UNIVERSITY OF TWENTE.

Cumulative Assessment in University of Twente's Mechanical Engineering Curriculum:

A Pilot Study

M.J.M. (Marlou) Stinenbosch Msc.
Master Thesis

July 27th, 2017

Supervisors:
Prof. Dr. Ir. B.P. (Bernard) Veldkamp
A.I. (Anniek) van den Hurk Msc.
Prof. Dr. Ir. A. (André) de Boer

Faculty of Behavioural, Management and Social Sciences

Master of Science in Educational Science and Technology

Educational Design and Effectiveness

July 27th, 2017

University of Twente

Student:	M.J.M. (Marlou) Stinenbosch Msc.
Student number:	s1529544
E-mail:	m.j.m.stinenbosch@student.utwente.nl
Institute:	University of Twente
Faculty:	Behavioural, Management and Social Sciences
Programme:	Master in Educational Science and Technology
First supervisor:	Prof. Dr. Ir. B.P. (Bernard) Veldkamp
Second supervisor:	A.I. (Anniek) van den Hurk Msc.
External supervisor:	Prof. Dr. Ir. A. (André) de Boer

Acknowledgements

I would first like to thank my first supervisor Bernard Veldkamp for his continuous support of my study, for his patience, motivation and immense knowledge. His guidance helped me in all times of the research.

Besides my first advisor, I would like to thank Anniek van den Hurk for being my second supervisor. Her guidance and support helped me to complete the research.

I would also like to thank André de Boer. His door was always open when I ran into a trouble or had a question about the continuation of the study. Without his passionate participation and input, the study could not have been successfully conducted.

Finally, I would like to express my profound gratitude to my roommates in HT4.15 for their openness and willingness to always explain me complicated Mechanical Engineering materials and for solving many puzzles during coffee breaks. A special thank goes to Bjorn and Sebastiaan, who helped me a lot with the development of questions and elaborations.

Thank you.

Marlou Stinenbosch
July 2017

Management Summary

Currently, students studying the bachelor program Mechanical Engineering (ME) at the University of Twente have a tendency to forget basic terms and theories necessary for consequent courses, according to the teachers. Teachers need to spend time on previously taught materials during their courses. The use of cumulative assessment (CA) can be a powerful promoter for students to increase knowledge retention.

This study developed CAs to measure students' knowledge retention of previously covered knowledge and applied this in a pilot study. The pilot study tested how a formative CA tool can be developed to gain insight in the current long-term knowledge retention of students. First, the basic learning goals were identified by teachers and multiple-choice (MC) questions were developed by them that were included in the CAs. Feedback on test results and elaborations was given to the students. All bachelor students of ME were asked to participate. Students were asked to participate twice during academic year 2016-2017, once at the end of quartile 3 (n = 269) and once at the end of quartile 4 (n = 95).

Data was collected by an open question to teachers to identify basic learning goals, MC questions and elaborations. Two CAs for students were used that consisted of MC questions measuring the basic knowledge they were expected to have. Reliability analysis on the developed CAs were done as well as analysis on the items in the CA used.

This study identified insufficient support throughout the faculty of ME for implementing CA. Support of all staff members is necessary when items need to be developed to be used in CAs. Furthermore, it appeared that teachers had no or less experience with developing MC questions for the CA. Analysis on the developed items in the CA used in this study showed an unreliable scale for all the six developed CAs. Item analysis revealed few questions that were identified as sufficient, most items were identified as being too easy, too difficult or that another alternative answer seemed to suit better. Furthermore, formative CA was done to provide the participating students with feedback on their test results and elaborations. The results indicated that students scored on average insufficient on the tests, but they wanted to read the elaborations and check how they answered the previous CA.

Keywords: Cumulative assessment; Formative assessment; Pilot testing; Higher education; Knowledge retention

Table of contents

List of tables

List of figures

List of abbreviations

AfL:	Assessment for Learning
CA:	Cumulative Assessment
DBDM:	Yield-Oriented Working
DT:	Diagnostic Testing
ECTS:	European Credit Transfer System
FA:	Formative Assessment
MC:	Multiple-Choice
ME:	Mechanical Engineering
NCLB:	No Child Left Behind
UT:	University of Twente
Rir:	Item Response Rest Correlation
Rit:	Item Response Total Correlation
SA:	Summative Assessment

1 Introduction: the rationale for this study

Students tend to postpone studying for a test and this increases risks for behaviours of suboptimal learning (Berg & Hofman, 2005; Bruinsma & Jansen, 2009; Schmidt et al., 2010; Van Eerde, 2000, 2003). Students studying for a test are depended on motivation, since motivation drives people towards reaching a goal and goals can only be achieved by displaying some kind of behaviour that makes the goal more achievable (Heckhausen, 1977). Hence, the motive to pursue academic studies drives students to study in order to pass exams successfully and thereby moving closer to their final goals (Schouwenburg & Groenewoud, 2001). However, motivation is not the only motive that drives students' behaviour. Assessment is seen as an important stimulator and motivator for the study behaviour of students (Bruijns, 2014; Gibbs, 2010). Moreover, assessment is the most powerful instrument for making a difference in study behaviours of students towards spending time on the course they study (Gibbs, 2010), as students' study behaviour is not only influenced by teaching, but also by the way they perceive the demands of the assessment system (Cohen-Schotanus, 1999). Additionally, providing feedback to support student learning is another powerful single influencer on learning and achievement (Hattie & Timperley, 2007).

Currently, teachers at the bachelor of Mechanical Engineering (ME) at the University of Twente (UT) experience that many students forget basic terms and theories that they need for subsequent courses. Courses within the bachelor are interrelated and knowledge retention is indispensable to complete successive courses. A lack of knowledge retention results in teachers needing to spend time on already taught materials instead of new materials. In the worst case, students fail courses. This results in study delay, which is an undesirable course of events for both the student and the teachers and the faculty.

In order to stimulate students to study more frequently and continuously, the number of tests can be increased, which might lead to more knowledge retention (Cohen-Schotanus, 1999; Jansen, 2004; Kerdijk, Cohen-Schotanus, Mulder, Muntinghe, & Tio, 2015). An assessment method that offers this possibility is cumulative assessment (CA). CA is a form of assessment that comprises multiple tests throughout one course and has been shown to be a powerful promoter for students to distribute their learning activities over a course and increases the time students spend on self-study (Bruijns, 2014; Kerdijk et al., 2015). Based on previous research, CA is defined as a comprehensive assessment reflecting the previously learned basic-objectives of the curriculum in this study (Roediger & Karpicke, 2006; Swanson, Holtzman, & Butler,

2010; Vleuten, Verwijnen, & Wijnen, 1996; Wrigley, Van Der Vleuten, Freeman, & Muijtjens, 2012).

CA combines several theories about assessment to stimulate students' learning to increase the students' test performances, since spending more time on self-study will result in better test performances. In addition, CA increases test performance, because CA assesses the learning materials of all preceding weeks and the same content will be repeated, tested and studied repeatedly (Cohen-Schotanus, 1999). Earlier studies confirmed that repeated testing of materials has a positive impact on knowledge retention (Bekkink, Donders, van Muijen, & Ruiter, 2012; Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). Moreover, research shows that if teachers determine the critical knowledge and skills students should know at the end of a course, emphasize these points in class and repeatedly test these, students will acquire the required critical knowledge (Roediger & Karpicke, 2006).

While much is known about the benefits of CA, implementing CA into a study program remains a challenge. This study investigates the development of CA in the bachelor of ME at the UT, since no other research was found that implemented formative CA into a technical study through end-of-course assessments. Research was found that implemented CA, but mandatory coursework was eliminated (Haugan, Lysebo, & Lauvas, 2017), assessments were done at the end of each week (De Paola & Scoppa, 2011; Kerdijk et al., 2015), or CA was done in summative way (De Paola & Scoppa, 2011; Kling, McCorkle, Miller, & Reardon, 2005; Mitra, Nagaraja, Ponnudurai, & Judson, 2009). Before implementing a CA in a study program, it is important to conduct a pilot study. This pilot study will pre-test the research instrument, which is a critical element prior to the implementation of CA in the study program (Van Teijlingen & Hundley, 2001). The pilot study is undertaken to improve the understanding of the main innovation being delivered by CA (Turner, 2005). Conducting this pilot study reduces the risk and the uncertainty associated with the implementation and increase the likelihood to succeed (Turner, 2005; Van Teijlingen & Hundley, 2001). For that reason, it is the goal of this study to develop and apply a pilot version of CA for the ME program at the UT.

2 Theoretical framework, research questions and scientific and practical relevance

This chapter describes the theoretical framework and starts with an elaboration of aspects of CA. Different types of assessment will be described next, followed by the research question and sub questions. Finally, the scientific and practical relevance will be discussed.

2.1 Aspects of cumulative assessment

The main goal of CA is to encourage students to retain previously learned material and attend to the interrelatedness among topics as new material is learned (Roediger & Karpicke, 2006; Swanson et al., 2010). This is achieved by CA used to systematically retest critical knowledge from previously covered topics as well as critical knowledge from the current courses (Swanson et al., 2010). The CA samples basic knowledge expected of students on completion of their course and the test provides feedback to assess their state of development to what is required. The test provides longitudinal assessment, comprehensive assessment and include feedback for students, which is in line with research of Wrigley et al. (2012) and Vleuten et al. (1996).

First, CA is longitudinal because it offers an opportunity to compare performance over time. Frequent measurement of all subsequent courses during the academic year of the program helps maximise the reliable tracking and monitoring students' developmental progression (Freeman, Van Der Vleuten, Nouns, & Ricketts, 2010; Wrigley et al., 2012). Research provided further support for testing learning multiple times per year instead of once at the end of a course (Wade et al., 2012). The longitudinal data of CA can also serve as a benchmarking instrument for the faculty by which to measure the quality of educational outcomes (Wrigley et al., 2012).

Second, the CA is comprehensive in that the assessments are developed by sampling from the entire curriculum (Wrigley et al., 2012). Students are tested on previously covered materials as well as the just-completed module (Swanson et al., 2010). Because of this comprehensiveness, CA discourages rote memorization and cramming and thus will increase the deep long-term knowledge retention (Vleuten et al., 1996; Wrigley et al., 2012).

Third, the CA can be made publicly available for the students after the tests and feedback with elaborations about each item can be provided. Students can assess their state of development in relation to what is required at the end of each module. These CAs will be valuable in analysing deficiencies and suggestions for remediation for students (Vleuten et al., 1996). This type of assessment is also called formative assessment (FA), because it provides

information about the test results to students and refers to assessment generating feedback on performance to improve and accelerate learning (Sadler, 1998).

2.2 Types of assessments

To structure the review of CA systems, it is common to distinguish that CA can be done in two ways: summative and formative. Summative assessment (SA) aims at identifying the learning state of students to decide about selection, classification, placements or certifications (Sadler, 1998). The point where the judgement about the state of students is given is a finality at the point of judgement (Garrison & Ehringhaus, 2007). The process of assessment is a single process, meaning that the judgement of FA is made according to the same process of SA (Taras, 2005). However, for an assessment to be formative, feedback is required that identifies the existence of a gap between the actual level of the work assessed and the required standard (Taras, 2005). FA also requires an indication of how the work can be improved to reach the goal and can be used to guide the learning process of the students (Garrison & Ehringhaus, 2007; Sadler, 1998). In this study, FA is used to assess the students because the goal is not to give a judgement about progress, but to identify a gap and helping the learner forward and therewith activating them to be owners of their own learning (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006).

A substantial research of Black and Wiliam (1998) showed that FA 'works', since it is effective in promoting student learning across several educational settings like disciplinary areas, types of outcomes and levels. Black and Wiliam (1998) developed a model that includes five key strategies on different levels (see Table 1), displaying three aspects of FA (where the learner is going, where the learner is right now, how to get there) that applies to different agents (teacher, peer, learner). Since FA is applied to CA, this model can be used to describe the different steps that needs to be undertaken for this study. However, FA includes more strategies than CA, like both peers and students that needs to be activated as instructional resources and owners for learning (Black & Wiliam, 1998). Therefore, only the first three key strategies of the model developed by Black and Wiliam (1998) were used for the of CA.

The first key strategy is to clarify and share learning intentions and criteria for success, which is the starting point for a successful FA. This first strategy applies to both teacher and learner to get insight in where the learner should go. The second strategy consists of engineering learning tasks that elicit the evidence of student understanding, which is subsequent to the first strategy. CA was developed to elicit this evidence and therewith describe the understanding of

the learner about where they are right now. Multiple-choice (MC) questions were developed by for the CA based on the basic learning goals as formulated in strategy 1 and done by teachers. The third strategy provides feedback to move the students forward. This was done by providing the learner its test results and elaborations, which aims at helping them understanding the materials better. Strategy 4 and 5 can only be done when the previous strategies are successfully completed and were no part of this study, since the focus was on the development of CA.

Table 1

Aspects of formative assessment (Black & Wiliam, 1998, p. 5)

<table>
<tr><th></th><th>Where the learner is going</th><th>Where the learner is right now</th><th>How to get there</th></tr>
<tr><td>Teacher</td><td rowspan="3">1. Clarifying learning intensions and criteria for success

Understanding and sharing learning intensions and criteria for success

Understanding learning intensions and criteria for success</td><td>2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding</td><td>3. Providing feedback that moves students forward</td></tr>
<tr><td>Peer</td><td colspan="2">4. Activating students as instructional resources for one another</td></tr>
<tr><td>Learner</td><td colspan="2">5. Activating students as the owners of their learning</td></tr>
</table>

Thus, this study aims to develop a pilot version for the implementation of CA in the bachelor curriculum of ME at the UT, since no other research was found that implemented formative CA in a technical study by end-of-course CAs. This study will identify where the students are going, followed by the identification of where the students are right now and finally, feedback will be given to move the students forward.

2.3 Research questions

Based on the theoretical framework, this study aims at developing and piloting a formative CA tool to measure students' knowledge retention in the ME curriculum and therefore, the main question is:

"How can a formative cumulative assessment tool be developed to measure students' knowledge retention level in University of Twente's Mechanical Engineering bachelor curriculum?"

To answer this main research question, three sub questions were developed:

1. Which knowledge should students possess at the end of specific bachelor Mechanical Engineering courses?
2. What is the reliability of items in the test used in the formative cumulative assessment for bachelor Mechanical Engineering students?
3. How to provide Mechanical Engineering students with feedback in a manner that they can learn from it?

2.4 Scientific and practical relevance

To the best of the author's knowledge, no CA has been carried out at a technical study program. Findings from the current study adds to the existing theories on whether CA will be applicable to a technical study. In the absence of similar project, this study can be considered as a base project for implementing CA in a technical study program. It could serve as a benchmark for the further development of a validated formative CA.

Additionally, this study provides the faculty of ME insight into the development process of implementing CA. First, data was collected on the basic learning goals as defined by teachers. Currently, the learning goals were defined for a whole course. These overall learning goals needed to be reduced to only the critical basic knowledge students should at least possess at the end of a course and even after some time. Basic learning goals were not drafted by teachers yet, so these were an added value for the faculty. Second, the CA tool provides valuable in-depth information on students' current level of retention and how to assess this. Third, the formative way of CA show how to provide students with feedback so they can learn from it. These aspects lead to additional insights for future implementation of CA in the curriculum.

3 Research design and methodology: developing formative cumulative assessments, data collection and analysing it

This chapter describes the research design and methodology used in this study starting with the research design, followed by a description of the respondents and the sampling method that was used. The instrumentation will be described afterwards and a description of the procedure follows. Finally, the data analysis will be discussed.

3.1 Research design

This study is a design study that develops a formative CA tool to measure students' knowledge retention level. The study can be identified as a design study, since it analysed a specific problem identified by the faculty of ME, designed a system that supports solving this problem, validated the design, and reflected about the lesson learned in order to refine the design guidelines (Sedlmair, Meyer, & Munzner, 2012). The design study is the appropriate research design since it guided this study towards its main question aiming to develop a formative CA tool.

This study used a mixed method type of research to answer the main question and the three sub questions. The first sub question aimed at clarifying the basic learning materials. It was answered through an open question asked to ME teachers. This was a nonexperimental qualitative method of data collection relying on teacher responses to the specific question. This type is used to gain understanding of what information the teachers had on the basic learning goals (Boudah, 2010).

The second sub question aims at identifying the reliability of the items of the test used in the CA. This is done by conducting the CAs that identified the knowledge students possessed, which is a quantitative method. This type of research was descriptive and used to provide a broad description of the condition of students' knowledge retention level. This was done without manipulating the students and will not determine any causes or effects (Boudah, 2010). This question is considered as nonexperimental, since there was no treatment. However, remaining observations were done on the extent students possess the basic learning materials as identified in the first research question. Hence, two moments were chosen to assess the students and to see if they have better test results the second time. The test results of both CAs were analysed.

The third research question aimed at providing feedback to students. The feedback was developed by the teachers who also developed questions for the CA. The teachers were asked through an open question to provide elaborations of the MC questions they developed. This is a non-experimental qualitative method of data collection, since it relied on teachers' responses. To get insight in the perceived use of feedback, questions were developed by the researcher to give understanding of the way the students perceived the feedback. This is a quantitative descriptive type of research, since it got the opinion from the students through MC questions. Furthermore, the grades of the students were analysed of all CAs. Comparison was done to check if there was a difference between the two moments of testing among the academic years.

This pilot study is instrumental of nature because it aimed to develop a CA tool (Boudah, 2010). This study develops the CA, documented the responses, analysed the data and described the nature of students based on the responses. The descriptive research result provides a check as a basis for further experimentation to determine possible cause and effects.

3.2 Respondents and sampling

For the qualitative data collection of this study, which is the identification of the knowledge that students should possess at the end of specific ME courses, criterion sampling was chosen. The goal of this qualitative research was to obtain insights into educational practices according to the ME teachers. Therefore, criterion sampling was used to select the teachers. The selection was based on the courses included in the research (Onwuegbuzie & Leech, 2007). This is an appropriate approach, since these teachers are the content experts and were therefore selected to develop questions and elaborations. In total, 21 teachers participated in this study.

For the quantitative data gathering of this study, non-probability sampling was used and does not rely on the use of randomization techniques (Bernard & Bernard, 2012). Two types of this non-probability sampling were used, namely criterion and convenience sampling. First, the criterion sampling method was chosen to select the students that meet criteria including: the student must be a bachelor ME students and aged 18 or older (Onwuegbuzie & Leech, 2007). Second, the convenience sampling method was used for the qualitative part of this study. The study involved selecting individuals that happen to be available and were willing to participate one or two of the scheduled meetings (Onwuegbuzie & Leech, 2007). In total, 269 students participated in the primary CA and 95 in the secondary CA. Prior to proceeding this study, the Ethical Committee of the UT was asked for approval of this study and this was granted.

3.3 Instrumentation

To answer the first research question regarding the basic knowledge that ME students should possess at the end of a course, this study used open questions asked to ME teachers. An open question was useful in this situation, since it gave the teachers the opportunity to think and reflect to give the required information on basic learning goals (Boudah, 2010).

To answer the second research question that assesses the reliability of the developed CA, at least four MC questions per course were developed by the selected course teachers. The test results of the students were used for the analysis of these items. For each academic year, two

CAs were developed. One assessed at the end of third quartile and one at the end of the fourth quartile. Since there are three academic years in the bachelor, six CAs were developed in total. The CAs consisted of MC questions and is an appropriate method, which are suitable for a large number of participating students.

Finally, the teachers who developed questions for the CAs were asked to develop elaborations on the questions. These elaborations included the correct answer, explanation about why the alternative answers were wrong and its relation to the identified basic learning materials. To answer the third sub research question regarding suitable feedback, students were asked to fill in a short survey consisting of three MC questions that was added to the secondary CA. Students were asked to analyse the feedback developed by the teachers on the CA received by mail. The results of the CAs were used to compare the grades of the students.

3.4 Procedure

3.4.1 Identifying the knowledge students should possess

The current curriculum of ME is divided in three years, each year in four quartiles (which are also called modules) and each module in three to four courses (see Appendix 1 for an overview of the curriculum 2016-2017). After having received approval of the Ethical Committee of the UT, the data collection started by reflecting the curriculum of ME by the program director to identify courses that needed to be included in the CAs. The final curriculum existed of 21 courses (see Appendix 2). There were no courses in Module 9, 10 and 12. Module 9 and 10 were used for students to choose a minor resulting in different materials for each student and these modules were excluded for testing. Module 12 was used to complete a bachelor assignment and this module was excluded for testing too.

In February 2017, a mail was sent to the teachers related to the 21 remaining courses. The mail asked for basic learning goals and explained the difference between the overall learning goals and basic learning goals. In total, basic learning goals of 20 courses were identified and are displayed in Appendix 3. The number of learning goals per course varied between two and four. This provided in-depth information about the knowledge that students should possess at the end of the courses, which helps answering sub research question 1.

3.4.2 Assessing the reliability of items of cumulative assessment

In March 2017, a follow-up mail was sent to the teachers asking for MC questions measuring the learning goals they identified. Only four teachers responded after two weeks and several reminders were sent. Six teachers did not response to the request of sending questions prior to the primary CA. Six modules missed one course (2, 3, 4, 5, 6 and 7) and these courses were not measured in the primary CA. By March 2017, the data base existed of questions measuring 14 courses. Each teacher developed two or more MC questions, but these did not necessarily measure all learning goals. However, each question was linked to at least one learning goal.

Primary CA

An exam with nine questions randomly chosen from the item bank was tested with three students who finished the modules that were included in the CA. The main goal was to identify the time it took them to complete the questions. Table 2 shows that the average time per question varied between 0.40 and 3.30 minutes. The total average time for the three students to complete the CA was about 14 minutes. The students answered respectively four, five and two questions correct. Students provided feedback on the questions, for example that one question was on two pages and that they needed to flip the page. Their feedback was considered and changes were made to questions.

Table 2

Results try-out exam (n = 3)

Question number		*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*Total*
R1	Time	1.35	1.02	2.06	4.36	0.33	0.57	4.00	1.49	1.05	16.23
	Points	0	0	1	1	0	0	1	0	1	4
R2	Time	1.01	1.07	1.38	3.29	0.32	0.44	4.19	2.52	0.35	14.57
	Points	1	0	0	0	1	1	1	0	1	5
R3	Time	1.28	1.47	1.52	2.24	0.56	0.52	1.21	1.21	1.07	11.08
	Points	0	0	0	0	1	0	0	1	0	2
	Average time	*1.21*	*1.19*	*1.65*	*3.30*	*0.40*	*0.51*	*3.13*	*1.74*	*0.82*	*13.96*

The try-out exam showed that the average time was about two minutes per question. The time scheduled for the final CA was ninety minutes. Thus, the maximum number of questions was 45. Per course, two questions were included in the CA and for one course three, since these three covered all the learning goals of that specific course. A concept version of the CA was

drafted for the first, second and third year students. The director and the educational assistant of ME carefully read through the CAs and provided feedback. Their feedback was considered and improvements were made. The final CAs for the first, second and third year students consisted of respectively 15, 22 and 28 questions. An overview of the courses that were included in the primary CA can be found in Appendix 2. The overview of the questions that were included in the CAs can be found in Appendix 3, an overview of the use of the questions in both primary and secondary CA can be found in Appendix 4 and finally, Appendix 5 shows the overview of all questions in the item bank.

Secondary CA

After the primary CA, all teachers were contacted again to ask for two new questions for the secondary CA. Only seven teachers sent new questions and thus 40 percent of the new questions were developed by the teachers. The remaining 60 percent of the questions were developed by the programme director and a master student. The overview of these percentages can be found in Table 3. The programme director checked all the questions when finished. By June 2017, the data base existed of questions of two more courses than the primary CA, since two teachers responded after the primary CA and were thus too late to be included in that CA. Thus, 16 courses were included in the primary CA. An overview of the courses that were included in the secondary CA can be found in Appendix 2. The overview of the questions that were included in the CA can be found in Appendix 3.

Table 3

Overview of question and elaboration development and who developed it

Done by	*Primary CA*				*Secondary CA*			
	Question	*Part (%)*	*Elaboration*	*Part (%)*	*Question*	*Part (%)*	*Elaboration*	*Part (%)*
Teacher	15	100.0	12	80.0	7	41	6.5	38.2
Director	0	0.0	1	6.7	4.5	26	1.5	8.8
Student	0	0.0	2	13.3	5.5	32	9	53.0

Assessing the primary and secondary CAs

After the development of the CAs, data collection was done through assessing the CA. The students were all invited and at the beginning of both CAs, the researcher started with telling

the students that their participation is voluntarily and that they can withdraw from the CA at all times. Moreover, they were told that their test results will be treated with confidentially and anonymity. Only the researcher has insight in the individual test results of the CA.

The CAs were conducted at the University of Twente, between or after lecture hours. The CAs were planned while considering the normal college hours and the colleges the students had. The maximum duration of one CA was 90 minutes, equal to a normal lecture duration. The researcher was present during the CAs for questions and several teachers were present to guide the CA and answer possible questions.

3.4.3 Provide students with feedback

The teachers were asked to develop elaborations of the questions they developed, so that students could learn from the CA. Several reminders were sent to the teachers, but still not all teachers sent elaborations. In total, 80 percent of the teachers sent elaborations for the primary CA. The remaining elaborations were made by the director and a master student. Only 40 percent of the teachers sent elaborations for the secondary CA, where the remaining elaborations were developed by the director and for more than half by a master ME student (Table 3 in the previous section shows the overview of these percentages)

To send the mails, an e-mail address was created by the ICT department of the UT. Mail merges were developed by the researcher and the mails were tried to send. Each mail merge included a test mail that the researcher should receive to check if the mails were send. The test mail was received from the first and third year students mail merge, but not from the second-year mail merge. Since the ICT department could not identify or solve the problem, the mail merge of the second-year students was successfully send via a private mail account. During the secondary CA, students from the first and third year complained that they did not receive the test results. After a check, it appeared that these mail merges were not send by the special created e-mail address. These mails were send the next day, but it was too late for the students to use the test results to learn from it prior to the secondary CA.

After each CA, feedback was given to the students about their test results using a three-level feedback model. First, it gives feedback to each student on what questions were answered correctly and incorrectly. Next, an overview of the learning goals per question was provided so that students could check what basic learning goals are part of what course. Finally, the corresponding elaboration per question was given so that students could understand why the

answers are wrong or right. The elaborations and the CAs were uploaded on Blackboard so that students can check the CA and the elaborations.

Additional questions were added to the secondary CAs to gather the students' feelings and thoughts on the CAs to answer the third sub research question. It asked if the learner read the elaborations send by mail and if they checked which questions they answered correctly and incorrectly. This was done by dichotomous questions so that it could easily be analysed. These questions will help answering sub research question 3. Furthermore, analysis was done on the grades of the students. Overall comparison was done between the primary and secondary CAs of year 1, 2 and 3 to identify differences between the average results of the students.

3.5 Data analysis

3.5.1 Identifying the knowledge students should possess

First, the data gathered by the open questions send to the teachers was documented. A list with basic learning goals was developed including the courses that are selected for the CA. The list of basic learning goals was sorted based on the year the courses were included in the curriculum. This documentary was necessary to answer the first sub research question.

3.5.2 Assessing the reliability of items of cumulative assessment

In order to analyse the quality of the items used for the CAs, Cronbach's alpha was calculated to test the internal consistency of the items in the CA and thus the reliability of each CA. For this "ConTest" was used, a statistical program developed at the University of Twente. The CAs were further analyzed by checking the distinctiveness of an item, which is necessary to determine the appropriateness of the CA (Ebel, 1972). The Item rest correlation (Rir-value) and the Item total correlation (Rit-value) were used to do so. The Rit-value determines how the item makes a difference between students with a higher grade and students with a lower grade. The value defines the distinctiveness of an item. Furthermore, the value shows the extent to which the item separately measures the test against the test as whole. The Rir-value is more accurate than the Rit-value, because it does not take its own value into consideration. In summary, the Rir-value identifies how well the item fits in the CA. Literature defines a value of 0.19 and lower as bad, between 0.20 and 0.29 as doubtable, between 0.30 and 0.39 as good and 0.40 and higher as very good (Ebel, 1972). However, the Rir-value should be analyzed in the context of the *p*'-value to get a view of the items. The *p'*-value is the transformed *p*-value with taking the

guessing change into consideration, where the *p*-value determines the probability of answering a question correct without a guessing chance.
Since the CA exists of MC items, the guessing chance needed to be taken into consideration. The *p'*-value was calculated according to the following equation:

$$p' = 1\frac{1}{3}p - \frac{1}{3}$$

Thus, the transformed *p*-value results in a *p'*-value varying between $-\frac{1}{3}$ to 1, where the average is at $\frac{1}{3}$. A *p'*-value lower than $\frac{1}{3}$ means that the item is difficult, since less than the average of $\frac{1}{3}$ of the students answered the item correct. Thus, only a small part of the students answered the item correct. When the *p*'-value is higher than $\frac{1}{3}$, the item is too easy, since more than the average of $\frac{1}{3}$ of the students answered the item correct. Analysis of the *p'*-value was done in comparison with the Rir-value to analyze the items used for the CA. This analysis was used to answer the second sub research question.

3.5.3 Provide students with feedback

Finally, in order to answer the third sub research question the additional questions that were included in the secondary CA were analysed. These additional questions asked the students if they used the feedback they received and if they learned from it. Descriptive analysis was done to describe and summarize the answers of the students. The question regarding the use of feedback could have been answered as follows:

a) I skimmed the elaborations and I did not check how I answered to the items at the previous cumulative assessment
b) I skimmed the elaborations, while I checked how I answered to the items at the previous cumulative assessment
c) I read the elaborations thoroughly and I did not check how I answered to the items at the previous cumulative assessment
d) I read the elaborations thoroughly, while I checked how I answered to the items at the previous cumulative assessment
e) I did not read the elaborations

Overall comparison was done between the students in the different academic years and between the primary and secondary CAs. The differences between the groups were calculated by the independent *t*-test that is used in situations in which there are two conditions and different participants (Field, 2013).

4 Results

This chapter provides the results of the study. First, the results related to gathering data for the first sub research questions are discussed. This is followed by the results obtained to answer the second sub research questions and finally, the results for the third sub research questions are discussed.

4.1 Identifying the knowledge students should possess

The basic learning goals were identified by an open question send to the teachers. Out of 21 teachers, 20 responded and send their basic learning goals. These goals were all mapped and an overview was made. These basic learning goals per course can be found in Appendix 3.

4.2 Assessing the reliability of items of cumulative assessment

4.2.1 Reliability check of the assessments

Six CAs were developed to measure the extent students' master the required basic knowledge. To validate the CA, the reliability of the scale was calculated. Reliability means that the CA should consistently reflect the construct that it is measuring, in this case the basic learning materials (Field, 2013). The idea behind this is that individual items or a set of items should produce consistent results with the overall assessment. This was done by calculating the Cronbach's alpha, which is a common way to measure the scale reliability (Field, 2013). A Cronbach's alpha with a value between .7 to .8 is an acceptable value according to Field (2013), lower values indicate an unreliable scale.

The Cronbach's alpha for the six CAs were calculated and an overview is displayed in Table 4. It shows no acceptable value for any of the CAs. The lowest value is a negative value, which is because the magnitude of the negative covariance is bigger than the magnitude of the positive ones in this extreme case (Field, 2013). These unacceptable values show that the questions cannot be reduced to one score on an underlying dimension, which was supposed to

be the basic learning materials. It might imply that the CAs include too many dimensions that cannot be measured properly with the questions that were used, since the internal consistency is too low.

Table 4

Cronbach's alpha of the six CAs and its required replication to get an acceptable value

Year	*CA number*	*Chronbach's alpha*	*Required replications to get* $\alpha = 0.7$
1	1	.08	28.27
	2	-.18	*
2	1	.13	15.03
	2	.06	35.45
3	1	.09	24.36
	2	.15	12.79

4.2.2 Item analysis of primary assessment

The primary CA developed for students in the bachelor of ME at the UT. Since the bachelor includes three academic years, three CAs were developed for the primary CA. These CAs were conducted at the end of the third module of the academic year. This section elaborates on the item analysis of the items developed for this CA.

The item distinctiveness is generated by a comparison between the Rir- and *p'*-value. The *p'*-value differs between $-\frac{1}{3}$ and 1. The negative *p'*-values were found, because the *p'*-value is calculated with the chance score taken into consideration. Where the *p*-value probability lies between 0 and 1, the *p'*-value lies between $-\frac{1}{3}$ and 1 (see for further explanation Chapter 3.5.2).

CA for academic year 1

The CA consisted of 15 questions and after a chance score taken into consideration, the students should have least get 10 points to pass the CA. In total, 93 students participated to the first CA, whereof 9 (9.68%) passed the CA with a 5.5 or higher and 85 (90.32%) failed the CA.

Table 5 shows the statistical details of the results. The students' grades varied between a 1.00 and a 6.80, with an average of 3.62. The table shows the *p*-values of the results. A low *p*-value means that a question is hard and a higher *p*-value means that a question might be too easy. Too easy and too hard questions are undesirable in a CA and therefore a *p*-value around

.3 would be optimal and the corrected *p*-value of this CA is about .29 which is close to an optimal value.

Table 5

Statistical details of primary CA in year 1

	Statistical details			
	Average	*St. dev.*	*Lowest*	*Highest*
Points	6.98	1.85	1.00	11.00
P-value	.47	.19	.11	.84
Corrected P-value	.29	.26	-.19	.78
Item-rest correlation	.03	.06	-.06	.15
Grade	3.62	1.39	1.00	6.80

The CA can be further analyzed by checking the distinctiveness of an item. Figure A6.1 in Appendix 6 shows the analysis of the Rir-value with the *p*'-value. It shows that question 4, 7 and 8 have an exceptional low p*'*-value combined with a low Rir-value, which is identified as a wrong key or another alternative may be plausible. For example, Table A6.1 shows the entire CA analysis and shows that only 18 students chose D which was the correct answer at question 4, where 52 chose option C. It seems like the alternative is more plausible, which can be seen in Table A6.2.

CA for academic year 2

The second-year CA existed of 22 questions and 144 students voluntarily participated. Of the total number of students, 26 (22.81%) students had a grade of a 5.5 or higher and 88 (77.19% failed). In total 14 points were necessary to get at least a 5.5, when taking the chance score into consideration.

The statistical details of the CA can be found in Table 6. The students got on average 11.60 points which is equal to a grade of 4.25. The grades varied between 1.70 and 7.20. The table also shows the corrected p-value at .37, which is close to the optimal value. This means that the average questions were not too hard or not too easy.

Table 6

Statistical details of primary CA in year 2

	Statistical details			
	Average	*St. dev.*	*Lowest*	*Highest*
Points	11.60	2.33	7.00	17.00
P-value	.53	.18	.24	.86
Corrected P-value	.37	.24	-.02	.81
Item-rest correlation	.03	.11	-.17	.24
Grade	4.25	1.29	1.70	7.20

Further analysis can be found Appendix 7, showing the distinctiveness of the items. It first shows a figure that displayed the analysis of the Rir-value against the p'-value, with a table displaying the item analysis report of the entire CA showing the number of students that choose a certain answer. Finally, the conclusions per questions are displayed in a table with the corresponding Rir- and p'-value per item.

Figure A7.1 in Appendix 7 shows the analysis of the Rir-value with the *p*'-value. It shows that question 4, 7 and 8 might have an exceptional low p'-value combined with a low Rir-value. For example, Table A7.1 shows the entire CA analysis and shows that only 18 students chose D which was the correct answer at question 4, where 52 chose option C. It seems like the alternative is more plausible, which can be seen in Table A7.2.

CA for academic year 3

The general results of the CA that exist of 28 items was conducted among students in their third year. In total, 62 students participated and 4 (6.25%) of them passed the CA, against 58 (93.55%) failures. After taking the chance score of 7.75 into consideration, the students needed to have at least 18 points to pass the CA.

The statistical details of the CA can be found in Table 7. The average number of points was 13.42, equal to a grade of 3.68. The number of points varied between 1.30 and 6.10. The table shows a *p'*-value of .31, which is close to the acceptable value.

Table 7

Statistical details of primary CA in year 3

	Statistical details			
	Average	*St. dev.*	*Lowest*	*Highest*
Points	13.42	2.46	8.00	19.00
P-value	.48	.23	.11	.90
Corrected P-value	.31	.30	-.18	.87
Item-rest correlation	.02	.13	-.17	.38
Grade	3.68	1.07	1.30	6.10

The item analysis report is displayed in Appendix 8. It shows how the items relate to the rest of the CA (see Figure A8.1). For example, several items seem to be too hard since they have a negative Rir-value in combination with a negative *p'*-value like question 20, 24 and 28. Table A8.*1* shows that only ten students chose the correct answer at item 20 and 44 chose the alternative D. In total, four questions seem to be too easy, among item 23 that was answered correct by 52 students (see Table A8.2).

Overall Rir- and *p'*-value analysis

The developed CAs were further analyzed by checking the distinctiveness of all items. The Rir-values were compared to the *p'*-values. An item was scored to be *too easy, too hard, another alternative may be plausible, wrong key or another alternative may be plausible* or *OK*. The score depends on the level of the *p'*-value combined with the Rir-value. An overview was made to compare the Rir- and *p'*-value analysis of the CAs and can be seen in Table 8.

The table shows that there were few questions that were rated as *OK* in the first year, meaning that these items score good on distinctiveness. All remaining items score thus lower on distinctiveness and were not a good item for the CAs. In the second and third year, more questions were identified as *OK*. Almost half of the questions were identified as *OK* in the second CA of the second year. Many questions were identified as *too easy* or that *another alternative was possible*. Few questions were identified as *too difficult*.

Table 8

Overview of Rir- and p'-value analysis of the six developed CAs

Year	*CA number*	*Total items*	*Analysis result*				
			OK	*Too easy*	*Too difficult*	*Another alternative*	*Wrong key or alternative plausible*
1	1	15	1	2	0	9	3
	2	18	5	4	1	6	2
2	1	22	6	4	0	10	2
	2	30	13	4	0	9	4
3	1	28	8	4	0	9	7
	2	32	11	2	3	7	9

4.2.3 Item analysis of secondary cumulative assessment

This section shows the results of the secondary CA that was conducted in the fourth module of the academic year. For this assessment, three CAs were developed. The item analysis of these CA can be found in this section.

Assessment for academic year 1

In total, 40 students voluntarily participated in the secondary CA. In total, 8 (20%) students passed the CA and 32 (80%) failed. The CA included 18 questions and after taking the chance score into consideration, 12 items needed to be answered correctly.

The statistical details of the secondary CA can be found in Table 9. It shows the average mark of the students, which is at 4.62 and varies between 2.00 and 7.30. Furthermore, it shows a corrected p-value of .40, which is slightly higher than the required value of .30. The number shows that the average questions are towards being too easy instead of too difficult.

Table 9

Statistical details of secondary CA in year 1

	Statistical details			
	Average	*St. dev.*	*Lowest*	*Highest*
Points	9.93	1.75	6.00	14.00
P-value	0.55	0.22	0.23	0.98
Corrected P-value	0.40	0.29	-0.03	0.97
Item-rest correlation	-0.02	0.16	-0.24	0.28
Grade	4.62	1.17	2.00	7.30

The item analysis report is displayed in Appendix 9. Figure A9.1 shows how the items relate to the rest of the CA. For example, question 7 seems to be too hard, since it has a negative Rir-value in combination with a negative *p'*-value, which exact data can be found in Table A9.1 and Table A9.2. In this case, question 5 seems to be too easy, since 31 students chose the correct answer. In total, three other questions seem to be too easy. Further results can be found in the appendix.

CA for academic year 2

In total, 47 second-year students participated in the secondary CA, whereof 6 (12.77%) passed and 41 (87.23%) failed. Students needed at least 19 points to pass the CA that included 30 questions.

In Table 10 is shown the statistical details of the test results. It shows the average number of 15.53 points, which is graded with a 4.21. The grades varied between 2.40 and 6.40. The p'-value shows an average of .36, which is close to the acceptable value.

Table 10

Statistical details of secondary CA in year 2

	Statistical details			
	Average	*St. dev.*	*Lowest*	*Highest*
Points	15.53	2.46	11.00	21.00
P-value	.52	.25	.15	.98
Corrected P-value	.36	.33	-.13	.97
Item-rest correlation	.02	.16	-.32	.26
Grade	4.21	.98	2.40	6.40

The item analysis report is displayed in Appendix 10. Figure A10.1 shows how the items relate to the rest of the CA. For example, question 7 seems to be too hard, since it has a negative Rir-value in combination with a negative *p'*-value, which exact data can be found in Table A10.2 and Table A10.1. Question 5 seems to be too easy, since 31 students chose the correct answer. In total, three other questions seem to be too easy. Further results can be found in the appendix.

CA for academic year 3

In total, 8 students participated and none of them passed the CA that existed of 32 questions. Table 11 shows the statistical details of the CA. The average grade is a 2.48 and the grades varied between 1.00 and 4.00. The *p'*-value of 0.16 is lower than the acceptable value, indicating that the items were too easy for this test.

Table 11

Statistical details of secondary CA in year 3

	Statistical details			
	Average	*St. dev.*	*Lowest*	*Highest*
Points	11.88	2.67	8.00	16.00
P-value	.37	.21	.00	.88
Corrected P-value	.16	.28	-.33	.83
Item-rest correlation	.05	.37	-.92	.84
Grade	2.48	.99	1.00	4.00

The item analysis report is displayed in Appendix 11. It shows how the items relate to the rest of the CA (see Figure A11.1). In total, four question seem to be too easy, as displayed in Table A11.1 and Table A11.2. Some of the questions have both a low Rir-value and a low p'-value. Further results can be found in the appendix.

4.3 Provide students with feedback

4.3.1 Use of feedback and elaborations

CA for academic year 1

Table 12 shows the results of the additional questions answered by the first-year students. It shows that only one student prepared for the CA (item 1) and that most the students did not do anything with the elaborations that they received after the first CA (*n = 23)*. In total, 12 students read the elaborations and 4 students said that they changed their study behavior after the primary CA, where 36 did not.

Table 12

Additional questions second CA in year 1

		A	*B*	*C*	*D*	*E*
Item	*Open*	*f*	*f*	*f*	*f*	*f*
1	0	1	39	-	-	-
2	5	12	0	0	0	23
3	0	4	36	-	-	-

CA for academic year 2

Table 13 shows the result of the additional questions. Only one student said that he or she prepared for the CA. There were 24 students who skimmed the elaborations, whereof 8 did not use their test results and 16 did. In total, 5 students read the elaboration thorough, whereof 1 did not check the test results and 4 did. So, in total 9 did not check the test results while reading the elaborations and 17 did not read the elaborations at all.

Table 13

Additional questions second CA in year 2

		A	*B*	*C*	*D*	*E*
Item	*Open*	*f*	*f*	*f*	*f*	*f*
1	1	1	45	-	-	-
2	5	8	16	1	4	17
3	1	2	44	-	-	-

CA for academic year 3

In Table 14 is shown what the results of the additional questions are. In total, seven students said that they did not prepare for the CA. Two students said they skimmed the elaborations, without checking the test results. Four said that they did not read it at all. In total, four students said that they did not change their study behaviour and four did not answer the question.

Table 14

Additional questions second CA in year 3

		A	*B*	*C*	*D*	*E*
Item	*Open*	*f*	*f*	*f*	*f*	*f*
1	1	0	7	-	-	-
2	2	2	0	0	0	4
3	4	0	4	-	-	-

4.3.2 Knowledge retention level of students

The developed CAs measured the knowledge retention level of the students twice. The first time at the end of the third module, the second time at the end of the fourth module. The students were not aware of this CA when they were asked to make the CA, while they already had elaborations and awareness of the CA when they were asked to make the secondary CA. Therefore, the knowledge retention level of the students was analyzed per CA, but also in comparison to each other to establish any significant differences.

Overall comparison

First, Figure 1 displays the average grades of the six CAs differentiated per academic year. The students in year 1 had on average a higher mark on their secondary CA than on the primary CA, where the students in year 2 and 3 had on average a higher grade in their primary CA than in their secondary CA. The next section will elaborate on the differences between these years and if there is any significant difference found.

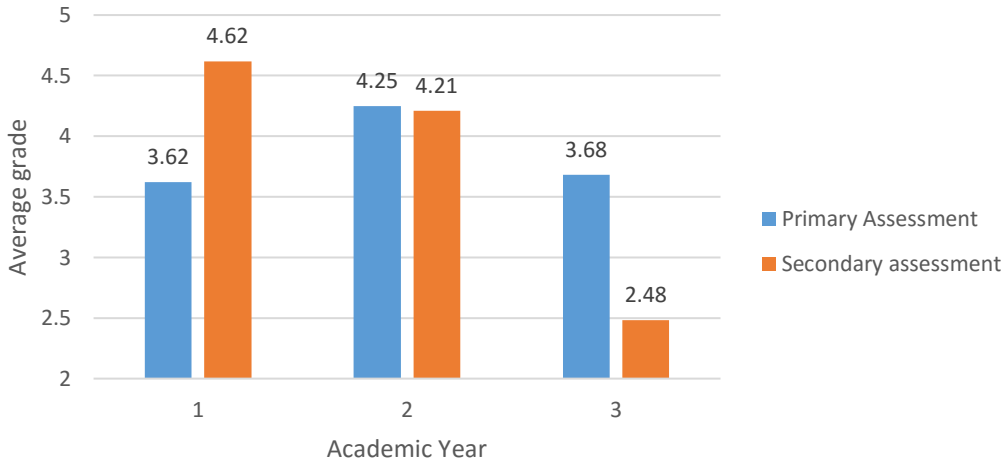


Figure 1 Comparison between primary and secondary CA of year 1, 2 and 3 students

Primary and secondary CA: year 1

Comparison was done to check if there is a difference between the two moments of testing among the first-year students. The independent *t*-test is used in this situation in which there are two conditions and different participants (Field, 2013). The output of the analysis showed that on average, students had a higher grade on the secondary CA (M = 4.62, SE = .14) than on the primary CA (M = 4.62, SE = .19). This difference was not significant t (-4.10) = .49, $p > .05$.

Primary and secondary CA: year 2

The outputs of the analysis showed that on average, students had on average a higher grade on the primary CA (M = 4.21, SE = 0.12) than on the secondary CA (M = 4.21, SE = 0.14). This difference was significant t (110.96) = .18, $p < .05$.

Primary and secondary CA: year 3

The outputs of the analysis showed that the students on the primary CA had on average a higher grade (M = 3.68, SE = .14) than students on the secondary CA (M = 2.48, SE = .37). This difference was not significant t (68) = 3.03, $p > .05$.

Secondary CA comparison between year 1 and 2

A comparison was made of the secondary test results between year 1 and year 2, because students in year 1 did not get test results where year 2 did get their test results. Both groups got elaborations on the questions. The results showed that the students in year 1 had on average a higher grade (M = 4.62, SE = .19) than students in year 2 (M = 4.21, SE = .15). This difference was not significant t (85) = 1.73, $p >.05$.

Secondary CA comparison between year 2 and 3

A comparison was also made for the difference in the secondary CA between year 2 and 3. The output showed that students in year 2 had on average a higher grade (M = 4.21, SE = .15) than students in year 3 (M = 2.48, SE = .37). The difference is not significant t (9.23) = 4.34, $p > .05$.

5 Conclusions, discussions, limitations and recommendations for further research

This section starts with answering the formulated research questions. This is followed by the limitations of this study. Finally, the recommendations for further research are given.

5.1 Conclusions

5.1.1 Research question 1: Which knowledge should students possess at the end of specific bachelor Mechanical Engineering courses?

This section answers the first research question: "Which knowledge should students possess at the end of specific bachelor Mechanical Engineering courses?"

The course teachers were contacted and asked to differentiate between overall learning goals and basic learning goals. The overall learning goals were already identified for each course and only a selection was needed. Out of 21 teachers, 20 responded to the request and send basic learning goals. These goals were all translated so they started with "After the course, the student is able to …". It appeared that teachers were perceiving this not as hard or time consuming, since they responded quickly. The documentation of all basic learning goals can be found in Appendix 3.

5.1.2 Research question 2: What is the reliability of items in the test used in the formative cumulative assessment for bachelor Mechanical Engineering students?

This section answers the second research question: "What is the reliability of items in the test used in the formative cumulative assessment for bachelor Mechanical Engineering students?"

Scale reliability

For each of the CAs, the Cronbach's alpha was calculated. It showed no valuable value for either one of the CAs. The highest alpha was .15, while an acceptable value is between .7 and .8 (Field, 2013). The alpha for the secondary CA was lower for the first and the second year, where the alpha for the third year was higher. Thus, the reliability of the scales of the developed CAs was insufficient for all CAs. It indicates that that the set of items that were produced did not consistently reflect the construct that it was measuring: the basic learning materials.

Item analysis

Statistical analysis was done on all items the assess the developed items for the CA. First, the *p'*-values were identified to determine whether a question was too hard or too easy. A good *p'*-value should have been about $\frac{1}{3}$. The calculated values show that almost all *p'*-values were at that value, except the second CA for the third-year students. That value showed a lower *p'*-value, indicating that the items were too easy.

Item distinctiveness

The item distinctiveness was analyzed by comparing the Rir- and *p'*-values. It shows that hardly any items were scored with OK. It appeared that at most items another alternative was more likely to be true, or the item was found to be too easy. The results showed that the low scoring students had a high probability of answering the item correct, where it is normal that high scoring students have a higher probability of answering items correctly. This means that overall, the items did not accurately measure the basic learning goals in a correct manner.

5.1.3 Research question 3: How to provide Mechanical Engineering students with feedback in a manner that they can learn from it?

This section answers the third research question: "How to provide Mechanical Engineering students with feedback in a manner that they can learn from it?"

Formative CA

Students were provided with feedback about their test results and related elaborations of the CA items were included. That way, students could assess their state of development in relation to what is required at the end of each course. According to Black and Wiliam (1998), providing feedback that moves students forward is key strategy 3. However, due to a failure the feedback was only send to students in their second year.

Comparison between academic years in secondary CA

All students were provided with elaborations, but only the second-year students with their test results. First, looking at the use of elaborations among the first-year students, 41 percent read the elaborations and 59 percent did not. Among the second-year students, 57 percent read the elaborations and 43 percent did not. Moreover, about 70 percent of the students who read the

elaborations, read it while they were checking how they answered the previous CA items. Among the third-year students, 43 percent read the elaborations and 57 percent did not.

Conclusion

Even though that only the second-year students received their test results, it appeared that more than half read the elaborations. The percentage students that read the elaborations was the highest among the second-years. Moreover, 70 percent of them checked the elaborations while they were looking to the test results of the previous CA. These factors indicate that students want to read the elaborations and when they have the test results, they check how they answered the previous CA. Thus, the feedback analysis shows that students want to read elaborations and used their test results to read them.

Discussion about knowledge retention level

Looking at the overall test results, it was concluded that the students do not master the learning materials the teachers expects them to master. There can be several possible explanations.

First, the test and item analysis showed that the CAs did not meet the standard requirements for a sufficient test according to the COTAN guidelines (Evers, Lucassen, Meijer, & Sijtsma, 2009). COTAN is an assessment system that identified the quality of a test and includes for example the Rir-analysis as part of the analysis (Evers et al., 2009).

Second, since only two questions were used to measure one course it is impossible to accurately draw a conclusion on the extent to which a course is correctly or incorrectly measured. Adding more items would give a better picture of the dimension that should be measures.

Third, the extent to which student should possess basic learning goals can be discussed. As being one of the reasons for this study, teachers were complaining about students who were not recognizing previously learned materials which they were expected to still recognize. In the CAs, basic learning goals were measured to check the extent to students possess these basic learning goals of previously completed courses. Since the researcher was present at the CA, some of the students talked to the researcher after the CA. They told her that if they would have had their books, they would know what book to use to solve the question and they would be able to solve most of the questions. So, where the teachers complained about students who did not recognize course contents, these students did. Teachers should discuss whether students should know how to solve questions by heart or if recognizing is enough.

Fourth, it might be that the students did not take the CA seriously. During the CA, some students were discussing questions together or might have been just randomly guessing when did they not know the answer quickly. When there would be a consequence to this CA, student might have taken it more serious and would have tried harder to get the correct answer. That would have resulted in a higher grade and would have given a more accurate picture of their long-term knowledge retention.

A comparison of the test results of the students in the primary and secondary CA was done in these remaining observations. The comparison revealed no significant differences between primary and secondary test results of the first- and third-year students. However, a significant difference among the second-year students between the primary and secondary CA was found, where the grades were higher on the primary CA. The reason for this might be that students received their results prior to the secondary CA and could check the questions they answered (in)correctly, where the other students could not do this prior to their secondary CA. Another explanation of this significant difference is that most of the students who read the elaborations and test results, belonged to the second-years. However, these are only suggestions, since this study did not investigate any cause and effect.

5.2 Limitations

Limitations in this study were identified as characteristics that potentially impacted or influenced the interpretation of the findings of this study, since it is unclear to what degree the different factors limited the findings. One limitation is that the e-mail with the test results was not send to first and third year students after the first CA, due to an unfortunate failure of the e-mail program that was used. These failures are part of the limitations of this study. Despite these failures, additional analysis could be done to investigate differences between the test results of the second-year students in comparison to the first- and third-year students.

The low number of participating students of the secondary CA may also have influenced the findings. At the secondary CA, the number of first year reduced from 93 to 40, the second years from 144 to 47 and the third years from 62 to 8. It is unclear how this influenced the findings. However, it might be that only the motivated students voluntarily participated the second time. This hypothesis would stand at the first-year students, since the second CA average grades is higher than the primary. The hypothesis does not hold the second and third-year students, since both secondary test results are lower than the previous.

This study did not use a control group, because of ethical issues. Therefore, no conclusions can be drawn on the effects of CAs between the primary and the secondary CA. It is unclear which other aspects of CAs have influenced students' study behaviour. Further research is needed to unravel how each aspect of CAs influences the study behaviour of students and what their actual long-term knowledge level is.

A final limitation is that the researcher had no knowledge on ME materials. The checks on the content of the CAs and the elaborations was done by the director and a master student, where the researcher could not check the contents. The research was dependent on others for checking the items, elaborations and basic learning goals. The process might have been faster when the researcher was not dependent on others, but had knowledge on ME materials.

5.3 Recommendations

Full support of all teachers throughout the faculty is required

During the collection of all items, it appeared that implementing CAs was not supported by all the staff members throughout the faculty. Some send negative reactions per mail or did not reply at all. However, to be able to successfully implement CAs, support throughout the entire faculty is required. Furthermore, none of the staff developed basic learning materials before this study. They all could quickly draft them, but as they said, many of them never thought of it before. It might be even more helpful when teachers talk about the basic learning goals to make students aware of them.

Some of the questions and elaborations were not developed by the teachers, but by master students and the program director. The teachers are content experts and have more knowledge on the learning goals they developed, so it would have been better if they developed all the questions and recommendations.

Define basic learning goals throughout the ME curriculum

Teachers identified basic learning goals for their own courses, these were only documented. No analysis was done, since the researcher was not a content expert. It is recommended that the teachers discuss with each other what basic learning goals per module are, instead of only the courses. Since some courses relate, the learning goals should relate as well. The faculty should discuss what the exact knowledge is that students are required to possess at the end of a course, but also at the end of subsequent courses or even years.

The identification of basic learning goals is in coherence with the first strategy of Black and Wiliam (1998), being the first key strategy and showing its importance. Before a continuation of the development of reliable CAs, the basic learning goals should be identified throughout the entire ME curriculum. Some items were developed by master students or program director; no analysis was done to identify possible differences. This analysis was not executed because of time boundaries. Analysis can be done to identification the differences, which would give insight in the way the teachers developed the items in comparison to the students and program directed. This would not have an added value for the future development of new items, since it is recommended for the teachers to develop all the items.

Development of reliable MC items in item bank

It appeared that teachers had no or less experience with MC items. This was illustrated by the results of the CA analysis, where not one of the CAs had a sufficient reliability and most of the items scored insufficient on the item analysis. One of the teachers developed a 'trick question', where most students got to the correct answer, but they were tricked by another answer. This makes it questionable if the goal was to measure the basic learning goals or how to tackle trick questions. The development of good items is crucial prior to an actual implementation. Further research is recommended on the reliability of the items with a factor analysis to determine possible underlying dimensions. Moreover, statistics can help to determine the level of the questions and to identify sufficient items, but the teachers should go back to the drawing phase to develop new items. Developing CAs did not take a lot of time, but the collection of items and elaborations did. This was because teachers did not have enough time to develop them or they had other priorities. A question cannot be used in different CAs for the same students. Teachers should take time to develop multiple questions for their courses. Therefore, it is recommended to develop an item bank that includes reliable items prior to implementing CAs.

It might also be that MC questions is not the appropriate way to measure the basic learning goals of the curriculum, since ME is a technical study and which is the reason that teachers had no or less experience with developing MC questions. MC questions require much more attention than the development of an open question. Therefore, the faculty should discuss what type of questions is the best to use to assess the basic learning goals of students through CAs. This development is in line with key strategy 2 of Black and Wiliam (1998), stating that the teacher should be able to determine where the learner is right now and thus to elicit evidence of student understanding.

Consequences for the CA need to be considered

The theoretical elaboration on successes of CAs showed that students study depending on motivation driving towards the goal of being rewarded (Heckhausen, 1977). This CA has no goal, like study points, for the students and moreover, the participation was voluntarily. That might have been the cause for the lower number of participating students during the secondary CA and might have also influence the test results. CA combines theories to stimulate students to study regularly and to increase the students' test performances. However, without consequences this influence might not happen. Students might change the way they prepare for the CA and even the way they make the CA. If a consequence for the CA would be that if a sufficient grade would grand access to the next study year, more students would have participated and would have taken it more serious. So, an important boundary condition concerns the consequences for the CA.

Reflect on the way feedback was given

Research showed the added value of formative CA over summative CA. FA can be used not only to give judgement about progress, but identifies a gap and helps the learner forward (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006). This will help the learner to become owners of their own learning, which is in line with key 5 of Black and Wiliam (1998). The final recommendation is to reflect on the way the feedback was given, since the program used for sending individual test results did not work. Another program should be considered, but attention is recommended to pay on the execution of the formative part of CA.

References

Bekkink, M. O., Donders, R., van Muijen, G. N., & Ruiter, D. J. (2012). Challenging medical students with an interim assessment: a positive effect on formal examination score in a randomized controlled study. *Advances in Health Sciences Education, 17*(1), 27-37.

Berg, M. N. V. D., & Hofman, W. H. (2005). Student success in university education: A multi-measurement study of the impact of student and faculty factors on study progress. *Higher education, 50*(3), 413-446.

Bernard, H. R., & Bernard, H. R. (2012). *Social research methods: Qualitative and quantitative approaches*: Sage.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education: principles, policy & practice, 5*(1), 7-74.

Boudah, D. J. (2010). *Conducting educational research: Guide to completing a major project*: Sage.

Bruijns, V. (2014). Het effect van tussentijds toetsen op studierendement: een literatuurstudie. *Onderzoek van Onderwijs*(43), 15-20.

Bruinsma, M., & Jansen, E. P. (2009). When will I succeed in my first-year diploma? Survival analysis in Dutch higher education. *Higher Education Research & Development, 28*(1), 99-114.

Cohen-Schotanus, J. (1999). Student assessment and examination rules. *Medical Teacher, 21*(3), 318-321.

De Paola, M., & Scoppa, V. (2011). Frequency of examinations and student achievement in a randomized experiment. *Economics of Education Review, 30*(6), 1416-1429.

Ebel, R. L. (1972). Essentials of educational measurement.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2009). *COTAN beoordelingssysteem voor de kwaliteit van tests (geheel herziene versie)*.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*: Sage.

Freeman, A., Van Der Vleuten, C., Nouns, Z., & Ricketts, C. (2010). Progress testing internationally. *Medical Teacher, 32*(6), 451-455.

Garrison, C., & Ehringhaus, M. (2007). Formative and summative assessments in the classroom.

Gibbs, G. (2010). *Using assessment to support student learning*: Leeds Met Press.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Haugan, J., Lysebo, M., & Lauvas, P. (2017). Mandatory coursework assignments can be, and should be, eliminated! *European Journal of Engineering Education*, 1-14.

Heckhausen, H. (1977). Achievement motivation and its constructs: A cognitive model. *Motivation and emotion, 1*(4), 283-329.

Jansen, E. P. (2004). The influence of the curriculum organization on study progress in higher education. *Higher education, 47*(4), 411-435.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *science, 319*(5865), 966-968.

Kerdijk, W., Cohen-Schotanus, J., Mulder, B., Muntinghe, F. L., & Tio, R. A. (2015). Cumulative versus end-of-course assessment: effects on self-study time and test performance. *Medical education, 49*(7), 709-716.

Kling, N., McCorkle, D., Miller, C., & Reardon, J. (2005). The impact of testing frequency on student performance in a marketing course. *Journal of Education for Business, 81*(2), 67-72.

Mitra, N., Nagaraja, H., Ponnudurai, G., & Judson, J. (2009). The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1, multidisciplinary summative tests. *IeJSME, 3*(1), 2-7.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education, 31*(2), 199-218.

Onwuegbuzie, A. J., & Leech, N. L. (2007). A call for qualitative power analyses. *Quality & Quantity, 41*(1), 105-121.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in education: principles, policy & practice, 5*(1), 77-84.

Schmidt, H. G., Cohen-Schotanus, J., Van Der Molen, H. T., Splinter, T. A., Bulte, J., Holdrinet, R., & Van Rossum, H. J. (2010). Learning more by being taught less: a "time-for-self-study" theory explaining curricular effects on graduation rate and study duration. *Higher education, 60*(3), 287-300.

Schouwenburg, H. C., & Groenewoud, J. (2001). Study motivation under social temptation; effects of trait procrastination. *Personality and Individual Differences, 30*(2), 229-240.

Sedlmair, M., Meyer, M., & Munzner, T. (2012). Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics, 18*(12), 2431-2440.

Swanson, D., Holtzman, K., & Butler, A. (2010). Cumulative achievement testing: Progress testing in reverse. *Medical Teacher, 32*(6), 516-520.

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British journal of educational studies, 53*(4), 466-478.

Turner, J. R. (2005). The role of pilot studies in reducing risk on projects and programmes: Elsevier.

Van Eerde, W. (2000). Procrastination: Self-regulation in initiating aversive goals. *Applied Psychology, 49*(3), 372-389.

Van Eerde, W. (2003). A meta-analytically derived nomological network of procrastination. *Personality and Individual Differences, 35*(6), 1401-1418.

Van Teijlingen, E. R., & Hundley, V. (2001). The importance of pilot studies.

Vleuten, C. V. D., Verwijnen, G., & Wijnen, W. (1996). Fifteen years of experience with progress testing in a problem-based learning curriculum. *Medical Teacher, 18*(2), 103-109.

Wade, L., Harrison, C., Hollands, J., Mattick, K., Ricketts, C., & Wass, V. (2012). Student perceptions of the progress test in two settings and the implications for test deployment. *Advances in Health Sciences Education, 17*(4), 573-583. doi:10.1007/s10459-011-9334-z

Wrigley, W., Van Der Vleuten, C. P., Freeman, A., & Muijtjens, A. (2012). A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Medical Teacher, 34*(9), 683-697.

Appendix 1 – Overview curriculum Mechanical Engineering

The table below shows the Mechanical Engineering curriculum 2016-2017.

Table A1.1 Overview curriculum Mechanical Engineering 2016-2017

B1	Design and Production	100 %	15 EC	Energy and Materials	100 %	15 EC	Energy and Sustainability	100 %	15 EC	Design and Mechanics	100 %	15 EC
	Mathematics A + B1	26.7 %	4.0	Mathematics B2	20.0 %	3.0	Mathematics C1	20.0 %	3.0	Mathematics D1	20.0 %	3.0
	Statics & Modelling and Programming 1	20.0 %	3.0	Eng. Thermodynamics 1 & Mod. and Prog. 2	26.7 %	4.0	Eng. Thermodynamics 2 & Mod. and Prog. 3	20.0 %	3.0	Mechanics of Materials & Mod. and Prog. 4 (201500519)	30.0 %	4.5
	Production Systems 1	16.6 %	2.5	Material Science 1 (201400307)	20.0 %	3.0	Material Science 2 (201400308)	13.4 %	2.0			
	Technical Drawing	10.0 %	1.5	Proj. Analysis Energy Syst. & Ac. Skills 2 (≥ week 4)	26.7 %	4.0	Life Cycle Analysis	23.3 %	3.5	Machine Parts (2014002306)	26.7 %	4.0
	Proj. Design Machine & Ac. Skills 1	26.7 %	4.0	Proj. Design Machine & Ac. Skills 1 (week 1,2)	6.6 %	1.0	Proj. Design Energy Syst. & Ac. Skills 3	23.3 %	3.5	Proj. Design Construction & Ac. Skills 4	23.3 %	3.5
B2	**Dynamic Systems**	**100 %**	**15 EC**	**Product Design**	**100 %**	**15 EC**	**Fluid Mechanics & Heat Transfer**	**100 %**	**15 EC**	**Mechatronic Design**	**100 %**	**15 EC**
	Mathematics D2	13.3 %	2.0	Tribology (201600123)	13.4 %	2.0						
	Dynamics 1 (201400488)	23.3 %	3.5	Elasticity Theory (201600060)	13.4 %	2.0	Fluid Mechanics 1 (201500391)	23.3 %	3.5	Dynamics 2 (201500496)	30.0 %	4.5
	System Analysis (201400377)	26.7 %	4.0	Processes and Properties of Polymers (201600124)	20.0 %	3.0	Heat Transfer (201500390)	23.3 %	3.5	System and Control 1 (201500497)	26.7 %	4.0
	Proj. Precision Mechanisms & Ac. Skills 5	36.7 %	5.5	Proj. Product Design & Ac. Skills 6	53.2 %	8.0	Proj. Fluids Engineering & Ac. Skills 7	53.4 %	8.0	Proj. Mechatronics & Ac. Skills 8	43.4 %	6.5
B3	**Module 9**		**15 EC**	**Module 10**		**15 EC**	**Production Systems Engineering**	**100 %**	**15 EC**	**ME Bachelor Assignment**	**100 %**	**15 EC**
	Minor	100.0%	15.0	Minor	100.0%	15.0	Statistics	16.7 %	2.5	BSc assignment	80.0 %	12.0
							Intro. Finite Element Method (201400311)	23.3 %	3.5	Academic Research & Skills 2	20.0 %	3.0
							Academic Research & Skills 1	20.0 %	3.0			
							Proj. Production Systems Engineering	40.0 %	6.0			

Appendix 2 – Courses selected to be measured

The table below shows the courses in green that were selected to be included in the CAs.

Table A2.1 Selected courses for the CAs

B1	Ontwerp en Productie	100 %	15 EC	Energie en Materialen	100 %	15 EC	Energie en Duurzaamheid	100 %	15 EC	Ontwerp en Mechanica	100 %	15 EC
	Mathematics A + B1	26.7 %	4.0	Mathematics B2	20.0 %	3.0	Mathematics C1	20.0 %	3.0	Mathematics D1	20.0 %	3.0
	Statica & Modelleren en Programmeren 1	20.0 %	3.0	Tech. Thermodynamica 1 & Mod. en Prog. 2	26.7 %	4.0	Tech. Thermodynamica 2 & Mod. en Prog. 3	20.0 %	3.0	Sterkteleer & Mod. en Prog. 4 (201500519)	30.0 %	4.5
	Productiesystemen 1	16.6 %	2.5	Materiaalkunde 1 (201400307)	20.0 %	3.0	Materiaalkunde 2 (201400308)	13.4 %	2.0			
	Werktuigbouwkundig tekenen	10.0 %	1.5	Proj. Analyse Energiesyst. & Ac. Skills 2 (≥ week 4)	26.7 %	4.0	Ketenbeheer	23.3 %	3.5	Werktuigonderdelen (2014002306)	26.7 %	4.0
	Proj. Ontwerp Werktuig & Ac. Skills 1	26.7 %	4.0	Proj. Ontwerp Werktuig & Ac. Skills 1 (week 1,2)	6.6 %	1.0	Proj. Ontwerp Energiesyst. & Ac. Skills 3	23.3 %	3.5	Proj. Ontwerp Constructie & Ac. Skills 4	23.3 %	3.5
B2	**Dynamische Systemen**	**100 %**	**15 EC**	**Productontwerpen**	**100 %**	**15 EC**	**Fluid Mechanics & Heat Transfer**	**100 %**	**15 EC**	**Mechatronisch Ontwerpen**	**100 %**	**15 EC**
	Mathematics D2	13.3 %	2.0	Tribology (201600123)	13.4 %	2.0						
	Dynamica 1 (201400488)	23.3 %	3.5	Elasticity Theory (201600060)	13.4 %	2.0	Fluid Mechanics 1 (201500391)	23.3 %	3.5	Dynamica 2 (201500496)	30.0 %	4.5
	Systeemanalyse (201400377)	26.7 %	4.0	Processes and Properties of Polymers (201600124)	20.0 %	3.0	Heat Transfer (201500390)	23.3 %	3.5	Systeem en Regeltechniek 1 (201500497)	26.7 %	4.0
	Proj. Dynamische Systemen & Ac. Skills 5	36.7 %	5.5	Proj. Consumentenproducten & Ac. Skills 6	53.2 %	8.0	Proj. Fluids Engineering & Ac. Skills 7	53.4 %	8.0	Proj. Mechatronica & Ac. Skills 8	43.4 %	6.5
B3	**Module 9**		**15 EC**	**Module 10**		**15 EC**	**Production Systems Engineering**	**100 %**	**15 EC**	**ME Bachelor Assignment**	**100 %**	**15 EC**
	Minor	100.0%	15.0	Minor	100.0%	15.0	Statistics	16.7 %	2.5	BSc assignment	80.0 %	12.0
							Intro. Finite Element Method (201400311)	23.3 %	3.5	Academic Research & Skills 2	20.0 %	3.0
							Academic Research & Skills 1	20.0 %	3.0			
							Proj. Production Systems Engineering	40.0 %	6.0			

Appendix 3 – Basic learning goals of Mechanical Engineering bachelor curriculum

Module 1: Statica en Modelleren en Programmeren 1

De student is in staat om:

1. Een goede Vrije Lichaam Structuur (VLS) te maken.
2. Een evenwichtsvergelijking op te stellen.
3. De dwarskracht en buigend momentenlijn te tekenen.

Module 1: Productiesystemen 1

De student is in staat om:

1. Een overzicht te geven van de bestaande productieprocessen en de bijbehorende kenmerken.
2. Een passend productieproces te kiezen voor een gegeven (deel van een) product.

Module 1: Werktuigbouwkundig tekenen

De student is in staat om:

1. Een technische tekening en schets te maken en interpreteren.
2. De in de werktuigbouwkunde gebruikelijke tekenregels toe te passen.
3. Het 3D CAD-systeem SolidWorks te gebruiken om een productontwerp in werktekeningen om te zetten.

Module 2: Technische Thermodynamica 1

De student is in staat om:

1. Verschillende vormen van thermodynamische energieën te onderscheiden en (wiskundig) te beschrijven hoe deze in elkaar omgezet worden.
2. Het gebruiken van gassen en vloeistoffen bij verschillende drukken en temperaturen te verklaren en uit te leggen en fasediagrammen te maken en te interpreteren.
3. De hoofdwetten van de thermodynamica uit te leggen, deze toe te passen op thermodynamische systemen en de gevolgen ervan verklaren en interpreteren.
4. Een ingewikkeld (samengesteld) thermodynamisch systeem om arbeid en/of warmte/koude te produceren te herkennen en de configuratie en werking te verklaren.

Module 2: Materiaalkunde 1

De student is in staat om:

1. Relevante mechanische eigenschappen te bepalen uit een trekproef en een hardheidsmeting
2. Te verklaren hoe materialen falen en plastisch vervormen bij verschillende productie- en gebruiksomstandigheden
3. Te verklaren wat de invloed van warmtebehandelingen is op de materiaalstructuur en –eigenschappen

Module 3: Technische Thermodynamica 2

De student is in staat om:

1. Het verschil tussen pad-en toestandfuncties uit te leggen, te bepalen in welke categorie een grootheid of energie valt en uit te leggen wat de gevolgen daarvan zijn voor de grootheid.
2. De gebruikte thermodynamische formules af te leiden en de thermodynamische diagrammen en tabellen te verklaren met behulp van fundamentele wiskunde (partiële differentialen).

Module 3: Materiaalkunde 2

De student is in staat om:

1. De microstructuur van eenvoudige ijzerlegeringen en andere legeringen te voorspellen op basis van fasediagrammen en tijd-temperatuur-transformatiediagrammen.
2. Te verklaren hoe faseovergangen zoals stollen, precipitatie en martensietvorming verlopen.

Module 3: Ketenbeheer

De student is in staat om:

1. Domein specifieke referentiekader voor levenscyclusanalyses te begrijpen en toe te passen.
2. De elementaire transformaties om van product karakteristieken te komen tot impacts op midpoint effecten.

3. De volgende hoofdbegrippen te kennen: functionele eenheden, allocatie berekeningen, en profiling (met name karakterisatie en normalisatie).

Module 4: Sterkteleer

De student is in staat om:

1. Spanningen en vervormingen te bepalen van balken en andere slanke elementen onder invloed van trek en druk, buiging, afschuiving en torsie.

Module 4: Werktuigonderdelen

De student is in staat om:

1. Werktuigbouwkundige ontwerp vraagstukken te analyseren door gebruik te maken van werktuigbouwkundige ontwerpprincipes, alternatieve concept oplossingen aan te bieden en hieromtrent een weloverwogen besluit nemen.
2. Werktuigbouwonderdelen in werktuigbouwkundige ontwerpen te benoemen en typeren en ze in relatie tot systeemspecificaties te analyseren.
3. Eenvoudige werktuigbouwkundige constructies te evalueren en een verantwoorde selectie te maken door gebruik te maken van gestandaardiseerde werktuigbouwonderdelen

Module 5: Dynamica 1

De student is in staat om:

1. Een goede Vrije Lichaam Structuur (VLS) te maken.
2. Bewegingsvergelijken op te stellen.
3. De wet van behoud van arbeid en energie toe te passen.
4. Kinematica te analyseren.

Module 5: Systeemanalyse

De student is in staat om:

1. Vanuit een Ideal Physical Model (IPM), en via een Free Body Diagram (FBD, Nederlands: Vrij Lichaam Schema (VLS)), het kunnen opstellen van een blokschema (Engels: Block diagram) van een dynamisch electro-mechanisch systeem.

2. Het, uit een blokschema, kunnen afleiden van een Transfer Function (TF, Nederlands: overdrachtfunctie).
3. Het kunnen analyseren van een Transfer Function in het tijddomein (staprespons) en het frequentiedomein (Bodediagram).

Module 5: Project Dynamische Systemen

De student is in staat om:

1. Vrijheidsgraden van simpele 3D mechanismen te analyseren.
2. Met basis elastische elementen (bladveren, sprieten, etc.) een translerend of roterend 3D mechanisme te kunnen ontwerpen.

Module 6: Tribology

De student is in staat om:

1. De volgende begrippen te beheersen: contact mechanica, wrijving, slijtage en smering.

Module 6: Elasticity Theory

De student is in staat om:

1. Vectoren, tensoren en lineaire algebra te gebruiken en ermee om te gaan.
2. Spanning, vervorming en materiaaleigenschappen “Elastic Isotopic Materials” te gebruiken.

Module 6: Processes and Properties of Polymers

De student is in staat om:

1. Aan te geven op welke manier de chemische en fysische opbouw van de molecuulketens de eigenschappen van een polymeer kan beïnvloeden.
2. Bestaande modellen voor (tijdsafhankelijke) kleine vervormingen in kunststof onderdelen toe te passen (lineaire elastische theorie inclusief Boltzmann en tijdtemperatuur-superpositie) in de berekening van vervorming of spanning.

Module 7: Fluid Mechanics 1

De student is in staat om:

1. Partiële differentiaalvergelijkingen te kunnen manipuleren met behulp van de productregel, de kettingregel, en de Einstein sommatie conventie, met als doel stromingseigenschappen te analyseren.
2. Een dimensie analyse op basis van een gegeven probleemstelling met een reeks van dimensievolle parameters te kunnen uitvoeren.

Module 8: Dynamica 2

De student is in staat om:

1. Een bewegingsvergelijking van een dynamisch systeem op te stellen
2. Inzicht te tonen in het dynamisch gedrag van een systeem (eigenfrequenties en –modes)
3. Een knikanalyse uit te voeren

Module 8: Systeem en Regeltechniek 1

De student is in staat om:

1. Elektromechanische systemen met meerdere vrijheidsgraden op een systematische manier modelleren in de vorm van blokschema's en daar overdrachtsfuncties uit afleiden.
2. Stabiele gesloten lus systemen ontwerpen in het frequentiedomein (Bode diagrammen) met gewenste cross-over frequentie en demping.
3. Instellingen van P(I)D regelaars voor elektromechanische systemen bepalen op basis van de cross-over frequentie die nodig is om een gewenste nauwkeurigheid te krijgen voor de responsie op tweede- of derdegraads opzetfuncties.
4. (Robuuste) stabiliteit evalueren van gesloten lus systemen met behulp van Nyquist diagrammen.

Module 11: Introduction Finite Element Method

De student is in staat om:

1. Aan te geven en te beschrijven wat de wiskundige en mechanische grondslagen zijn achter de Eindige Elementen Methode.
2. Op een efficiënte manier een Eindige Elementen model te maken van een constructie en deze te analyseren met behulp van het Eindige Elementen programma.
3. De resultaten van Eindige Elementen simulaties te interpreteren en een inschatting te geven van de nauwkeurigheid van de berekening.

Appendix 4 – Overview questions and elaborations primary and secondary assessment

The table below show the overview per course and the corresponding data for the CAs. The table shows an *x* when the data was collected. Column *Who* shows who the person was that made the questions or elaborations: *T* means Teachers, *D* means director and *S* means master student.

Table A4.1

Overview questions and elaboration primary and secondary CAs

Quartile	*Course*	*Goals*	*Primary CA*				*Secondary CA*			
			Questions	*Who*	*Elaboration*	*Who*	*Questions*	*Who*	*Elaboration*	*Who*
1	Statica & Modelleren en Programmeren 1	x	x	T	x	T	x	T	x	T
	Productiesystemen 1	x	x	T	x	T	x	S	x	S
	Werktuigbouwkundig tekenen	x	x	T	x	T	x	S	x	S
2	Tech. Thermodynamica 1	x	x	T	x	T	x	S	x	S
	Materiaalkunde 1	x	-	-	-	-	x	T	x	T
3	Tech. Thermodynamica 2	x	X	T	x	T	x	D	x	S
	Materiaalkunde 2	x	-	-	-	-	x	T	x	T
	Ketenbeheer	x	X	T	x	T	x	T	x	T
4	Sterkteleer & Mod. en Prog. 4	x	X	T	x	D	x	D	x	S+D
	Werktuigonderdelen	x	-	-	-	-	-	-	-	-
5	Dynamica 1	x	X	T	x	S	x	S+D	x	S
	Systeemanalyse	x	-	-	-	-	-	-	-	-
	Project Dynamische Systemen	x	X	T	x	T	x	D	x	D
6	Tribology	x	-	-	-	-	-	-	-	-
	Elasticity Theory	x	X	T	x	T	x	T	x	S+T
	Processes and Properties of Polymers	x	X	T	x	T	x	S	x	S
7	Fluid Mechanics 1	x	X	T	x	S	x	S	x	S
	Heat Transfer	-	-	-	-	-	-	-	-	-
8	Dynamica 2	x	x	T	x	T	x	T	x	T
	Systeem en Regeltechniek 1	x	x	T	x	T	x	T	x	T
11	Intro. Finite Element Method	x	x	T	x	T	x	D	x	S

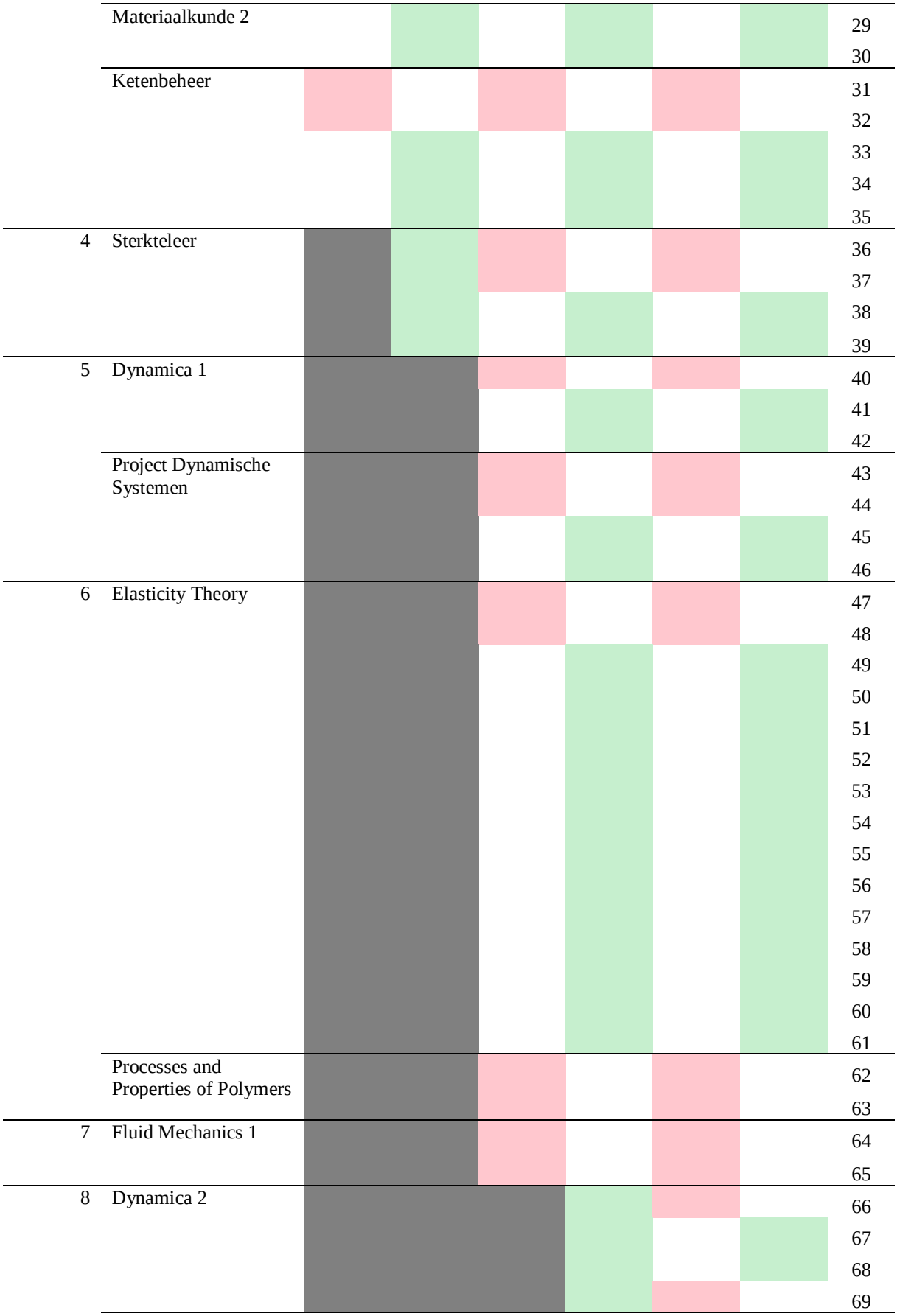
Appendix 5 – Question overview

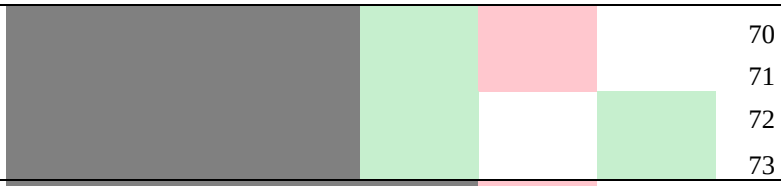
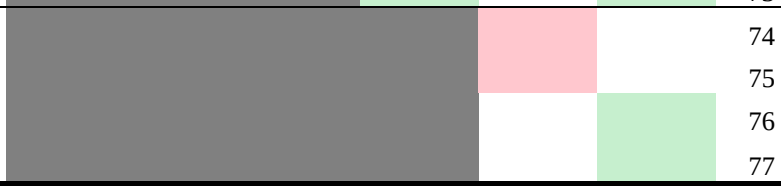
The table below shows the overview of all questions in the item bank and in which CA they were used. For example, question number 1 was used in CA 3, 7 and 11 and is therefore red: the second CA cannot use this question. The green questions are questions that could be used for the secondary CA.

Table A5.1

Question overview of all CAs

		CA number						
Module	*Course*	*3*	*4*	*7*	*8*	*11*	*12*	*#*
1	Statica en Modelleren en Programmeren 1							1
								2
								3
								4
								5
								6
	Productiesystemen 1							7
								8
								9
								10
	Werktuigbouwkundig tekenen							11
								12
								13
								14
								15
								16
								17
								18
2	Technische Thermodynamica 1							19
								20
								21
								22
	Materiaalkunde 1							23
								24
3	Technische Thermodynamica 2							25
								26
								27
								28


Materiaalkunde 2
Ketenbeheer
4 Sterkteleer
5 Dynamica 1
Project Dynamische Systemen
6 Elasticity Theory
Processes and Properties of Polymers
7 Fluid Mechanics 1
8 Dynamica 2
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69

	Systeem en Regeltechniek 1		70 71 72 73
11	Introduction Finite Element Method		74 75 76 77

Appendix 6 – Item analysis report of primary CA in year 1

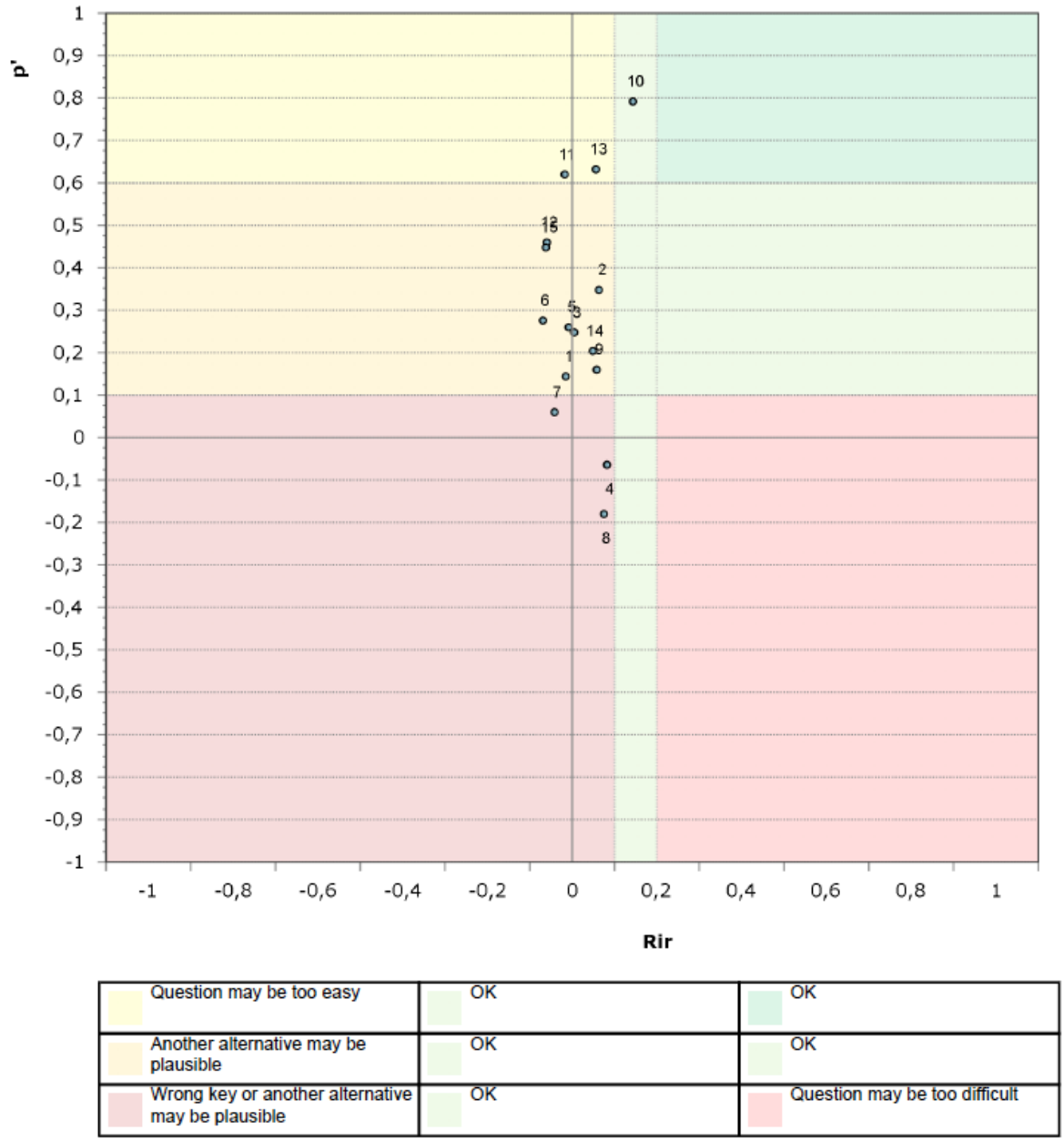


Figure A6.1 Rir- and p'-value analysis of primary CA in year 1

Table A6.1

Entire test analysis of primary CA in year 1

						A		*B*		*C*		*D*	
Item	*p*	*p'*	*Rir*	*Rit*	*Open*	*f*	*z*	*f*	*z*	*f*	*z*	*f*	*z*
1	0.35	0.14	-0.01	0.25	0	3	-0.35	34	0.14	**33**	**-0.01**	23	-0.15
2	0.51	0.34	0.07	0.34	0	0	*	3	-1.42	43	0.02	**47**	**0.07**
3	0.43	0.24	0.01	0.28	3	17	0.06	24	-0.16	**40**	**-0.01**	9	0.37
4	0.19	-0.08	0.09	0.30	1	14	0.01	8	0.31	52	-0.10	**18**	**0.16**
5	0.44	0.25	0.00	0.27	1	25	-0.16	8	0.75	18	-0.02	**41**	**-0.04**
6	0.45	0.27	-0.06	0.21	0	3	-0.29	**42**	**-0.07**	30	-0.07	18	0.32
7	0.29	0.05	-0.03	0.21	0	1	-0.38	52	-0.05	13	0.34	**27**	**-0.05**
8	0.11	-0.19	0.08	0.25	0	8	-0.90	39	0.07	**10**	**0.24**	36	0.06
9	0.37	0.15	0.07	0.32	2	12	-0.10	35	-0.04	**34**	**0.06**	10	0.08
10	0.84	0.78	0.15	0.34	4	2	0.16	3	-0.53	6	-0.03	**78**	**0.02**
11	0.71	0.61	-0.01	0.24	0	0	*	26	-0.04	**66**	**-0.01**	1	1.52
12	0.59	0.46	-0.05	0.22	1	5	0.09	**55**	**-0.08**	17	0.05	15	0.21
13	0.72	0.63	0.06	0.30	1	**67**	**0.04**	12	0.23	11	-0.35	2	-0.71
14	0.40	0.20	0.06	0.32	1	1	-0.33	**37**	**0.07**	46	0.07	8	-0.68
15	0.58	0.44	-0.05	0.21	1	28	0.19	**54**	**-0.05**	8	-0.30	2	0.05

Table A6.2

Entire test description of primary CA in year 1

Item	*Description*	*Rir*	*p'*
1	Another alternative may be plausible	-0,01	0,14
2	Another alternative may be plausible	0,07	0,34
3	Another alternative may be plausible	0,01	0,24
4	Wrong key or another alternative may be plausible	0,09	-0,08
5	Another alternative may be plausible	0,00	0,25
6	Another alternative may be plausible	-0,06	0,27
7	Wrong key or another alternative may be plausible	-0,03	0,05
8	Wrong key or another alternative may be plausible	0,08	-0,19
9	Another alternative may be plausible	0,07	0,15
10	OK	0,15	0,78
11	Question may be too easy	-0,01	0,61
12	Another alternative may be plausible	-0,05	0,46
13	Question may be too easy	0,06	0,63
14	Another alternative may be plausible	0,06	0,20
15	Another alternative may be plausible	-0,05	0,44

Appendix 7 – Item analysis report of primary CA in year 2

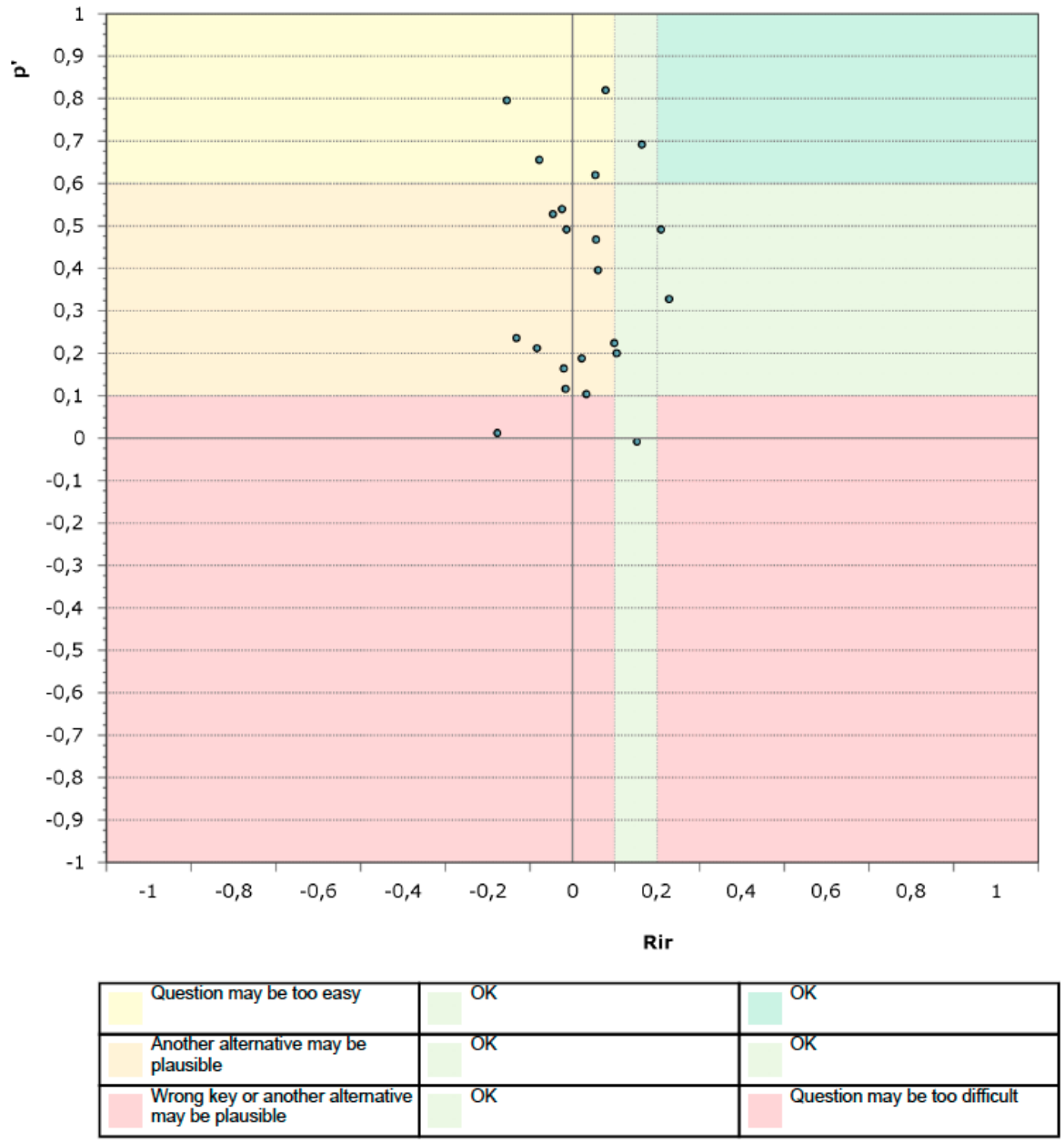


Figure A7.1 Rir- and p'-value analysis of primary CA in year 2

Table A7.1

Entire test analysis of primary CA in year 2

						A		*B*		*C*		*D*	
Item	*p*	*p'*	*Rir*	*Rit*	*Open*	*f*	*z*	*f*	*z*	*f*	*z*	*f*	*z*
1	0.39	0.18	0.03	0.24	3	30	0.02	**44**	**0.05**	13	-0.05	24	-0.1
2	0.24	-0.02	0.16	0.33	1	14	-0.01	16	-0.06	56	-0.11	**27**	**0.28**
3	0.65	0.53	-0.02	0.19	0	21	-0.08	5	-0.33	**74**	**-0.01**	14	0.3
4	0.84	0.79	-0.15	0.01	2	2	-0.94	8	0.34	6	0.55	**96**	**-0.04**
5	0.64	0.52	-0.04	0.17	0	28	0.17	12	-0.13	1	-1.28	**73**	**-0.03**
6	0.49	0.32	0.24	0.43	2	4	-0.29	38	-0.21	**56**	**0.23**	14	-0.26
7	0.4	0.2	-0.08	0.13	0	3	0.06	29	0.12	36	0.01	**46**	**-0.09**
8	0.42	0.23	-0.12	0.09	0	6	-0.29	9	-0.17	51	0.2	**48**	**-0.15**
9	0.74	0.65	-0.07	0.12	0	1	-0.37	28	0.08	**84**	**-0.04**	1	1.78
10	0.71	0.61	0.06	0.25	1	**81**	**0.04**	12	-0.13	18	-0.12	2	0.27
11	0.61	0.49	0.22	0.41	1	0	*	42	-0.26	**70**	**0.18**	1	-1.36
12	0.54	0.39	0.07	0.28	0	15	-0.11	**62**	**0.06**	24	0	13	-0.16
13	0.39	0.19	0.11	0.32	0	4	-0.2	**45**	**0.14**	17	-0.57	48	0.09
14	0.32	0.1	0.04	0.24	7	35	-0.22	17	0.1	**37**	**0.07**	18	0.18
15	0.33	0.11	-0.01	0.19	0	5	-0.29	**38**	**-0.01**	48	-0.09	23	0.26
16	0.6	0.46	0.06	0.27	1	5	-0.17	4	1.11	**68**	**0.06**	36	-0.21
17	0.61	0.49	-0.01	0.2	1	**70**	**0.01**	21	-0.19	9	0.51	13	-0.11
18	0.41	0.22	0.11	0.31	1	33	-0.14	**21**	**-0.12**	**26**	**0.34**	33	-0.05
19	0.86	0.81	0.09	0.23	1	8	-0.28	**98**	**0.02**	2	1.2	5	-0.51
20	0.76	0.68	0.17	0.35	1	5	-0.3	7	-0.45	**87**	**0.08**	14	-0.19
21	0.37	0.16	-0.01	0.19	0	3	-0.25	60	0.08	9	-0.39	**42**	**-0.02**
22	0.25	0.01	-0.17	0.01	7	**29**	**-0.3**	55	0.06	9	-0.01	14	0.39

Table A7.2

Entire test description of primary CA in year 2

Item	*Description*	*Rir*	*p'*
1	Another alternative may be plausible	0,03	0,18
2	OK	0,16	-0,02
3	Another alternative may be plausible	-0,02	0,53
4	Question may be too easy	-0,15	0,79
5	Another alternative may be plausible	-0,04	0,52
6	OK	0,24	0,32
7	Another alternative may be plausible	-0,08	0,20
8	Another alternative may be plausible	-0,12	0,23
9	Question may be too easy	-0,07	0,65
10	Question may be too easy	0,06	0,61
11	OK	0,22	0,49
12	Another alternative may be plausible	0,07	0,39
13	OK	0,11	0,19
14	Wrong key or another alternative may be plausible	0,04	0,10
15	Another alternative may be plausible	-0,01	0,11
16	Another alternative may be plausible	0,06	0,46
17	Another alternative may be plausible	-0,01	0,49
18	OK	0,11	0,22
19	Question may be too easy	0,09	0,81
20	OK	0,17	0,68
21	Another alternative may be plausible	-0,01	0,16
22	Wrong key or another alternative may be plausible	-0,17	0,01

Appendix 8 – Item analysis report of primary CA in year 3

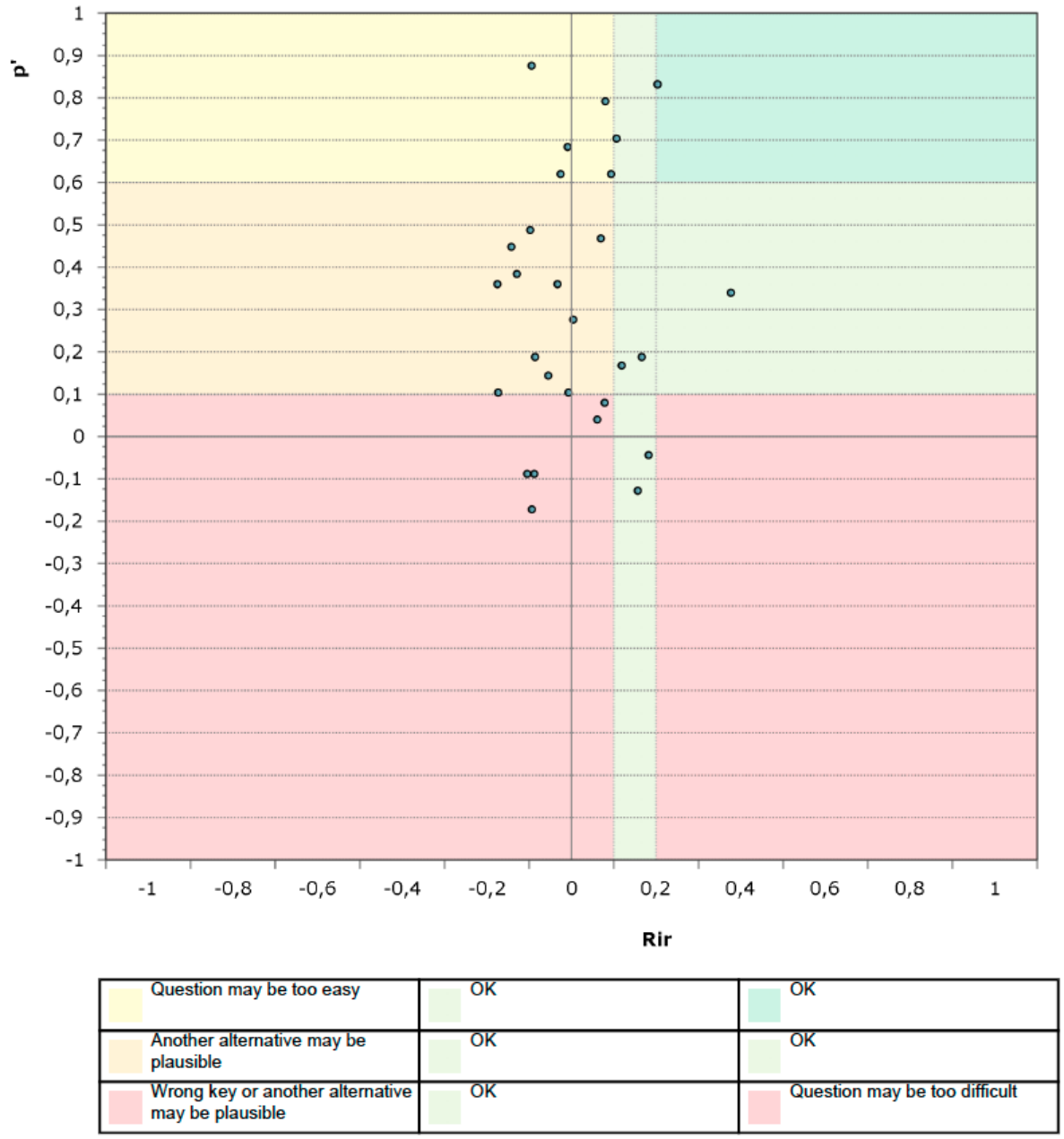


Figure A8.1 Rir- and p'-value analysis of *primary CA in year 3*

Table A8.1

Entire test analysis of primary CA in year 3

						A		*B*		*C*		*D*	
Item	*p*	*p'*	*Rir*	*Rit*	*Open*	*f*	*z*	*f*	*z*	*f*	*z*	*f*	*z*
1	0.60	0.46	0.08	0.27	0	**37**	**0.06**	6	0.15	11	0.00	8	-0.40
2	0.39	0.18	-0.08	0.12	2	24	0.05	**24**	**-0.08**	2	1.21	10	-0.16
3	0.15	-0.14	0.16	0.30	2	4	-0.74	8	-0.22	39	0.03	**9**	**0.39**
4	0.45	0.27	0.01	0.21	0	23	-0.02	3	-0.68	**28**	**0.01**	8	0.27
5	0.87	0.83	0.21	0.34	1	1	-1.52	5	-0.59	1	1.02	**54**	**0.06**
6	0.53	0.38	-0.12	0.08	0	15	-0.14	6	0.11	8	0.65	**33**	**-0.11**
7	0.61	0.48	-0.09	0.11	0	4	0.38	18	0.17	**38**	**-0.07**	2	-0.94
8	0.37	0.16	0.13	0.32	0	3	0.12	15	-0.08	21	-0.14	**23**	**0.16**
9	0.27	0.03	0.07	0.25	0	9	-0.01	9	-0.01	27	-0.06	**17**	**0.11**
10	0.71	0.61	0.10	0.28	0	**44**	**0.06**	6	0.26	12	-0.37	0	*
11	0.35	0.14	-0.05	0.15	1	2	-0.46	**22**	**-0.08**	31	0.09	6	0.02
12	0.71	0.61	-0.02	0.17	0	2	-0.70	16	0.12	**44**	**-0.01**	0	*
13	0.52	0.35	-0.17	0.03	0	4	-0.46	**32**	**-0.16**	24	0.11	2	2.24
14	0.76	0.68	0.00	0.17	0	0	*	13	0.11	**47**	**0.00**	2	-0.69
15	0.52	0.35	-0.03	0.18	0	2	-0.17	**32**	**-0.02**	4	-0.79	24	0.18
16	0.32	0.10	0.00	0.19	1	27	-0.09	6	0.16	**20**	**-0.01**	8	0.21
17	0.50	0.33	0.38	0.55	1	20	-0.29	3	-0.40	**31**	**0.39**	7	-0.72
18	0.39	0.18	0.17	0.36	1	20	-0.03	13	-0.47	**24**	**0.22**	4	0.31
19	0.31	0.08	0.09	0.27	0	1	-0.05	**19**	**0.13**	23	-0.45	19	0.42
20	0.18	-0.10	-0.08	0.08	0	5	0.96	2	0.71	**11**	**-0.17**	44	-0.10
21	0.58	0.44	-0.13	0.07	0	**36**	**-0.11**	11	0.51	10	-0.06	5	-0.18
22	0.90	0.87	-0.09	0.03	0	0	*	**56**	**-0.03**	1	0.60	5	0.20
23	0.84	0.78	0.09	0.24	0	**52**	**0.04**	4	0.49	3	-1.21	3	-0.10
24	0.18	-0.10	-0.10	0.06	0	0	*	43	0.07	8	-0.10	**11**	**-0.21**
25	0.21	-0.05	0.19	0.35	0	40	-0.13	**9**	**0.10**	**4**	**0.97**	9	0.05
26	0.32	0.10	-0.17	0.02	0	**20**	**-0.24**	36	0.21	5	-0.28	1	-1.24
27	0.77	0.70	0.11	0.28	0	1	-0.27	2	-0.06	**48**	**0.06**	11	-0.23
28	0.11	-0.18	-0.09	0.04	0	25	0.28	13	0.03	17	-0.34	**7**	**-0.24**

Table A8.2

Entire test description of primary CA in year 3

Item	*Description*	*Rir*	*p'*
1	Another alternative may be plausible	0.08	0.46
2	Another alternative may be plausible	-0.08	0.18
3	OK	0.16	-0.14
4	Another alternative may be plausible	0.01	0.27
5	OK	0.21	0.83
6	Another alternative may be plausible	-0.12	0.38
7	Another alternative may be plausible	-0.09	0.48
8	OK	0.13	0.16
9	Wrong key or another alternative may be plausible	0.07	0.03
10	OK	0.10	0.61
11	Another alternative may be plausible	-0.05	0.14
12	Question may be too easy	-0.02	0.61
13	Another alternative may be plausible	-0.17	0.35
14	Question may be too easy	0.00	0.68
15	Another alternative may be plausible	-0.03	0.35
16	Wrong key or another alternative may be plausible	0.00	0.10
17	OK	0.38	0.33
18	OK	0.17	0.18
19	Wrong key or another alternative may be plausible	0.09	0.08
20	Wrong key or another alternative may be plausible	-0.08	-0.10
21	Another alternative may be plausible	-0.13	0.44
22	Question may be too easy	-0.09	0.87
23	Question may be too easy	0.09	0.78
24	Wrong key or another alternative may be plausible	-0.10	-0.10
25	OK	0.19	-0.05
26	Wrong key or another alternative may be plausible	-0.17	0.10
27	OK	0.11	0.70
28	Wrong key or another alternative may be plausible	-0.09	-0.18

Appendix 9 – Item analysis report of secondary CA in year 1

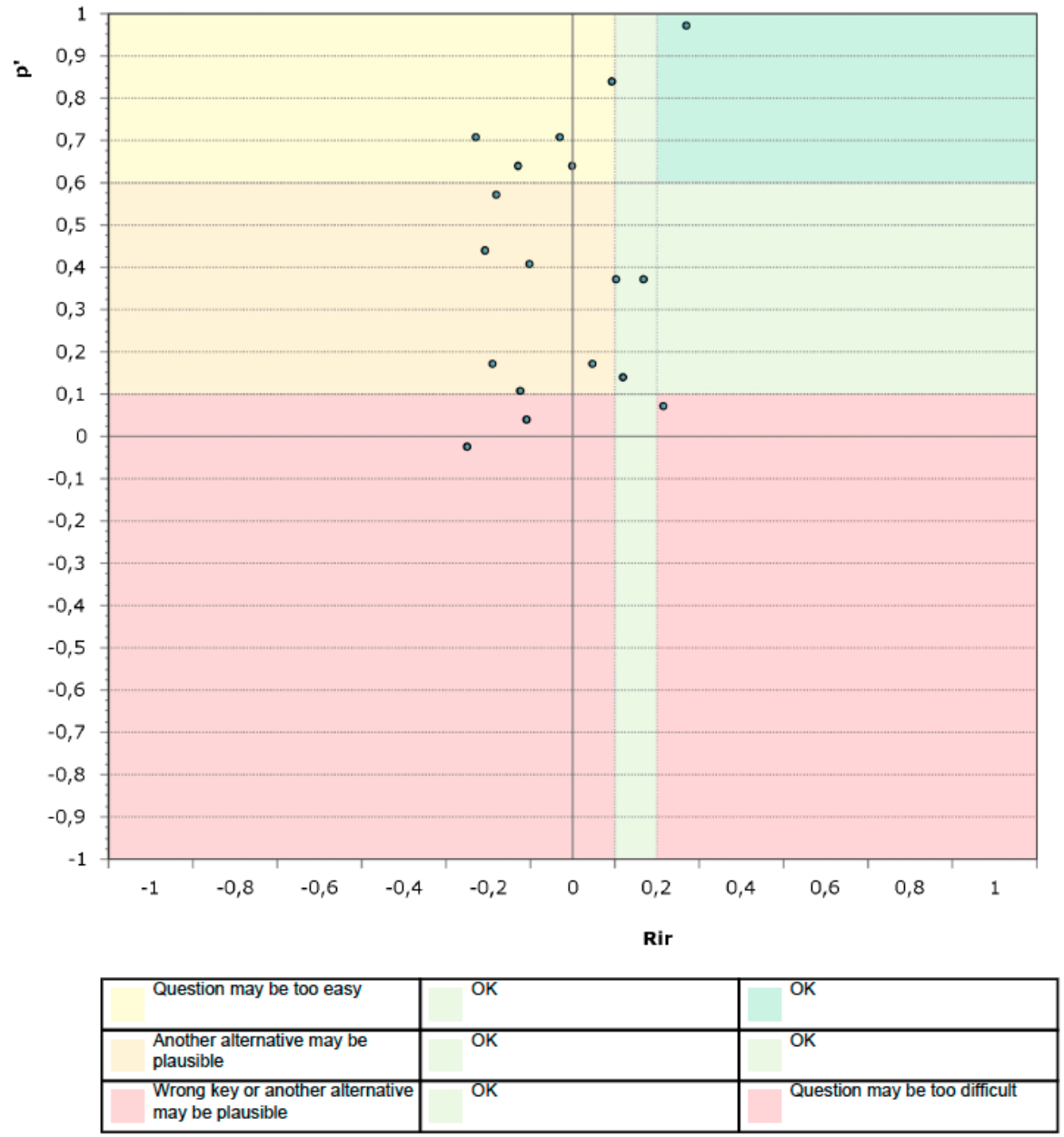


Figure A9.1 Rir- and p'-value analysis of secondary CA in year 1

Table A9.1

Entire test analysis of secondary CA in year 1

						A		B		C		D	
Item	*p*	*p'*	*Rir*	*Rit*	*Open*	*f*	*z*	*f*	*z*	*f*	*z*	*f*	*z*
1	0.98	0.97	0.28	0.36	0	0	*	0	*	**39**	**0.04**	1	-1.73
2	0.58	0.43	-0.20	0.08	0	1	0.93	14	0.08	**23**	**-0.17**	2	0.93
3	0.38	0.17	0.05	0.33	3	7	0.19	**15**	**0.25**	7	-0.60	8	-0.12
4	0.55	0.40	-0.09	0.19	0	**22**	**-0.09**	3	0.55	10	0.01	5	0.01
5	0.78	0.70	-0.22	0.01	0	8	0.47	**31**	**-0.12**	1	-0.08	0	*
6	0.30	0.07	0.22	0.46	1	15	-0.10	6	0.07	**12**	**0.38**	6	-0.57
7	0.23	-0.03	-0.24	-0.01	0	5	-0.61	11	0.52	15	0.09	**9**	**-0.45**
8	0.28	0.03	-0.10	0.15	0	15	0.09	**11**	**-0.16**	3	0.01	11	0.04
9	0.53	0.37	0.18	0.45	0	12	-0.04	3	-0.25	4	-0.56	**21**	**0.17**
10	0.68	0.57	-0.17	0.09	0	1	-1.83	9	0.30	**27**	**-0.12**	3	0.80
11	0.73	0.63	0.01	0.26	0	1	-1.89	**29**	**0.00**	3	0.47	7	0.05
12	0.38	0.17	-0.18	0.09	0	**15**	**-0.23**	6	-0.03	11	0.56	8	-0.31
13	0.73	0.63	-0.12	0.13	0	**29**	**-0.07**	8	0.03	2	1.03	1	-0.11
14	0.78	0.70	-0.02	0.22	0	4	-0.38	1	-0.09	4	0.50	**31**	**-0.01**
15	0.88	0.83	0.10	0.29	0	0	*	**35**	**0.04**	3	-0.03	2	-0.62
16	0.33	0.10	-0.12	0.15	0	4	-0.20	7	0.31	16	0.05	**13**	**-0.17**
17	0.35	0.13	0.13	0.39	0	7	-0.09	17	-0.03	**14**	**0.17**	2	-0.66
18	0.53	0.37	0.11	0.39	0	7	0.11	**21**	**0.11**	8	-0.55	4	0.37

Table A9.2

Entire test description of secondary CA in year 1

Item	Description	*Rir*	*p'*
1	OK	0,28	0,97
2	Another alternative may be plausible	-0,20	0,43
3	Another alternative may be plausible	0,05	0,17
4	Another alternative may be plausible	-0,09	0,40
5	Question may be too easy	-0,22	0,70
6	Question may be too difficult	0,22	0,07
7	Wrong key or another alternative may be plausible	-0,24	-0,03
8	Wrong key or another alternative may be plausible	-0,10	0,03
9	OK	0,18	0,37
10	Another alternative may be plausible	-0,17	0,57
11	Question may be too easy	0,01	0,63
12	Another alternative may be plausible	-0,18	0,17
13	Question may be too easy	-0,12	0,63
14	Question may be too easy	-0,02	0,70
15	OK	0,10	0,83
16	Another alternative may be plausible	-0,12	0,10
17	OK	0,13	0,13
18	OK	0,11	0,37

Appendix 10 – Item analysis report of secondary CA in year 2

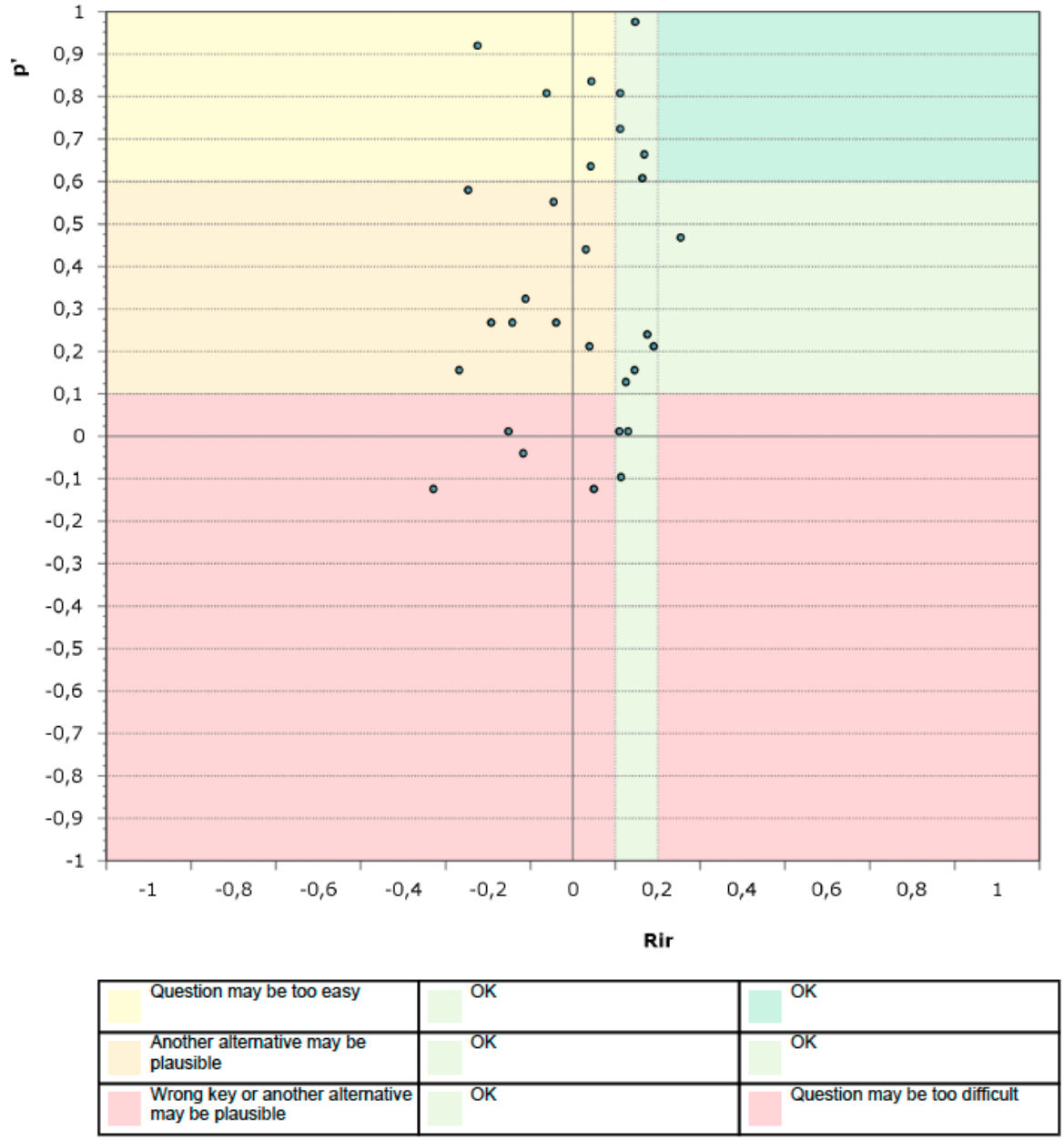


Figure A10.1 Rir- and p'-value analysis of secondary CA in year 2

Table A10.1

Entire test analysis of secondary CA in year 2

						A		*B*		*C*		*D*	
Item	*p*	*p'*	*Rir*	*Rit*	*Open*	*f*	*z*	*f*	*z*	*f*	*z*	*f*	*z*
1	0,87	0,83	0,05	0,19	0	0	*	1	-0,69	**41**	**0,02**	5	-0,02
2	0,40	0,21	0,20	0,39	0	8	-0,43	20	-0,06	**19**	**0,24**	0	*
3	0,34	0,12	0,13	0,32	2	11	0,21	**16**	**0,17**	10	-0,15	8	-0,43
4	0,49	0,32	-0,10	0,10	2	11	0,02	3	0,69	**23**	**-0,11**	8	0,03
5	0,57	0,43	0,04	0,24	1	15	-0,06	0	*	**27**	**0,07**	4	-0,27
6	0,85	0,80	0,12	0,26	0	3	0,13	1	-1,12	**40**	**0,05**	3	-0,42
7	0,43	0,23	0,18	0,37	2	**20**	**0,19**	8	0,04	8	-0,12	9	-0,35
8	0,26	0,01	-0,14	0,03	0	19	-0,05	13	0,17	**12**	**-0,25**	3	0,56
9	0,40	0,21	0,05	0,25	0	25	-0,05	3	0,09	0	*	**19**	**0,06**
10	0,66	0,55	-0,04	0,16	0	4	0,26	**31**	**-0,03**	3	0,19	9	-0,08
11	0,72	0,63	0,05	0,23	0	**34**	**0,03**	12	0,05	1	-1,59	0	*
12	0,36	0,15	-0,26	-0,07	2	3	-0,17	**17**	**-0,32**	12	-0,08	13	0,53
13	0,85	0,80	-0,05	0,09	0	1	0,13	**40**	**-0,02**	2	-0,28	4	0,33
14	0,60	0,46	0,26	0,44	2	5	0,00	7	-0,70	**28**	**0,19**	5	-0,09
15	0,98	0,97	0,15	0,21	0	0	*	**46**	**0,02**	1	-1,05	0	*
16	0,45	0,26	-0,18	0,01	1	0	*	**21**	**-0,20**	1	1,16	24	0,12
17	0,26	0,01	0,12	0,29	1	**12**	**0,16**	3	-1,03	25	-0,18	6	0,92
18	0,36	0,15	0,15	0,34	0	**17**	**0,20**	23	0,08	2	-0,93	5	-0,67
19	0,15	-0,13	0,06	0,20	1	6	-0,34	8	0,07	**7**	**0,16**	25	0,01
20	0,21	-0,05	-0,11	0,06	0	2	0,07	16	-0,10	19	0,19	**10**	**-0,21**
21	0,74	0,66	0,18	0,35	1	5	-1,09	5	0,09	1	1,36	**35**	**0,10**
22	0,94	0,91	-0,22	-0,12	0	2	0,56	**44**	**-0,06**	1	1,36	0	*
23	0,15	-0,13	-0,32	-0,19	0	14	0,21	24	0,09	**7**	**-0,77**	2	0,05
24	0,45	0,26	-0,13	0,07	2	8	0,24	7	0,23	**21**	**-0,18**	9	0,03
25	0,26	0,01	0,14	0,31	1	8	0,01	6	0,12	20	-0,20	**12**	**0,26**
26	0,70	0,60	0,17	0,35	1	3	-0,11	**33**	**0,08**	4	0,04	6	-0,40
27	0,68	0,57	-0,24	-0,06	0	3	-0,07	**32**	**-0,16**	6	0,45	6	0,45
28	0,45	0,26	-0,03	0,17	0	17	0,21	5	-0,04	4	-0,65	**21**	**-0,04**
29	0,17	-0,11	0,12	0,27	0	0	*	**8**	**0,27**	25	-0,03	14	-0,09
30	0,79	0,72	0,12	0,28	0	3	0,11	5	-0,73	**37**	**0,06**	2	0,53

Table A10.2

Entire test description of secondary CA in year 2

Item	*Description*	*Rir*	*p'*
1	Question may be too easy	0,05	0,83
2	OK	0,20	0,21
3	OK	0,13	0,12
4	Another alternative may be plausible	-0,10	0,32
5	Another alternative may be plausible	0,04	0,43
6	OK	0,12	0,80
7	OK	0,18	0,23
8	Wrong key or another alternative may be plausible	-0,14	0,01
9	Another alternative may be plausible	0,05	0,21
10	Another alternative may be plausible	-0,04	0,55
11	Question may be too easy	0,05	0,63
12	Another alternative may be plausible	-0,26	0,15
13	Question may be too easy	-0,05	0,80
14	OK	0,26	0,46
15	OK	0,15	0,97
16	Another alternative may be plausible	-0,18	0,26
17	OK	0,12	0,01
18	OK	0,15	0,15
19	Wrong key or another alternative may be plausible	0,06	-0,13
20	Wrong key or another alternative may be plausible	-0,11	-0,05
21	OK	0,18	0,66
22	Question may be too easy	-0,22	0,91
23	Wrong key or another alternative may be plausible	-0,32	-0,13
24	Another alternative may be plausible	-0,13	0,26
25	OK	0,14	0,01
26	OK	0,17	0,60
27	Another alternative may be plausible	-0,24	0,57
28	Another alternative may be plausible	-0,03	0,26
29	OK	0,12	-0,11
30	OK	0,12	0,72

Appendix 11 – Item analysis report of secondary CA in year 3

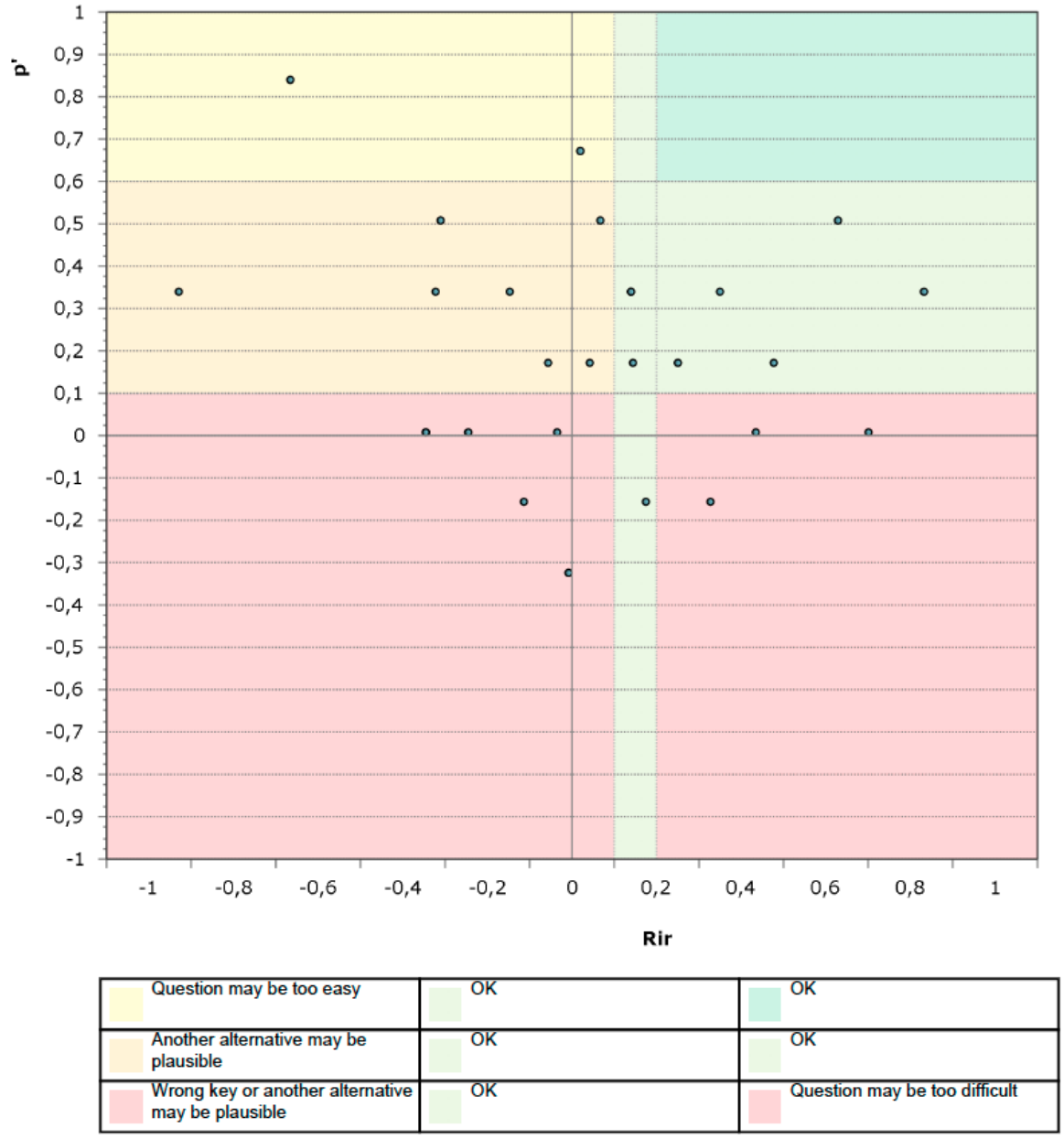


Figure A11.1 Rir- and p'-value analysis of secondary CA in year 3

Table A11.1

Entire test analysis of secondary CA in year 3

						A		B		C		D	
Item	*p*	*p'*	*Rir*	*Rit*	*Open*	*f*	*z*	*f*	*z*	*f*	*z*	*f*	*z*
1	0.75	0.67	0.03	0.19	0	0	*	0	*	**6**	**0,02**	2	-0,05
2	0.25	0.00	-0.34	-0.19	0	1	-0,58	5	0,35	**2**	**-0,58**	0	*
3	0.13	-0.17	0.18	0.30	0	4	-0,68	**1**	**0,48**	2	0,29	1	1,64
4	0.50	0.33	-0.14	0.05	0	2	-0,14	2	0,42	**4**	**-0,14**	0	*
5	0.38	0.17	0.05	0.23	0	4	-0,38	1	1,35	**3**	**0,06**	0	*
6	0.50	0.33	0.36	0.52	0	0	*	3	-0,02	**4**	**0,36**	1	-1,38
7	0.50	0.33	0.15	0.33	0	**4**	**0,15**	1	-1,33	2	0,64	1	-0,54
8	0.50	0.33	0.15	0.33	0	2	0,25	1	-0,54	**4**	**0,15**	1	-0,54
9	0.50	0.33	-0.92	-0.89	0	3	0,95	1	0,84	0	*	**4**	**-0,92**
10	0.50	0.33	0.84	0.89	0	0	*	**4**	**0,84**	3	-0,91	1	-0,62
11	0.38	0.17	0.15	0.33	0	**3**	**0,20**	2	-0,20	2	0,59	1	-1,37
12	0.63	0.50	-0.30	-0.13	0	0	*	**5**	**-0,23**	2	-0,09	1	1,35
13	0.38	0.17	0.26	0.42	0	0	*	**3**	**0,33**	4	-0,60	1	1,40
14	0.50	0.33	0.15	0.33	0	1	-0,54	3	-0,02	**4**	**0,15**	0	*
15	0.25	0.00	-0.34	-0.19	0	5	0,35	0	*	**2**	**-0,58**	1	-0,58
16	0.88	0.83	-0.66	-0.58	0	0	*	**7**	**-0,25**	1	1,74	0	*
17	0.13	-0.17	0.34	0.44	0	0	*	**1**	**0,89**	0	*	7	-0,13
18	0.25	0.00	-0.34	-0.19	0	**2**	**-0,58**	1	-0,58	2	1,03	3	-0,10
19	0.38	0.17	0.26	0.42	0	**3**	**0,33**	4	-0,30	0	*	1	0,20
20	0.63	0.50	0.64	0.74	0	0	*	2	-0,96	**5**	**0,49**	1	-0,54
21	0.00	-0.33	0.00	0.00	0	2	-1,08	5	0,27	1	0,80	**0**	*****
22	0.50	0.33	-0.31	-0.14	0	3	0,82	0	*	1	-1,21	**4**	**-0,31**
23	0.00	-0.33	0.00	0.00	0	1	-1,45	7	0,21	**0**	*****	0	*
24	0.25	0.00	-0.24	-0.08	0	1	-0,59	5	0,28	**2**	**-0,41**	0	*
25	0.13	-0.17	-0.11	0.02	0	3	-0,90	2	1,03	**1**	**-0,28**	2	0,47
26	0.25	0.00	-0.03	0.14	0	0	*	1	0,14	5	-0,01	**2**	**-0,05**
27	0.38	0.17	0.48	0.62	0	1	-0,63	**3**	**0,63**	1	-1,46	3	0,07
28	0.25	0.00	0.71	0.78	0	3	-0,41	**2**	**1,23**	2	-0,69	1	0,16
29	0.38	0.17	-0.05	0.13	0	3	-0,06	2	0,19	0	*	**3**	**-0,06**
30	0.25	0.00	0.44	0.57	0	1	1,79	**2**	**0,77**	3	-0,94	2	-0,26
31	0.00	-0.33	0.00	0.00	0	3	-0,08	4	-0,14	**0**	*****	1	0,80
32	0.63	0.50	0.07	0.25	1	0	*	1	1,35	**5**	**-0,13**	1	-0,70

Table A11.2

Entire test description of secondary CA in year 3

Item	*Description*	*Rir*	*p'*
1	Question may be too easy	0,03	0,67
2	Wrong key or another alternative may be plausible	-0,34	0,00
3	OK	0,18	-0,17
4	Another alternative may be plausible	-0,14	0,33
5	Another alternative may be plausible	0,05	0,17
6	OK	0,36	0,33
7	OK	0,15	0,33
8	OK	0,15	0,33
9	Another alternative may be plausible	-0,92	0,33
10	OK	0,84	0,33
11	OK	0,15	0,17
12	Another alternative may be plausible	-0,30	0,50
13	OK	0,26	0,17
14	OK	0,15	0,33
15	Wrong key or another alternative may be plausible	-0,34	0,00
16	Question may be too easy	-0,66	0,83
17	Question may be too difficult	0,34	-0,17
18	Wrong key or another alternative may be plausible	-0,34	0,00
19	OK	0,26	0,17
20	OK	0,64	0,50
21	Wrong key or another alternative may be plausible	0,00	-0,33
22	Another alternative may be plausible	-0,31	0,33
23	Wrong key or another alternative may be plausible	0,00	-0,33
24	Wrong key or another alternative may be plausible	-0,24	0,00
25	Wrong key or another alternative may be plausible	-0,11	-0,17
26	Wrong key or another alternative may be plausible	-0,03	0,00
27	OK	0,48	0,17
28	Question may be too difficult	0,71	0,00
29	Another alternative may be plausible	-0,05	0,17
30	Question may be too difficult	0,44	0,00
31	Wrong key or another alternative may be plausible	0,00	-0,33
32	Another alternative may be plausible	0,07	0,50