university teacher of philosophy who teaches mercantilism to students in a faculty of economics, business, and technology studies. After initial presentation of the task to estimate the possible expertise of a person regarding the concept of mercantilism based on the domain in which the participants work, a correlation of 0.87 was found between both raters which was deemed sufficient to use the original domain classification as a measure for the participants’ existing expertise in the concept of mercantilism.

7.8 Materials

Each participant received one A4 piece of paper containing the description of the task and indicating how text and test items should be entered in an online word processor that recorded all keystrokes. Because the instructions were printed on paper, the participants could work on the task and consult the idea leaflet, the task, and the paragraph about mercantilism at the same time. The participants were seated in a computer room to perform the task. The experimenter was the lead author of this article. The experimenter was always present during the experiment to keep track of the timing of the experiment, answer any questions, or address any technical issues. Answering questions regarding the idea leaflet by the experimenter was restricted to clarifying information about the meaning of specific terms.
7.9 Analysis

The study was set up as an experimental between-subjects design. Summary statistics were generated for the number of test items developed and the mean degree of test item originality for each participant and each condition for both phases.

One participant in the prior support conditions failed to design selected response test items for both phases. Additionally, one participant for the midway support condition failed to design selected response test items for phase one and two participants for the midway support condition failed to design selected response test items for phase two. Failing to design selected response test items means that either participants were unable to design test items as such or produced open-ended items.

For phases in which no selected response test items were developed, values were treated as zero for both number and mean of developed items. Data were treated as missing for mean originality calculations and regression analyses.

Regression analyses were performed using MPLus Version 6 software with maximum likelihood (ML) estimation. Two full models were built for the dependent variable mean originality of test items produced. One model was built for the mean originality produced in phase one and one model for the mean originality produced phase two. The independent variables were domain relatedness of the participants, condition of support, and the interaction between the conditions and domain relatedness of the participants. To test for the effect of the condition, the condition to which the participants were assigned was recoded into three dummy variables: no support, prior support, and midway support (0 = not a member, 1 = member in all cases). The interaction between condition and domain relatedness was included because of the hypothesis that the intervention could vary in effectiveness depending on the domain knowledge of a participant. Variables were centered before the creation of these interaction terms. Further, mean originality of phase one was included as a covariate in the regression analyses for predicting the mean originality of phase two, as it is likely that the initial level of originality was a predictor for the level of originality in phase two. As the combined effects of the
independent variables on the outcome were unknown, backward elimination was applied by successively removing the least significant variables from the model until only significant predictive variables remained.

7.10 Results

Descriptive Statistics

Table 3 presents summary statistics for the number of developed test items for each phase and each condition as well as the mean originality of test items for each phase and condition.

Table 3
Summary statistics for the sum and mean of the number and mean of originality of developed test items for the two phases and three conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Phase one</th>
<th></th>
<th>Phase two</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of</td>
<td>Originality of</td>
<td>Number of</td>
<td>Originality of</td>
</tr>
<tr>
<td></td>
<td>items</td>
<td>items</td>
<td>items</td>
<td>items</td>
</tr>
<tr>
<td></td>
<td>Particip.</td>
<td>Mean (st. dev.)</td>
<td>Particip.</td>
<td>Mean (95% CI)</td>
</tr>
<tr>
<td>No support</td>
<td>7</td>
<td>8.29 (2.98)</td>
<td>7</td>
<td>1.91 (1.38-2.45)</td>
</tr>
<tr>
<td>Prior support</td>
<td>13</td>
<td>7.00 (3.37)</td>
<td>12</td>
<td>2.16 (1.77-2.54)</td>
</tr>
<tr>
<td>Midway support</td>
<td>21</td>
<td>7.29 (4.08)</td>
<td>20</td>
<td>2.16 (1.90-2.36)</td>
</tr>
</tbody>
</table>

Effect of the interventions

First, a full model predicting mean originality for phase one showed only domain relatedness as a significant predictor \((p < 0.05)\). Further successive models in which the least significant variables were eliminated in each new model yielded no significant remaining predictors.
Second, Table 4 presents the results of regression models predicting the mean of originality of test items for phase two. By successively eliminating the least significant variables from the model, a final model significantly predicting the mean originality for test items for phase 2 remained, with $R^2 = .25$, $F(2, 34) = 5.45$, $p < .01$, with two significant predictive variables: covariate mean originality phase one, $b = 0.517$, $p < .01$ ($\beta = 0.38$), and midway support condition, $b = 0.433$, $p < .05$ ($\beta = 0.30$).

Table 4

Summary of the regression analysis for variables predicting the mean originality of test items in phase two.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Full model</th>
<th></th>
<th>Final model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain relatedness of the participant</td>
<td>0.361</td>
<td>0.194</td>
<td>0.063</td>
<td></td>
</tr>
<tr>
<td>Mean originality of test items of phase one</td>
<td>0.428</td>
<td>0.194</td>
<td>0.027</td>
<td>0.517</td>
</tr>
<tr>
<td>Prior support</td>
<td>0.356</td>
<td>0.375</td>
<td>0.318</td>
<td>0.433</td>
</tr>
<tr>
<td>Midway support</td>
<td>0.442</td>
<td>0.290</td>
<td>0.128</td>
<td></td>
</tr>
<tr>
<td>Domain relatedness x prior support</td>
<td>-0.580</td>
<td>0.673</td>
<td>0.398</td>
<td></td>
</tr>
<tr>
<td>Domain relatedness x midway support</td>
<td>-0.798</td>
<td>0.512</td>
<td>0.119</td>
<td></td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.376</td>
<td></td>
<td>0.247</td>
<td></td>
</tr>
<tr>
<td>Aikaike (AIC)</td>
<td>248.278</td>
<td></td>
<td>189.332</td>
<td></td>
</tr>
</tbody>
</table>

*Note. N participants = 37; b = unstandardized regression coefficient; SE = standard error; p = two-sided p-value; all three support conditions were coded as dummy variables, with 1 indicating a member and 0 indicating not a member.
7.11 Conclusion

For research question one, concerning the extent to which providing participants, prior to a test item design task, with an intervention to increase the mean degree of originality of test items designed, no effect was found.

For research question two, regarding the effect of providing participants after initial exhaustion of ideas during the task with support intervention, the experiment showed that the originality of test items for phase two did increase. The main conclusion of this study therefore is that the intervention as designed has been effective when provided to the participants after an initial phase of exhaustion of ideas, but not effective when provided from the start.

7.12 Discussion

In view of the results of this study, a number of issues regarding the setup of the study and a better understanding of the task of test item generation can be identified.

First, the findings of the current study are in line with the studies of Mayenga’s (2009) work on the difficulty of developing test items. In particular, Mayenga established that the development of plausible distractors was found to be ‘the hardest’ part of test item design and took the most time in relative terms, especially the third distractor of a four-option multiple choice test item. This is in line with the present study’s findings that it takes extended effort to find more original ideas from instructional texts.

Second, in all conditions the participants received two verbal statements to “keep on going as well as you can.” This implies that in effect, the effect of the idea leaflet itself was discerned from the effect of the two oral statements and the leaflet combined and hence the effect of the idea leaflet itself was measured. This is an encouraging finding, as it could have been the case that the verbal encouragement to “keep on going as well as you can” on its own had an impact on the increase of the degree of originality by generating extended effort. The results, however, showed that encouragement alone did not result in an effective exertion of effort, as
there was no effect on the degree of originality for the control and prior-support conditions.

Third, the findings provide support for the proposition of Draaijer et al. (Chapter 2) that developing test items can be regarded a creative problem-solving task. In particular, the findings from creativity research that extended effort in creative problem solving results in a higher degree of originality (Beaty & Silvia, 2012; Basadur & Thompson, 1986; Parnes, 1961) were also found in the current study. In line with those findings, this study provides more substance to the cognitive process model described in Chapter 2. The factors of duration, quality of instruction, and extended effort must indeed be regarded as important factors that have an impact on the result of a test item design task. Additionally, the study supports the notion that the cognitive process model helps in becoming aware of opportunities for generating feasible interventions that can bring about an effect in practical design situations.

Fourth, the study revealed important insights into and further questions about the fine details of providing support to teachers. We point especially to the fact that only the condition in which the idea leaflet was introduced midway through the design process resulted in a significant increase in the degree of originality. Though it was expected that providing the idea leaflet prior to the task would have an effect, the analysis did not bear this out. The failure to observe any effect could be due to the limited power of the experiment with regard to phase one, as only a limited number of variables were included to explain the level of originality achieved by an individual. However, we have already speculated that participants in the prior support condition could be too cognitively overloaded in the first phases of idea generation to process the intervention effectively (Kirsh, 2000). This possibility suggests that the effect of the intervention could depend on the level of understanding of the task and problem that needs to be solved. In particular, at the end of the initial phase of idea generation, the perspective on the problem could have shifted sufficiently from a state of feeling simply the need to generate ideas, to a richer understanding of the need for additional help to be able to generate new ideas.
At the same time, though, the idea leaflet was also at the disposal in the second phase for participants in the prior support condition, and was apparently not effective. It seems that explicit presentation and use of the idea leaflet midway the task is needed to invoke more effective cognitive processing and divergent production. As teachers in practice are likely often in a situation that the preliminary and obvious ideas for test items have already been generated, the situation is more likely that teachers have reformulated their test item design problem into actively seeking a support intervention to stimulate divergent production of ideas. Hence, the finding that an increase in originality was found only as a result of extended effort must be acknowledged with respect to the usefulness of the designed intervention in practice.

**Limitations and further research**

This study has certain limitations that should be taken into consideration when interpreting the results.

First, the amount of allotted time provided to the participants could call for further attention; in the actual practice of higher education, it is likely that teachers could not spend twenty minutes on such a task or that the particular act of reinterpreting the material with the intervention will not take place. In addition, reaching a stage of depletion of initial ideas was operationalized in the study by setting a fixed time limit. The time to reach depletion of initial ideas, however, can and almost certainly does vary between individuals. An experiment in which the amount of development time or a cutoff point for individual participants based on a more precise measure of exhaustion would be valuable for further research.

Second, although the experiment revealed significant effects of the intervention, the statistical power of this experiment was limited, so it is possible that other effects were present but could not be detected in this experimental setup. For example, no interaction effects between expertise and condition was observed, suggesting that both novice and expert participants could take advantage of the intervention proposed. The absence of interaction effects should therefore not be assumed. Additionally, using more variables to explain the level of original ideas generated
for phase one could lead to more insight regarding the effect of those variables. For phase two, this limitation is much less of a problem, as the strong covariate of the originality of the ideas of phase one was incorporated into the regression model.

Third, the intervention was used by the participants only once in a limited time span. It would be valuable to learn what the intervention’s longitudinal use and effects might be. Would for example repeated employment of the idea leaflet lead to internalization of the stimuli and thus to improved performance by teachers? Would repeated employment invoke more effective cognitive processing and divergent production, as was shown in the second phase of the mid-way support group? Research investigating the longitudinal effects of creativity training suggest that these are certainly possible if not probably outcomes (Scott et al., 2004). The latter question is obviously very important, as the bulk of the literature on learning test item design emphasizes skills improvement through multiple rounds of designing and evaluating the quality of test items.

**Practical implications**

Finally, this study has implications for test item design training and support for teachers in higher education. First, by paying attention to the divergent aspect of test item design in support and training for teachers, it raises the awareness among teachers that developing test items entails the intentional use of creative effort beyond their own domain knowledge and the protocols for question construction. Further, the developed and easily-deployable creativity technique provides teachers with an additional support for finding original test item ideas when they exhaust their own ideas. When applied, this technique could raise the quality of examination in higher education.
Chapter 7

7.13 References


Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking


7.14 Appendix A

Examples of test items for the concept of mercantilism

Table 5

An example of the application of the scamper technique to the instructional paragraph of mercantilism.

<table>
<thead>
<tr>
<th>Question</th>
<th>Associations</th>
<th>Possible test items</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>What can I substitute in the mercantilism paragraph or for mercantilism itself?</td>
<td>Mercantilism $\rightarrow$ Capitalism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nations $\rightarrow$ Companies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nations $\rightarrow$ Political Parties</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Government measures $\rightarrow$ Natural Causes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nations $\rightarrow$ France, UK, China</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Companies $\rightarrow$ Ferrari, Bennet, Louis Vuitton, Aldi, Lidl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>War $\rightarrow$ Peace</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>&quot;How can I combine mercantilism with other activities/elements/functions/parts?&quot;</td>
<td>Mercantilism combined with democratic, autocratic, theocratic, totalitarian, oligarchic, technocratic, or communist principles.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Import/export combined with illegal trade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fair trade combined with mercantilism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monarchy combined with mercantilism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Education combined with mercantilism</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>Associations</td>
<td>Possible test items</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Mercantilism combined with nature.</td>
<td>What do mercantilism and capitalism have in common? Both … A) regard the concept of capital as centrally important. B) place the national interest first. C) are founded on international legal agreements. D) put the rights of religious groups first.</td>
<td></td>
</tr>
<tr>
<td>A “What can I adapt or copy from ideas other than mercantilism?”</td>
<td>Mercantilism adapted by principles of communism Mercantilism adapted by principles of fair trade Mercantilism adapted by principles of monarchy or democracy Mercantilism and legal matters</td>
<td></td>
</tr>
<tr>
<td>M “What can I magnify, minimize, or emphasize about mercantilism?”</td>
<td>Emphasize the role of government measures Emphasize the role of war Emphasize the role of the transport of goods Emphasize the role of the monarch Emphasize the role of the women’s power</td>
<td>The women’s rights movement is a predecessor of mercantilism. A) True B) False</td>
</tr>
<tr>
<td>P “How can I put mercantilism to other uses?”</td>
<td>Apply mercantilism to computer technology Apply mercantilism to religion Apply mercantilism to business organization Apply mercantilism to sports</td>
<td>As a phenomenon, what is sports most similar to? A) Democracy B) Mercantilism C) Socialism D) Capitalism E) Individualism</td>
</tr>
<tr>
<td>E “What can I eliminate or simplify in the concept of mercantilism?”</td>
<td>Would mercantilism be possible without government? Would mercantilism be possible without money or banks? Would mercantilism be possible without trade?</td>
<td>Without religious believers, mercantilism would be impossible. A) True B) False</td>
</tr>
<tr>
<td>Would mercantilism be possible without government? Would mercantilism be possible without money or banks? Would mercantilism be possible without trade?</td>
<td>Without a government, mercantilism would be useless. A) True B) False</td>
<td></td>
</tr>
<tr>
<td>R “How can I change, reorder, or reverse the Are countries selfish and does that induce mercantilism?”</td>
<td>Mercantilism will not work if three countries are involved in a trade transaction.</td>
<td></td>
</tr>
</tbody>
</table>

288
<table>
<thead>
<tr>
<th>Question</th>
<th>Associations</th>
<th>Possible test items</th>
</tr>
</thead>
</table>
| way in which mercantilism works.       | How many countries can be involved; one, two, three, or more? | A) True  
B) False  
Mercantilism is applicable only with three or more trading nations.  
A) True  
B) False |