Summary

Developing good tests and test items in higher education is extremely important. Test and test items determine to a large extent what students learn and how the domain or profession for which they are studying is experienced. Especially in the first years of higher education, selected response test items are used to keep grading within manageable bounds. Selected response test items include true-false questions, multiple-choice questions, multiple-correct questions, and ordering questions. In higher education, these test items are designed by teachers, and there is often criticism regarding their products. Too many context-free verbatim facts are queried or the test items are not clear enough. However, teachers often develop these test items without specific training and with very limited time. Further, it is difficult to work jointly and systematically with test experts to maintain skill levels. Teachers are often forced to engage in this activity alone. If training is offered, it proves not to be very effective in practice. The existing literature and support with regard to the design of test items is experienced by teachers in higher education as unhelpful and offers them little or no useful support in their daily practice. This situation is problematic and comes at the expense of the quality of assessment in higher education. Therefore, in this thesis, new ways to improve devising test items by teachers in higher education were examined.

The aim of this thesis was twofold. First, the objective was set out to understand and describe more fully how teachers generate and devise test items. Indeed, little is known about the conception and construction of test items from the teachers’ perspective or about the intensive cognitive process at work in that process. Without a proper understanding of this process, it makes little sense to develop methods to support the improvement of creating test items. Second, if the process of conceiving test items is understood more deeply, it may be more feasible to develop support for teachers that leads to better test items. The thesis examined what such improved support might be and whether it would be effective.

This thesis contains six studies that address in different ways the central research question. The study in Chapter 2 makes clear that the end product (the test item itself) is emphasized in the most of the literature for test item design; it presents
strict requirements that these items must comply with and are usually presented as lists of pitfalls and construction errors that should be avoided. This literature scarcely looks at methods that suggest how this result could be achieved. If methods for generating test items are addressed, then generating test items is presented as a more or less logical deductive process in which the test item designer, on the basis of an apparently clear starting point with respect to content representation and difficulty, needs to find a single correct test item that accords with that starting point. The design of test items is presented implicitly as a well-defined problem with a clear and correct solution.

Closer examination, however, reveals that the design of test items is based on a fundamentally different process. Designing test items can be understood as solving a poorly structured and poorly delineated and thus ill-defined creative design problem. Such problems do not have unambiguously determined starting conditions or clear conditions for the end point. Moreover, there are several ways to solve the problem. Multiple acceptable solutions are possible. In Chapter 2 of this thesis, these characteristics are described in fine detail.

Solving such problems requires a cognitive process of intensive divergent and convergent thinking. Divergent thinking refers to the process of generating multiple ideas and seeds for test items. Convergent thinking refers to the process of considering generated ideas critically and selecting potentially good ideas for further development and optimization. During this process, the perception of the problem changes continuously: "What is it exactly that I want to ask students and for what purpose? How do I articulate and adjust the test item so that the test item fits with what I think it should be querying and that the students will understand it as such? Which variant or modification of the test item fulfils this function best? How sure am I that competent students will likely choose the right answer to the test item and less competent likely will not?" A teacher has to make decisions on the basis of incomplete information and estimations. This process of problem solving is visualized in Chapter 2 in a simple figure based on the concept of creative problem solving as outlined by Isaksen and Treffinger (2004). Based on a description of this process, it was established that to create effective support for the process of test item design, such support should be geared to both the divergent
and convergent processes exhibited during the test item creation process. The divergent processes in creating test items have been especially underrepresented in the literature on test item development; the rest of the thesis therefore focuses mainly on the divergent thinking process for test item design.

In Chapter 3, an inventory and ordering was made of a number of inspirational guidelines that can help generate ideas for test items. The guidelines were inventoried from and intended for teachers in the domains of medicine and health and the life sciences. The structuring of the inspirational guidelines resulted in ten categories, each containing between three and ten relatively simple instructions. Each instruction may represent a source of inspiration, and thus can serve as an intervention or support resource, that can be used to create new test items. The inspirational guidelines were used by teachers and teacher assistants in 15 test item development projects. The use, the experiences, and the perceived usefulness of the inspirational guidelines were evaluated to establish under what conditions which guidelines were effective. It was concluded that four suggestions are always useful for any development project. However, if all suggestions are offered at once, teachers will likely feel overwhelmed. It was further concluded that being supported by an educational technologist is required to offer only a limited set of guidelines to teachers. These guidelines must be geared to specific situations to enable quick and effective cognitive processing and to prevent overload.

Chapter 4 dealt with a specific category of inspirational guidelines, the so-called design patterns. Design patterns are standard solutions for common design problems. Experts in a particular field often know nearly instantly which standard solutions fit for which categories of problems. Depending on the context and perception of the problem, the standard solution can be adapted. Identifying and describing design patterns for test items in higher education can offer teachers inspiration for generating their own test items. In Chapter 4 a descriptive model was developed for design patterns for test items. 10 design patterns based on the developed test items from the 15 test item development projects from Chapter 3 were fully developed.
In Chapter 5, the question was put forward as to how the computer can support teachers in the process of designing test items. The study presents a prototype of a computer program to support teachers in designing test items in which the central premise of the cognitive model as presented in Chapter 2 of this thesis was reflected in the features, content, and interactions of the user with the system. Specifically, the computer program offers support for orientation on the design problem, divergent thinking processes, and convergent thinking processes. The study also described how the interaction possibilities were designed in such a way that two main conflicting user interaction requirements are solved. These conflicting requirements are that support is not underused and that the user is not overloaded with information that hinders the user’s employment of the program. The program is used in the follow-up study in Chapter 6.

Finally, in Chapters 6 and 7, two experimental studies were presented in which the degree of effectiveness of specific creativity techniques were tested for their extent of improvement of divergent and convergent production of ideas for test items. These creativity techniques were identified in Chapters 2 and 5 of this thesis as possible effective interventions for the improvement of divergent and convergent production of such items.

Chapter 6 presents a study in which participants in an experiment needed to design test items for a specific topic in psychology. In the experimental condition, the participants were informed about the importance of divergent and convergent thinking to be able to design better test items, as a form of priming. To encourage divergent thinking even further, participants had to make a concept map and were provided with start sentences as leads for possible test items. In order to stimulate convergent thinking among participants in the experimental condition, a presentation about the importance of producing clear and unambiguous test items was provided and participants could consult a list containing common mistakes that should be avoided in that process. The results of the study showed that subjects in the experimental condition did not produce more test items, but on average did produce more original test items than subjects in the control condition, in which no support was offered. Compared to participants in the control condition,
participants in the experimental condition did not produce technically better test items. The conclusion of this study was that divergent intervention is effective, but convergent intervention is not.

Chapter 7 presents a study in which the effect of an intervention to encourage divergent thinking was investigated in a test item development task about a topic from economics. In that study, an intervention in the form of a job aid known as an idea leaflet was offered to participants in two experimental conditions. In the first, this idea leaflet was presented prior to the test item development task, while in the second it was offered after the subjects had produced the most obvious ideas and additional effort was needed to produce more ideas. The results of the study showed that subjects in the second experimental condition produced more original test items on average in contrast to participants in the control condition and in the first experimental condition. The conclusion of this study was thus that the intervention in the form of an idea leaflet is effective if test item designers have depleted their initial group of ideas and are urged to make an extra effort with the idea leaflet.

In the Summary and discussion chapter, the results of the studies are summarized and the significance of the results for educational practice and further research are discussed. All in all, it appeared that it is productive to consider designing test items as solving an ill-defined creative design problem. It offered a fundamentally different view of the problem of the test item design, leading to new opportunities to support teachers in the task. There is little reason to believe that this perspective is not useful for teachers in primary or secondary education as well. Specific support information, resources, and interventions were developed and studied for their degree of usefulness and effectiveness in either a practical or an experimental setting. It was found that the information, resources, and interventions can each be effective for improving the divergent production of ideas for test items, but not in all situations. Furthermore, support from educational developers is often needed to increase effectiveness.

Obviously, there remain other important questions open for further research. An important issue for improving developing test items in higher education remains
that supporting the convergent process requires extra attention. It is also necessary to examine whether and how the computer program would function if it were offered to teachers in real-world practice. In addition, the developed inspirational guides and the other interventions, such as the exact shape of priming, the various forms of possible task descriptions, forms of concept map task description, start sentences, and individual elements of the idea leaflet can and should all be scrutinized for their degree of effectiveness, in both experimental and practical situations.

From a theoretical and practical viewpoint, this thesis contributed to understanding and likely alleviating the widespread problem of designing test items in higher education. The developed process model makes the nature and complexity of the task of developing test items much clearer. In the model, the necessary cognitive processes involved in test item design from the perspective of the teachers are acknowledged; it can help in communicating more purposefully with teachers about this process and provide more guidance for their task. When the findings of this thesis reach teachers in higher education, it will likely contribute to a meaningful extent to improving the quality of testing and assessment in higher education.