Multidisciplinary Course Assessment with Multiple Assessors at TU/e

Mei, 2016

Dr. K. D. (Kelly) Meusen-Beekman

Department of Industrial Engineering & Innovation Sciences (IE&IS)

Innovation, Technology Entrepreneurship & Marketing (ITEM) group

Dr. J. J. L. (Jeroen) Schepers

Department of Industrial Engineering & Innovation Sciences (IE&IS)

Innovation, Technology Entrepreneurship & Marketing (ITEM) group

Dr. ir. P. A. M. (Ad) Kleingeld

Department of Industrial Engineering & Innovation Sciences (IE&IS)

Human Performance Management (HPM) group

This report was commissioned by *4TU CEE Centre of Engineering Education*.

Table of contents

Abstract

A challenge in the domain of multidisciplinary education is the assessment of students' work. Students have to apply theories and concepts from different academic disciplines, but safeguarding the educational and assessment quality is difficult, because the assessors are typically specialists, not generalists. The goal of this project was to examine the boundary conditions for and optimal design of multidisciplinary assessment. The course Industrial Engineering (IE) Quick Scan was chosen as a case study. Insight into the effectiveness of multidisciplinary assessment and its boundary conditions was gained through a literature study. Interviews with assessors and students then generated insights into the current assessment process in the IE Quick Scan. Consequently, a field experiment was conducted to examine whether lower expertise on a subject impairs an assessor to reliably assess students' work and make grading decisions. The results of this project provide features of effective multidisciplinary assessment, and provide information on the accuracy of grading procedures when specialist assessors assess multidisciplinary assignments. This results in clear guidelines to come to reliable, and valid assessments of such assignments, irrespective of the assessors' expertise. Further research is needed to explore whether holistic or analytic approaches are preferred within the multidisciplinary course IE Quick Scan.

1 Introduction

1.1 Background and context

The future role of engineers in society is multifaceted: many engineering challenges are multidisciplinary in nature and solving them requires input from the engineering, natural, and social sciences (Meijers and Den Brok 2013). Eindhoven University of Technology's current educational programming allows, and even stimulates, students to create a personalized and multidisciplinary educational profile by offering a large number of elective courses.

One of the dangers of the increased number of elective courses in students' curricula is that, in the end, a student ends up being a "jack of all trades" who does not know how to "connect the dots". In other words, although knowledge may be differentiated across courses, students should also learn how to integrate these pieces of the engineering puzzle.

Fortunately, many programs set up integrative courses, in which students have to apply theories and concepts from different academic fields and perspectives. For instance, the master program Innovation Management (IM) offers a "Design Project", in which students design a solution for a business problem based on multiple perspectives they have learned in the program. Similarly, the Industrial Engineering (IE) bachelor program offers the Quick Scan IE (1CK100)", in which students have to analyze the business processes of their internship company using multiple frameworks (from multiple areas of expertise) they have learned during their major. This elective course is taken by the majority of IE students (approximately 150 students per year).

Although these integrative courses are desirable, safeguarding their educational quality is a difficult endeavor. Many staff members are specialists in one field of interest and are unable to reliably grade the students' work in all parts of such integrative projects. For instance, in the IE Quick Scan, there are five areas of expertise: Accounting and Finance, Human Performance Management, Information Management, Operations Management, and Innovation Management. A specialist is typically an expert in one area, and perhaps somewhat knowledgeable in an additional area, but lacks knowledge of the three other areas. In some areas of expertise, students are even likely to be more knowledgeable than the assessor. This has two major implications.

First, this situation may compromise the validity and reliability of grading. Especially for educational visitations and accreditation boards, where focus is put on assessors' qualifications and grading procedure quality, the observation of such suboptimal procedures may cause problems. The responsible lecturer of the IE Quick Scan and the course design team have tried to remedy this situation by creating an assessor manual, organizing an assessor training day, and by drafting detailed answer models for each element of the IE Quick Scan to assist each assessor in his/her grading task. However, the team still has serious doubts about the ability of assessors to accurately grade students' performance in parts of the internship that fall outside their area of expertise. It could be that assessors extrapolate scores in their area of expertise to other parts of the report. In addition, central tendency and leniency errors may occur. Given the number of students and the variety of disciplines in a Quick Scan report, assessors are unlikely to ask other assessors (i.e., specialists) to help in grading. Such consultation would be infeasible in terms of personal resource availability.

Second, the situation may compromise the quality of feedback provided to students. An assessor can provide detailed feedback on the parts of the report that fall within his/her area of expertise, but may be unable to provide in-depth remarks in other parts. An assessor can indicate where a student's answer was inconsistent with the answer model for a specific element in the report, but is unlikely to provide more examples or details of what a perfect answer should have looked like. In fact, the responsible lecturer of the IE Quick Scan and the course design team fear that assessors will try to evade students' questions about their assessment, because they may be unable to come up with a satisfactory answer.

1.2 Aim of the study

Given the above, the aim of this study is to investigate whether the current assessment procedure in the IE Quick Scan attains the minimum requirements in educational quality, and whether and how this procedure can be redesigned to improve its validity and reliability. To to so, this project combines theory and empirics. First, a literature study on multidisciplinary assessment outlines the boundary conditions for effective assessment of multidisciplinary assignments. Second, interviews and a field experiment are conducted to provide more insights into the experiences of assessors, the validity and reliability of current assessment procedures, and potential improvement measures in the IE Quick Scan course.

More specifically, the following research questions were specified:
1. What is multidisciplinary assessment and how can it support student learning?
2. What criteria and standards are required to assess multidisciplinary assignments?
3. To what extent does the current assessment procedure for IE Quick scan meet the requirements of reliability and validity of grading?

2 Literature study

2.1 Introduction

To get an overview of the development of multidisciplinary assessment in higher education, recent international scientific literature on the topic was explored. Four electronic databases were used: Academic Search Elite, Education Resources Information Center (ERIC), Google Scholar and SpringerLink. Multidisciplinary assessment is associated with an abundance of terminology. Therefore, search terms such as 'multidisciplinary assessment', 'interdisciplinary assessment', 'transdisciplinary assessment' and 'integrated assessment' were employed. This search strategy resulted in 50,000 articles. Based on the review of learning skills interventions on student learning by Drake (1991), the following key words were further included: 'evaluation', 'learning inquiry', and 'instruction'. For the first research question, the search terms were combined with keywords such as 'higher education' and 'university education'. This search strategy resulted in 390 articles. The selected literature was screened to identify conditions, methods and outcomes with regard to the effectiveness of multidisciplinary assessment. Of these articles, 11 publications examined multidisciplinary assessment in university education. After this, references within the articles were used to find more studies. The articles included in this review were published between 2000 to 2015. The literature was used to write a narrative synthesis to provide a theoretical view.

2.2 Multidisciplinary learning

Multidisciplinary learning is a fuzzy concept (Klein 2000, 2002; Newell, 2001, 2002). According to Moss, Osborne, and Kaufman (2008) multidisciplinary learning is the capacity to integrate knowledge and modes of thinking drawn from more than one discipline to produce a cognitive advancement. Multidisciplinary learning develops several cognitive abilities (Schilling, 2001), such as developing and applying perspective-taking techniques and developing structural knowledge of problems appropriate to multidisciplinary inquiry.

The literature on multidisciplinary learning contains many attempts to organize the forms of multi-, inter- and transdisciplinary work into a coherent framework. A common framework used in multidisciplinary learning builds on the extent of topic integration, which is typically portrayed on a continuum moving from a disciplinary approach to increasing cross-topic connections and greater degrees of integration (Daly, Brown & McGowan, 2012; Drake, 2012; Fogarty & Pete, 2009). On the latter side of the continuum, a multidisciplinary approach indicates that a central theme is identified and used to organize and correlate the disciplines being integrated (Brough, 2012; Dowden, 2007). Key in multidisciplinary learning is that discipline integrity is preserved. Learning activities are shaped around subject areas so that unique discipline-based concepts can be used to study the area of interest (Drake, et al., 2015). Assessment is clear for each discipline. The assessor may or may not make the connections between subject areas explicit for the students. Sometimes the culminating activity requires the integration of the skills and knowledge of the subject areas (Drake et al., 2015). Disciplinary insights are not in conflict with multidisciplinary learning, rather they embody knowledge (Boix Mansila, 2005).

Then, how does multidisciplinary learning differ from interdisciplinary and transdisciplinary learning? As opposed to multidisciplinary work, in which discipline integrity is preserved, interdisciplinary work requires integration of knowledge and innovation. Interdisciplinary work applies to subjects that are relatively clear in their differences, but consists of shared key concepts, skills and attitudes and actions (Drake, et al., 2015). Lattuca, Voigt, and Fath referred to interdisciplinary learning outcomes such as "promoting the development of sophisticated views of knowledge and learning" and "building students' capacity to recognize, evaluate, and integrate differing (multiple) perspectives" (2004, p. 44). In interdisciplinary work sustainability, change and continuity, or complex interdisciplinary skills such as communication, critical thinking and problem-solving are emphasized. According to Drake et al. (2015), assessment should reflect the extent to which particular disciplinary standards are met. Sometimes each discipline receives a similar grade in the interdisciplinary aspect, for example by demonstrating collaborative problem-solving skills.

Within transdisciplinary work, students go beyond the disciplines. According to Richard and Bennett (2011) transdisciplinary work can be described as an examination of multiple disciplines to integrate knowledge and state that this is the highest level of integration. ‘Transdisciplinarity works to remove the notion that certain content matter is necessarily owned by any particular discipline, and we do not engage in multidisciplinary studies to meet outside requirements that identify exposure to specific content as the primary goal. Our goal is to find a problem or idea worth studying and bear the visions of multiple perspectives upon it in order to understand it more fully than if we were to observe it from a single vantage point. This understanding inevitably leads to content learning: in the process of using the disciplines in the same ways that a discipline expert would use them to view the world, students and teachers learn the content that attracted subject-area scholars to the discipline in the first place. However, the larger payoff is that students know how to use the content to continue to grow’ (Kaufman et al., 2003, p. 158).

Although multidisciplinary learning is an elusive concept, related to inter- and transdisciplinary learning, the starting point for a multidisciplinary approach in any course or assignment is to begin with standards from different disciplines and integrate these through the lens of a common theme (Drake, 2007). Multidisciplinary learning is informed by disciplinary expertise; it builds on knowledge and modes of thinking that are central to the work of experts in various domains. Students should be able to use knowledge and understand a concept when they are able to apply it accurately and flexibly in novel situations (Boix Mansila, 2005). In interdisciplinary learning, one explores common concepts or skills embedded in the standards, whereas transdisciplinary learning starts in a real world context, after which each context is shaped by the standards.

The conceptualization of multidisciplinary learning has implications for multidisciplinary assessment.

Over the years, researchers have become increasingly concerned about validity, reliability, and accountability of multidisciplinary assessment, which has led to a growing attention for quality measurement tools. However, literature is limited in its ability to indicate how multidisciplinary courses may best be assessed. Yet it is important to clearly articulate what are indicators of high-quality multidisciplinary work, and how such indicators can be measured.

2.3 Assessing multidisciplinary learning

The goal of multidisciplinary student work is to produce a cognitive advancement. However, assessing quality of multidisciplinary work is difficult, with regard to exchanging methods, translating categories, and testing outcomes against multiple standards of quality (Boix Mansila, 2005). Multidisciplinary assessment is not only about mastering multiple disciplines (Borja et al., 2008). Assessing multidisciplinary work requires carefully considering disciplinary grounding by a focus on disciplines and the use of knowledge (Wolfe & Haynes, 2003a). It involves considering how students selected some insights, such as particular theories, methods, tools, and forms of communication (Borja et al., 2008). In addition, articulating the level of understanding in students' work is important to weigh the affordances of several disciplinary perspectives against others, and against the overall purpose of the student's enterprise (Boix Mansila, 2005). Assessing students' work requires exploring the means by which goals were reached, and the limitations of students' work. On top of those conditions for effective multidisciplinary assessment, the assessment should be consistent, practicable, timely and effective in providing evidence course intended learning outcomes (Nikitina, 2006). Despite the benefits of multidisciplinary work to students, its assessment may be problematic. To improve the quality of multidisciplinary assessment guiding assessment questions are suggested (see Appendix A), and several key criteria are required. The following section outlines conditions and criteria for multidisciplinary assessment.

2.3.1 Criteria and standards

Multidisciplinary assessment requires the use of strategies to demonstrate competence in the subjects that are assessed and to provide an accurate evaluation of the student's overall proficiency (Fiels & Stowe, 2002; Boix Mansilla & Duraising, 2007; Stowe & Eder, 2002). Students should be able to build and demonstrate mastery of performance in a multidisciplinary assessment task. Features of quality of student multidisciplinary work can be grouped under several categories, such as incorporating multiple perspectives, critical and/or logical thinking, respect for disciplinary standards, and substantiated and grounded work (Borja et al., 2008; Moss, Osborn, & Kaufman, 2012).

The assessment must stand on valid indicators of what counts as accomplished student work and need to be reliable, valid and trustworthy (Wiggins, 1998). Clarity about indicators of quality is particularly evident in the assessment of student multidisciplinary work (Borja et al., 2008). Judging the quality of student work is generally based on the teachers' expectations (which cannot be fully communicated using criteria), and how other students have performed (despite protestations to the contrary) (Sadler, 2005). Within the context of assessment and grading in higher education, criteria and standards are crucial. Sadler (1987, pp. 82) defines 'standard' as 'a definite level of excellence or attainment, or a definite degree of any quality viewed as a prescribed object of endeavor or as the recognized measure of what is adequate for some purpose, so established by authority, custom, or consensus.' The terms 'criterion' and 'standard' are often used interchangeably, even though they are not fully equivalent. A criterion refers to a characteristic or to a minimum qualifying level (Johnson & Svingby, 2007). Of the two terms, criterion is broader in scope and therefore more inclusive.

Assessment criteria should establish the level of achievement that is required for a student to pass the course and should be directly related to the course learning outcomes (Boix Mansilla & Duraising, 2007). In addition, criteria for multidisciplinary assessment should include validity within and beyond the disciplines (Ackerman, 2000). Evidence of learning should be authentic and demonstrate valid learning (Boix Mansilla, 2005).

According to Ben-David (2000) the understanding of the criteria involved are crucial in producing agreement between assessors. It is emphasized to clarify and share criteria and standards between faculty and students (Boix Mansilla & Duraising, 2007).

In addition, effective multidisciplinary assessments provide consistent and constructive feedback on students' progress, processes and results, focusing on how to improve their work. In general, multidisciplinary assessments drive the students' learning processes (Boix Mansilla & Duraising, 2007).

2.3.2 How can the assessor be supported?

Essentially, all grading processes (criteria-based included) start with the professional judgments of lecturers as to the standards that are employed. The lecturer has to have an advantage over the students regarding superior knowledge and extensive experience. To support the assessor in assessing the quality of a task, assessment models are often used in which the quality of a product is assessed based on criteria, standards and rating scales (Van Berkel, 2012). The approaches of scoring can be distinguished in holistic and analytical judgments (Kuhlemeier, 2002; Mertler, 2004). Researchers are indecisive about which scoring model can be preferred. In holistic scoring, the rater makes an overall judgment about the quality of a product. Because learning and instruction are increasingly competence-based, holistic scoring to adequately determine competence is emphasized by several scholars (Van der Vleuten & Schuwirth, 2005; Oosterheert, Eldik, & Kral, 2007; Van Berkel (2012). Knowledge, skills and attitudes can be assessed separately and integratively, scoring both explicit and implicit expertise. In analytic scoring, the rater assigns a score to each of the criteria assessed in the product (Johnson & Svingby, 2007), after which the final score is summarized by all separate scores. The criteria then act as the primary reference points against which student submissions are judged (Sadler, 2005) and serve as the basis for the communication and feedback process. According to

Kuhlemeier (2002), not only holistic scoring but also analytic scoring is well applicable in difficult tasks.

Whether work is referred to as 'good', 'sufficient' or 'insufficient' should depend on a shared interpretation of the underlying criteria and, at least implicitly, the corresponding grade boundaries (Sadler, 2005). Hence, important in criteria-based assessment and grading is that students are informed in advance about how judgments of the quality of their performances will be made, and to assure them a degree of objectivity, excluding extraneous factors or the interference of previous achievements and performance history (Joughin, 2009). Unfortunately, students often do not gain access to the judgmental processes of the assessor, or the criteria that are applied.

2.4 Reliability of scoring

Different assessors should reach similar conclusions when assessing the same product. The more consistent the scores are across different raters and occasions, the more reliable the assessment is perceived to be (Moskal & Leydens, 2000). However, particularly multidisciplinary assessment can be perceived as complex with respect to reliability. Several findings support evidence of high reliability in assessment when students perform the same multidisciplinary tasks and scoring procedures are well defined. When students have diverging assignments, choose their own topics or produce unique solutions, then assessment reliability may be at risk (Jonsson & Svingby, 2007).

Variability in the assessment scores can appear due to variations in the raters' judgments on different criteria (Black, 2000). Judgment (in)consistency can occur either across raters, known as interrater reliability, or in the consistency of one single rater, called intra-rater reliability (Jonsson & Svinby, 2007).

2.4.1 Intra-rater reliability

Brown, Bull, and Pendlebury (1997) cite that the "major threat to reliability is the lack of consistency of an individual marker" (p. 235). Most of the studies investigating intra-rater reliability use Cronbach's alpha to estimate raters' consistency. According to Brown, Glasswell, and Harland (2004) alpha values above .70 are generally considered sufficient. Results from studies investigating intra-rater reliability indicate that rubrics seem to aid raters in achieving high internal consistency when scoring tasks (Jonsson & Svinby, 2007). A rubric is a document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing levels of quality from excellent to poor (Andrade, Wang, & Du, 2009). Also Brown, Glasswell and Harland (2004) emphasize that intra-rater reliability might not be a major concern when raters are supported by a rubric although the consensus agreement of raters depends heavily on the number of levels in the rubric. Fewer levels seem to increase the chance of agreement, also due to increased probability of agreement by chance.

2.4.2 Interrater reliability

There are several factors that underlie diverging assessments of different assessors. Besides obvious reasons for disagreement (such as differences in experience), it has been reported that factors such as lecturers' attitudes regarding students and content influence the rating of students' work (Davidson, Howell, & Hoekema, 2000).

When it comes to accuracy and consistency of scoring, interrater reliability is an important factor related to reliability and validity. Stemler (2004) refers to three main approaches to determine the accuracy and consistency of scoring. The first approach concerns consensus estimates, which measures the degree to which markers attach the same score to the same performance. Consensus estimates are based on the assumption that 'reasonable observers should be able to come to exact agreement about how to apply the various levels of a scoring rubric to the observed behaviors' (p.62). Consensus estimates are most useful when data are nominal and when different levels of the rating scale represent qualitatively different ideas, but are ordinal in nature (e.g., a Likert scale). Therefore, the importance is stressed of assessors being trained to the point where they agree on how to interpret a rating scale.

The second approach refers to consistency estimates of interrater reliability. Consistency estimates measure the correlation of scores among raters (Stemler, 2004). Within this approach, it is assumed that assessors do not have to share a common meaning of the rating scale, as long as each assessor is consistent in classifying the aspects according to his definition of the scale.

A consistency approach to estimate interrater reliability is the most useful approach with data that is continuous in nature.

The third approach concerns measurement estimates. An example of measurement estimates is the degree to which scores contribute to common scoring as opposed to error components. A measurement estimates approach assumes that an assessor is obliged to use all of the information available from all assessors (including discrepant ratings) in order to score for each respondent (Stemler, 2004).

Several factors influence inter-rater reliability. Multiple aspects can make an assessment more reliable, for example (Johnson & Sevingby, 2007):

- Standards increase agreement, but they should be chosen and defined with care (Denner, Salzman, & Harris, 2002; Popp, Ryan, Thompson, & Behrens, 2003);
- Analytical scoring is often preferable, except when separate dimension scores are summarized in the end; then the reliability decreases. The mean score is not always the sum of its component parts (Johnson, Penny, & Gordon, 2000, 2001; Penny, Johnson, & Gordon, 2000a, 2000b);
- Only under restrained conditions, two raters can be enough to produce acceptable levels of agreement (Marzano, 2002);
- Agreement is improved by discussion about interpretation of criteria and training in

assessment. However, training and discussion are not sufficient to eliminate differences (Johnsson & Svingby, 2007);

- Inter-rater reliability can be improved by augmentation of the rating scale (for example that the raters can expand the number of levels using + or − signs), although not for consensus agreements (Penny et al., 2000a, 2000b).
- A two-level scale (for example pass/fail) can be reliably scored with minimal training, whereas a four-level scale is more difficult to use (Williams & Rink, 2003).

2.4.3 Validity judgment of performance assessments

Reliability is not the only critical concept that has to be taken into account during assessment of multidisciplinary courses. Also the extent to which the assessment measures what it intents to measure need to be explored. The assessment should represent the course' learning objectives and content. The task needs to be consistent with the theory and the scoring structure, such as criteria or rubric must follow rationally from the domain structure (Allen & Tanner, 2006; Arter & McTighe, 2001). This requires an exploration of the concept of validity.

Validity concerns the question whether the assessment measures what it intends to measure. Validity in educational research is especially seen as an interpretation of the results (Borsboom et al., 2004; McMillan, 2004). Numerous aspects of validity are investigated and reported in the literature on assessment (Baartman, Bastiaens, Kirschner, & van der Vleuten, 2007). Most common are content and construct validity.

The content aspect is frequently examined. Nowadays, domain coverage is about both traditional content, as it is about thinking processes used during the assessment (Baartman et al., 2008). Within multidisciplinary assessment, the content refers to both the selection of insights as well as articulating the level of understanding in students' work. Therefore, attention has to be paid to the level of these cognitive processes in the assessment (Van de Watering & van der Rijt, 2006).

Construct validity determines content relevance and representativeness of the knowledge and skills revealed by the assessment (Baartman at al., 2007). The concern for how content is represented in assessment is due to the need for results to be generalizable to the construct domain, and not be limited to only the sample of assessed tasks. Construct validity includes evidence of both intended and unintended implications of score interpretation (Baartman et al., 2007).

2.5 Summary

In sum, our literature review shows that by means of multidisciplinary learning the capacity to integrate knowledge and modes of thinking, drawn from more than one discipline to produce a cognitive advancement, is improved. The key in multidisciplinary learning is that discipline integrity is preserved and assessment is clear for each discipline. Disciplinary insights are not in conflict with multidisciplinary learning, rather they embody knowledge (Boix Mansila, 2005). However, assessing the quality of multidisciplinary work is difficult, with regard to exchanging methods, translating categories, and testing outcomes against multiple standards of quality. In order to carefully assess multidisciplinary work, considering disciplinary grounding by a focus on disciplines, the use of knowledge and the level of understanding in students' work is required (Boix Mansila, 2005; Wolfe & Haynes, 2003). In addition, the assessment should be consistent, timely and effective (Nikitina, 2006).

To improve the quality of the assessments, our literature study reveals several requirements:

- The assessment must stand on valid indicators of what counts as accomplished student work (Wiggins, 1998).
- Criteria are crucial; they should establish the level of achievement that is required for a student to pass the course and should be directly related to the course learning outcomes (Boix Mansilla & Duraising, 2007).
- Criteria for multidisciplinary assessment should include validity within and beyond the disciplines (Ackerman, 2000).
- Evidence of learning should be authentic and demonstrate valid learning (Boix Mansilla, 2005).
- Understanding of the criteria is crucial in producing agreement between assessors.
- Clarifying and sharing criteria between assessors and students is required (Boix Mansilla & Duraising, 2007).
- The assessment should provide consistent and constructive feedback on students' progress, processes and results, focusing on how to improve their work.

To improve the intra-rater reliability for assessors, a rubric is suggested. Interrater reliability is improved when clear standards are defined, and interpretation of criteria is discussed. Also training in assessment is emphasized. But reliability is not the only critical concept that has to be taken into account during assessment of multidisciplinary courses. The task needs to be consistent with the theory, and the scoring structure (such as criteria or rubric) must follow rationally from the domain structure.

3. IE Quick Scan course

3.1 Introduction

Now that literature has been explored regarding the concept of multidisciplinary assessment, requirements of validity and reliability, the IE Quick Scan course in the bachelor program Industrial Engineering is considered as a case study. The IE Quick Scan aims to integrate knowledge and modes of thinking, drawn from more multiple disciplines to produce a cognitive advancement, but its assessment consists of several aspects that can be considered problematic from a literary point of view. For example, the assessors are aware of the importance of criteria and standards to evaluate the quality of student work, discussing and clarifying criteria to produce agreement between assessors, are no common practices in the course. Also, assessors vary in consistency and frequency with respect to providing feedback on students' progress, processes and results. To improve the effectiveness, reliability and validity of this multidisciplinary assessment and to gain insight in its boundary conditions, the IE Quick Scan was selected as a case study.

3.2 The Industrial Engineering Quick Scan

In the multidisciplinary course Industrial Engineering Quick Scan (course code 1CK100) students have to apply theories and concepts from five different academic fields and perspectives: Accounting and Finance, Human Performance Management, Information Management, Operations Management, and Innovation Management. The course is part of the coherent elective package "Introduction & Internship Industrial Engineering" in the Bachelor College. The objective of the course is to describe a company's business processes using the academic knowledge acquired during the three-year IE major program. Students engage in an internship, complete a number of preset subassignments, which are listed in the Quick Scan manual, participate in two peer-review group sessions, and write a report. The current study focuses on the assessment of this final written report, which consists of solutions to subassignments from five different IE disciplines (or "aspect systems"). The reports are assessed by staff members with a background in one of five academic fields. Most staff members are specialist in one discipline.

The IE Quick scan consists of a set of models and tools developed from different scientific disciplinary points of view. By applying these models and tools students develop a complete picture of the business setting. The Quick Scan results into a number of deliverables, for example answers to a number of questions from different scientific disciplinary points of view, which support the development of steady-state models for the key business process. Within scope are recommendations concerning possible improvements of current business processes based on reflection on the answers to the disciplinary questions and the steady-state models developed. The different assignments concern the following aspect systems: Accounting and Finance, Human performance management, Information Management, Operations Management and Innovation Management.

The tools used to assess a Quick Scan report consist of an answering model per subassignment (bundled in the Quick Scan teacher manual) and a grading form. Lecturers grade each subassignment as 1 (insufficient), 2 (sufficient) and 3 (good). The Quick Scan teacher manual provides an answer model and all lecturers are expected to carefully follow and apply it.

The assessment of the final document has a 70% weight in the final grade, the other 30% comes from performance in the peer review sessions and the company supervisor. The assessments for each subassignment are entered into an Excel sheet, which calculates the report grade, accounting for the weight of each subassignment. The students subsequently receive their score for the course and, depending on the lecturer, additional feedback is given either face-to-face, via email, or not at all.

Through a careful description of the assignments for different aspects and steady-state models, students are expected to display a sufficient insight into the situation of the organization within the process within scope to formulate possible improvements. Several assignments, formulated for each aspect system, are mandatory (18 in total). Students can increase their scores, through solid descriptions of general deliverables, such as the description of the AS-IS situation within scope in the form of a steady-state models as well as a swimming-lane diagram showing the timing and frequency of execution of business processes and information flows between business processes.
In addition, the scores of elective assignments (7 in total) can be added to the overall grade, under the condition that all mandatory assignments have been carried out. Elective assignments that are graded as insufficient do not contribute to the overall grade.

3.3 Design

Although the IE Quick Scan is an interesting and desirable course, safeguarding its educational quality is a difficult endeavor. Many staff members are specialists in one field of interest and feel ill-equipped to reliably grade the students' work in all parts of such integrative projects. To investigate the extent of the problem in this course, we take a two-step approach. First, we aim to get an in-depth view of lecturers' and students' perceptions of the course regarding its pedagogy and learning goals, and the validity and reliability of its multidisciplinary assessments. We thus conduct student and lecturer interviews as a qualitative grounding of the Quick Scan's standards, criteria and assessment processes. Second, we aim to empirically substantiate the (in)appropriateness of the current assessment procedure of the IE Quick Scan report in terms of validity and reliability, and to come up with potential suggestions for an alternative design. We thus setup an experimental design in which different lecturers assess a set of three reports.

4. Empirics: Interview Study

4.1 Method

4.1.1 Participants

In the autumn of 2015, participants in our study were interviewed, consisting of both assessors and students. First, we invited students based on their involvement in the course during the academic year 2014-2015, to discuss their perception of the course. To reflect the general picture across students, we invited students who delivered a wide range of quality work, varying between insufficient to excellent grading scores. Their reports were assessed by lecturers representing one of the five different academic fields (Accounting and Finance, Human Performance Management, Information Management, Operations Management, and Innovation Management). Twelve students responded positively to our request and participated in in-depth interviews.

To provide further context for our interviews, we invited several lecturers involved in the course in the academic year 2014/2015 or 2015/2016 to share their experiences and ideas. Our sampling strategy was to have a balanced sample of lecturers within Innovation, Technology Entrepreneurship & Marketing (ITEM) group, Operations, Planning, Accounting and Control (OPAC), Information systems (IS), Human Performances management (HPM). Ten lecturers (8 assistant and 2 associate professors) accepted our invitation and were interviewed.

4.1.2 Semi-structured interviews

Data collection consisted of semi-structured interviews with participants through face-to-face-interviews, videoconferencing, or telephone interviews. The interviews with students and assessors covered aspects of pedagogy and learning, and validity and reliability of multidisciplinary course assessments. Questions were based on the quality-criteria of assessment and our literature study (van Eggen, 2009; Evers, et al., 2010; Gerritsen-van Leeuwenkamp, 2012, Linn, 1990; Messick, 1995; Tillema, Leenknecht, & Segers, 2011; Van Berkel et al., 2014) (see Appendices B and C). Interviewees' perspectives on the use of knowledge and the level of understanding required within the IE Quickscan were part of the interviews. Also, reliability and validity of the assessment, and assessors' certainty during evaluation were discussed. In addition, we questioned the consistency, efficiency and effectiveness of the assessment. All interviews were conducted by the same interviewer and were approximately one hour in length. All information was treated strictly confidentially.

4.1.3 Data-analysis and procedure

Qualitative analysis of our interview data was used to elaborate on the third research question. The interviews were transcribed and clustered based on experiences, recommendations and obstacles in assessing or following the course, in relation to procedure, answering models, clarity amongst evaluation criteria, final judgments and feedback.

4.2 Results

4.2.1 Students' experiences

Overall, students found the IE Quick Scan to add value to their body of knowledge. Most students experienced multiple benefits of executing the project, for instance with regard to personal development, gaining practical experiences, and broadening their horizon. Students indicated that a certain amount of feedback (in addition to the overall grade) was given either via email, or face-to-face communication, depending on the assessor. Students noted differences between assessors, with regard to both quality and quantity of feedback. Sometimes feedback focused solely on the assessors' domain of expertise, but in general all aspect system assignments and the overall process were discussed.

However, students also reflected on the Quick Scan's negative points. The written reports consist of answers to the aspect system subassignments and not every assignment seemed to fit every company, requiring flexibility and creativity in solutions. Some of the subassignments were perceived as unclear, restrictive, or even infeasible, for instance when the company was unwilling to share required information. As a consequence the results did not always match the requirements, resulting in lower scores.

Also the unclear or even unknown criteria affected students' perceptions of the course. Students felt that assessment criteria should be objective and transparent. However, large differences between assessors appeared in this respect, with some assessors sharing and clarifying criteria either prior, during or only after finishing of the course. Finally, several students indicated that the distinction between grades should be more pronounced. It seems to be somewhat unclear, based on what grounds or criteria certain assignments were considered insufficient, sufficient or good. Clarification in this respect was requested. A summary is provided in table 1.

Table 1. Students' perspective

Top positive	*Negative*
• The IE Quick Scan adds value to students' body of knowledge; • The IE Quick Scan offers possibilities for personal development and gaining practical experiences; • The IE Quick Scan broadens students' horizon;	• A lack of clarity, restrictiveness, or even infeasibility of subassignments; • Unclear or even unknown criteria: more objective and transparent criteria are requested; • Sharing and clarifying criteria either prior, during or after finishing of the course largely depends on the various assessors; • Distinctions between grades are perceived somewhat unclear.

4.2.2 Lecturers’ opinions and experiences

The interviews revealed that the assessors differed in their assessment approach. Although prescribed and expected, an analytic approach was not always applied in the assessment of the Quick Scan report. Several lecturers seemed to prefer a holistic scoring approach, evaluating the student based on their overall impression of their efforts rather than the more detailed grading per subassignment. General opinions of students, past performances, quality of the subassignments within assessors’ expertise, or steady states could form the basis of the holistic assessment approach.

Similarly to the students’ responses, interviews with the lecturers revealed differences in sharing and clarifying preset grading criteria with their students. Criteria should be used as starting point, but students often were unaware of the interpretation of criteria.

To feel confident about scoring student work with accuracy, several assessors delved into topics both inside and outside their area of expertise. Only then lecturers felt confident to communicate grades to and share feedback with students. The communication of grades was rarely accompanied with arguments. Depending on their assessor, students were able to discuss the results to get a better insight into what elements they should improve.

A disadvantage of the current grading forms, mentioned by several lecturers, is the limitation in grading opportunities. Per subassignment only three grading options were provided (i.e., insufficient, sufficient, and good). Two of these options correspond to desirable grades; typically, a lecturer had to be really sure to score a subassignment as “insufficient”. This resulted in a limited variance in grades per subassignment as well as overall result.

Also opinions on the answer models were mixed. Some lecturers stated that the answer models for various assignments were usable. Others signaled the need for more elaboration and flexibility in grading the subassignments. An answer model should describe as transparantly and clearly as possible to what extent the student meets the criteria on each subassignment. However, for some of the aspect systems, the current answer models lack clarity and transparency. This affects scoring accuracy negatively.

Interpretation of criteria seemed to vary among lecturers, resulting in disagreement on the overall quality of the report. Some lecturers stressed the need to be trained in assessment. Several lecturers suggested an exercise to calibrate their findings. In addition, all lecturers interviewed reported feeling overwhelmed by an enormous workload of the current course. The guidance of students during their internship is doable, and can be efficiently organized. However, the assessment of the report is extremely time-consuming (approximately 7 hours per report), as are the administrative requirements.

Lecturer’s perceptions on workload and feasibility seem to have affected satisfaction with the grading process.

Overall, in agreement with the students, most lecturers found the IE Quick Scan valuable with several points of improvement. See table 2 for an overview of lecturers’ perspectives.

Table 2. Lecturers’ perspective

Top positive	*Negative*
• The IE Quick Scan is perceived valuable;	• Sharing and clarifying preset grading criteria with their students depends on the assessor;
• Assessors delved into topics both inside and outside their area of expertise to improve their guidance and the quality of their assessment;	• Assessment approaches vary among the assessors of the course;
• Assessors provide feedback, either oral or written, to improve students’ learning.	• Limitation in grading opportunities;
	• Answer models vary in usability;
	• Disagreement on the overall quality of the report, due to varying interpretation of criteria;
	• Workload of the current course, combined with other educational or research activities.

5. Empirics: Field experiment

We set up a field experiment to empirically substantiate the (in)appropriateness of the current assessment procedure of the IE Quick Scan report in terms of validity and reliability, and to come up with potential suggestions for an alternative design. The experiment focuses on the assessments of three reports written by students in the academic year 2014/2015.

5.1 Method

5.1.1 Participants

We invited several lecturers (a total of 12), representing the five capacity groups, to participate in our experiment. Despite the large monetary incentive offered, only four assessors were willing to participate to evaluate three written reports. Each assessor represents either one of the capacity groups (IS, ITEM, OPAC & HPM). Hence, each assessor has his/her expertise in a specific domain.

5.1.2 Experimental design

From a total set of approximately 70 Quick Scan reports, three reports were selected that were considered to be representative of the possible variation in reports. Selection criteria included a spread in grades and the extent to which the theoretical frameworks underlying the multidisciplinary assignments fitted the company context in which the project was conducted, both in terms of applicability of the frameworks and the degree to which the company could “help” the student by providing off-the-shelf analyses.

The assessment of three internship project reports was conducted by the four assessors. For each report, every assessor independently assessed the whole report, including the module (i.e., set of subassignments) for which he/she was considered an expert. We also asked assessors to indicate how confident (or: certain) they were on every grade provided for a subassignment. The expert-assessments of modules were used to establish a “golden standard”, assuming that an expert assessing his/her “own” module provides the most accurate and confident grade for this module. For the three reports, both the overall grades and the grades for each assignment by four assessors were then compared to the grade provided by the original assessor and the golden standard. This allows to draw conclusions of the accuracy of the grading and confidence of judgment.

The assessors received instructions regarding the assessment of the report. They were asked to thoroughly follow the answer model from the teacher manual as they evaluated successively every assignment. The results were filled in the Excel grading sheet. Importantly, the sheet did not calculate the final overall grade for the report for two reasons. First, the overall grade also contains information on a student’s performance in the peer review sessions and the company supervisor’s assessment.

Those were not available to our assessors. Second, providing the overall grade may have stimulated our assessors to take a holistic approach and complete the Excel sheet after establishing an overall grade. This was deemed undesirable.

5.1.3 Data- procedure and analysis

The data collection took place in a period of four weeks in March 2016. Data were explored by investigating the overall average scores of the three reports. In addition, means and standard deviations were explored based on scores by the original assessor, assessors in our experiment, and the golden standard. In addition, the assessors' confidence or certainty in each of their assessments was added. We also addressed to what extent assessors' scores differed from the golden standard.

5.2 Results

When we look closely at the results of the intervention, we can identify differences between assessors with regard to scoring, certainty of scoring, and overall results. Table 1 represents the overall grades for the reports in the experiment. The results show differences between overall grades provided by the four assessors on the three reports. Although there were hardly differences for report 1 (absolute range: 0.4), three of four assessors provided a substantially lower grade than the original assessor (difference >1.1) for report 2, and a very wide range of grades (5.9-9.4) for report 3. Although most assessors seem to agree on the quality differences between three reports, the grades appear to diverge.

Table 1. Overall report grades in experiment

	Report 1	Report 2	Report 3
Assessor 1	5.9	7.5	9.4
Assessor 2	5.7	6.5	8.9
Assessor 3	5.7	6.3	5.9
Assessor 4	5.5	6.4	8.2
Original	5.5	7.6	8.4

Table 2 provides the mean, maximum, and minimum values for different scores provided by the four assessors across the three reports. This table also shows a comparison of assigned grades in comparison to either the expert, original assessor or participants of this experiment.

The mean score of the grades suggest consensus over the quality of reports. This is not supported by the results on minimum and maximum scores, which reveal differences between assessors. Differences occur due to the assigned scores per subassignment, which reveal disagreement between the expert assessment, original assessment and assessors. This indicates that an overall agreement on quality work does not always reflect on quality work within the subassignment. The possibility that the amount of subassignments cover for disagreement between assessors must be taken into account.

Table 2. Means, deviations and certainty-scores of assessors, expert and original assessor of three reports

	Grade[1]	Certainty[2]	Expert-original[3]	Expert-assessors[4]	Assessors-original[5]	SD-assessors[6]
Mean	1.68	3.67	0.02	0.14	-0.12	0.45
Maximum	3	5	2.33	1.44	1,.89	1.28
Minimum	0	0.67	-2	-1.44	-1.89	0

[1] Grade per subassignment (weighted; scale: 1 to 3)
[2] Certainty of grade per subassignment (scale: 1 to 5)
[3] Difference between expert assessment and original assessment per subassignment (original scale: 1 to 3)
[4] Difference between expert assessment and the three other assessors per subassignment (original scale: 1 to 3)
[5] Difference between four assessors and the original assessment per subassignment (original scale: 1 to 3)
[6] Standard deviation between four assessors per subassignment

Next, Tables 3, 4 and 5 break down the overall grades from Table 1 per module (i.e., a bundle of subassignments in one specific area of expertise), per assessor for Report 1, 2, and 3 respectively.

Table 3. Results of report 1

Module	Golden Standard	Assessor 1 (Certainty)	Assessor 2 (Certainty)	Assessor 3 (Certainty)	Assessor 4 (Certainty)	Original Assessor
IS	2.2	2.1 (3.3)	2.1 (3.0)	2.1 (4.0)	*2.2 (5.0)*	2.3
HPM	1.9	1.6 (3.5)	*1.9 (4.5)*	1.6 (4.3)	1.7 (4.3)	1.7
OPAC	1.2	1.2 (3.1)	1.1 (3.8)	*1.2 (4.7)*	1.3 (4.0)	1.0
ITEM	1.3	*1.3 (4.8)*	1.5 (4.0)	1.3 (4.2)	1.5 (4.2)	1.7
Total	5.8	5.9	5.7	5.7	5.5	5.5

Note: Italic indicates that assessor provided golden standard for specific module.

Table 4. Results of report 2

Module	Golden Standard	Assessor 1 (Certainty)	Assessor 2 (Certainty)	Assessor 3 (Certainty)	Assessor 4 (Certainty)	Original Assessor
IS	1.8	2.3 (3.3)	2.7 (2.8)	2.4 (5.0)	1.8 *(4.5)*	1.8
HPM	2.0	2.1 (3.5)	2.0 *(4.3)*	1.6 (5.0)	1.8 (3.3)	2.0
OPAC	1.3	1.7 (1.9)	1.4 (2.3)	1.3 *(4.7)*	1.6 (2.8)	1.9
ITEM	1.3	1.3 *(4.2)*	1.4 (3.7)	1.3 (5.0)	1.4 (4.0)	1.6
Total	6	7.5	6.2	6.3	6.4	7.6

Note: Italic indicates that assessor provided golden standard for specific module.

Table 5. Results of report 3

Module	Golden Standard	Assessor 1 (Certainty)	Assessor 2 (Certainty)	Assessor 3 (Certainty)	Assessor 4 (Certainty)	Original Assessor
IS	1.9	2.1 (3.0)	2.3 (2.2)	2.1 (4.0)	1.9 *(3.2)*	2.1
HPM	2.3	2.5 (3.8)	2.3 *(4.5)*	2.0 (4.8)	2.2 (2.8)	2.5
OPAC	1.5	2.4 (2.4)	2.2 (2.8)	1.5 *(4.6)*	1.9 (2.3)	2.0
ITEM	2.1	2.1 *(4.0)*	2.3 (4.3)	1 (4.3)	2.5 (2.8)	2.0
Total	7.2	9.4	8.9	5.9	8.2	8.4

Note: Italic indicates that assessor provided golden standard for specific module.

The results of each report suggest that the consensus over a lower quality report (see Table 3) is more easily accomplished, than agreeing over mediocre or even excellent reports. The ranges of grading increases (see Table 4 and 5) as the reports' quality improve.

The results also indicate that differences occur between the scoring of modules. Grading the subassignments requires interpretations; often the model answer provided in the teacher manual is not sufficient to come to a grade with certainty. In addition, it seems that having expertise makes an assessor more strict in grading. This is most clearly evidenced by Table 4, which shows that the golden standard overall score is lower than all overall scores of the assessors. This effect is less pronounced for the lower quality report (Table 3) and the higher quality report (Table 5), although the latter is due to a very low score of assessor 3 on the ITEM module. It seems that when assessors have difficulty in grading the quality of parts of the students' work because it is outside the area of expertise, assessors become more risk averse and prefer to grade too high rather than too low. Indeed, during interviews lecturers mentioned that they try to avoid questions concerning certain domains by adapting their grading upward.

Although we find empirical support for our above contention, it is not overwhelming. Table 6 indicates for each module what percentage of subassignments were scored lower than, equal to, or higher than the expert assessor (i.e., golden standard) for that module. Assessors' scores are often comparable to the golden standard, are lower in 18% of the cases, and higher in 30% of the cases.

Table 6. Comparable scores from three assessors versus expert-scores

Module	Lower	Equal	Higher	Missing
IS	16%	37%	46%	1%
HPM	27%	50%	23%	
OPAC	16%	48%	32%	6%
ITEM	13%	67%	17%	3%
Total	18%	51%	30%	3%

We also addressed the certainty during the assessment of a multidisciplinary assignment in Tables 3, 4 and 5. The assessors are more certain about their scoring within their own area of expertise as compared to modules outside their area of expertise (see Table 6). Most likely each assessor is more flexible in interpreting answers of assignment within their field of expertise, except for assessor 3.

Table 7. Overall means of certainty

Module	Assessor 1 Certainty	Assessor 2 Certainty	Assessor 3 Certainty	Assessor 4 Certainty
IS	3.1	2.4	4.3	*3.6*
HPM	3.7	*4.4*	4.8	2.9
OPAC	2.2	2.6	*4.6*	2.5
ITEM	*4.1*	4.1	4.5	3.2

Note: Italic indicates certainty of assessor who provided golden standard for specific module.

Feedback on the various assignments needs to be based upon knowledge or expertise in the field and the ability to apply it in a multidisciplinary assignment. Answer models for HPM and ITEM are more extensive than answer models of OPAC and IS, providing for instance specific examples or explicit instruction to grading in a small rubric. This is reflected in the scores of certainty. An extensive answer model can be helpful to fill the gap between expertise, scoring and provided feedback. It seems as the elaboration of answers influences the scores of certainty.

6. Discussion

6.1 Most important findings

Assessing multidisciplinary student work is challenging. Students have to apply theories and concepts from different academic disciplines, but are generally assessed by specialists in certain fields, not generalists. Through a literature study and a field experiment, this study attempts to contribute toward recommendations for lecturers and program managers that want to apply multidisciplinary assessment in their courses.

Within the fuzzy concept of multidisciplinary learning, standards from different disciplines are the starting point for assessing a multidisciplinary approach. Integrates occurs through the lens of a common theme (Drake, 2007). The IE major program has found that common lens in the Quick Scan, which represents theories and standards from several disciplines.

Our literature study offers several suggestions to improve the assessment of the Quickscan and increase its reliability and validity. It is emphasized to clarify and share assessment criteria and standards between faculty and students. According to Ben-David (2000) the understanding of the criteria involved are crucial in producing agreement between assessors and crucial to provide an accurate evaluation of the student's overall proficiency. According to our interviews, several lecturers are unaware of the importance to share and clarify assessment criteria at the start of the course. This implicates that training in assessment might be useful, to enhance assessors with knowledge and skills with regard to multidisciplinary assessment.

In determining the grade, the lecturer preferably has an advantage over the students regarding superior knowledge and extensive experience. The interviewees stipulated the advantages of superior knowledge and extensive experience in a particular domain, in relation to the quality of feedback they have given to their students. However, several lecturers mentioned to feel less certain assessing subassignments outside their domain of expertise. According to Van Berkel (2012) assessors need to be supported in assessing the quality of a task. Although several lecturers seek support in improving their knowledge on domains outside their expertise through delving into research, they also express the need for extensive answer models and the possibility to consult colleagues. Therefore, elaborating on answer models can be helpful. Also, assessors should be facilitated to calibrate on the assessment of the Quickscan.

To support the assessor in assessing the quality of a task, assessment models are often used in which the quality of a product is assessed based on criteria, standards and rating scales (Van Berkel, 2012). Given that an assessment form is used, criteria, standards and rating scales are provided. However, according to our interviews, the answer models do not always clearly indicate what type of answer should be graded “insufficient”, “sufficient”, or “good”. This implicates that a more extensive grading guideline needs to be included.

Because learning and instruction are increasingly competence-based, holistic scoring to adequately determine competence is emphasized (Van der Vleuten & Schuwirth, 2005; Oosterheert, Eldik, & Kral, 2007; Van Berkel, 2012). However, many courses currently apply analytic judgment, such as the IE Quick Scan. The Quick Scan requires raters to assign scores to each of the criteria assessed in the product, after which the final score is summarized by all separate scores. The criteria act as primary reference points against which student submissions are judged, and then serve as the basis for the communication and feedback process. Analytic scoring is well applicable in this difficult task. However, when knowledge, skills and attitudes need to be assessed both separately and integrative, holistic scoring might be preferred with regard to grading explicit and implicit expertise. Further discussion on the main goals and the integrative grounding of this course are required to determine whether analytic judgment is emphasized.

In multidisciplinary assessment variability in the assessment scores can appear. Besides disagreement caused by differences in experience, it has been reported that factors such as lecturers’ attitudes regarding students and content influence the rating of students’ work (Davidson, Howell, & Hoekema, 2000). Also in the IE Quick Scan the major threat to reliability is the lack of consistency of the individual assessors. Therefore, developing assessment skills is necessary to improve the quality of individual assessors and to decrease inconsistency among them.

Multidisciplinary assessments tend to be more effective when consistent and constructive feedback on students’ progress, processes, results and limitations is provided. Students report a large variety in the quality of feedback in the IE Quick Scan. These results converge with findings that lecturers differ in the frequency and amount of feedback they tend to give on students’ progress, processes and results. A common understanding on qualitative feedback is necessary, to ensure that each student receives consistent and constructive feedback on their learning process.

In an ideal situation, the assessment and results are independent of those who score. Our experiment revealed difficulty in accomplishing homogeneity and reliability. Multiple assessors to assess each report can enhance quality assurance. Our empirical study shows that assessment by experts prevents inflation of grades. The possibility of assigning assessors to assess specific subassigments within their domain of expertise can improve reliability.

6.2 Limitations and future research

Results of the present study are conditional upon certain choices in the research design and procedure. Subsequently, some practical and methodological limitations can be addressed. First, due to time and resource constraints, our experiment was small in nature. Four assessors graded three reports, which may not be representative for a large amount of students and lecturers involved in this course. This limitation should be taken into account when considering the extent to which the results can be generalized.

Second, although the outcomes of the experiment are clear, this study was post-hoc and only focused on the assessment of the written report, which accounts for 70% of the final grade. A study with a longer time frame may investigate grading processes during the course's quartile and include reflections on the peer review sessions.

Third, although the Quick Scan course is described as integrative, the integration of knowledge and perspectives are not explicitly assessed. Considering the complex interdisciplinary skills such as communication, critical thinking and problem-solving that are required in authentic context, further research could assessment grading in a more interdisciplinary course, where reflection and integration of knowledge is expected.

Despite these limitations, our study provides promising results for educators who wish to enhance a multidisciplinary approach and assessment.

6.3 Managerial implications and recommendations

The IE Quick Scan should be supportive, flexible and responsive to both students' and lecturers' needs and priorities. We suggest the following recommendations for the IE Quick Scan;

- The IE Quick Scan seems to be a promising integrating platform through which students can consider all aspects of their professional roles and responsibilities as engineers in a multidisciplinary project, but a more flexible approach to assessing the quality of the final product is recommended. Due to a potential misfit between the description of some subassignments and the company context, answer models are not always suitable and students are not rewarded for their creativity or flexibility of their solution. Evaluation of flexibility and creativity can be included in the grading form.
- The IE Quick Scan can improve clarity about criteria and indicators of quality. Prior to, during and after the course, lecturers need to be aware of the assessment criteria, including validity for, within, and beyond the disciplines. Students need to be informed, beforehand, about the criteria and standards and the links to specific learning outcomes and objectives. Assessment

criteria should establish the level of achievement that is required for a student to pass the course and should be directly related to the course learning outcomes.

- Although the IE Quick Scan is multidisciplinary and thus features theories from several disciplines, there is little integration. It is suggested to enhance this integrative component to promote students' cognitive advancement. For example by reflecting on the integrative component of the course prior to the assessment, enhancing interdisciplinary learning by rich questioning, and adding interdisciplinary elements such as the level of interdisciplinary grounding into the assessment procedure.
- An assessor may indicate where a student's answer was inconsistent with the answer model for a specific element in the report, but is unlikely to provide more examples or details of what a perfect answer should have looked like, due to a lack of expertise in certain domains. More elaborate answer models, or adjustment of some of the assignments can help avoid that assessors evade student's questions about their assessment.
- Similarly, the interviews suggest that some lecturers/assessors feel ill equipped to establish meaningful assessments across multiple disciplines, due to the limited quality of the answer models or a lack of expertise in certain fields. It is imperative that answer models are improved and more clearly indicate what type of answer should be graded "insufficient", "sufficient", or "good". Currently there are too many answer models that provide an example or just provide some theoretical background, but do not include a grading guideline. A rubric can be used to support the assessments.
- To improve the quality of feedback by assessors, during and after the course, a meeting can be organized to present and discuss good and bad practices. Also, the rationale behind multidisciplinary assessment and the process and steps of assessment can be explained. In addition, continuity, as to which assessors are enrolled in the course, is emphasized.
- The written assignment holds for 70% of the score. In addition, at two stages during the project, students need to prepare and make an oral presentation, and participate in a peer group-review, where their own view(s) are subjected to constructive peer group criticism. The assessor judges students' efforts in the peer reviewed sessions. There are many ways in which peer review can contribute to the value of the course. Effective peer review can contribute to learning, for example when it is used as formative assessment. Peer review can be used during courses, according to Van Gennip et al. (2009), both when the assessors and assessees have equal or unequal ability. However, experience with peer review need to be built by means of training and an exercise round before the peer review actually accounted for grades (Thurlings. De Jong & Beijaard, 2015). It seems worthwhile to evaluate the impact of peer review on student capabilities. Depending on the wishes and needs, a training module in peer review can be added to the course.

- Our study showed challenges in accomplishing homogeneity and reliability. Multiple assessors or assessment by experts in certain domains can be considered to not only prevent inflation of grades, but also to improve the quality of assessments.
- Assessors also indicated to have too little time to assess the final reports, thus being unable to fulfill responsibilities. In addition, the high workload (balancing teaching activity with the high demands on research activity) leaves the assessors feeling subject to conflicting demands. It is strongly advised to provide assessors with more time in their schedules to supervise and assess Quick Scan students.

7 References

Ackerman, D.B. (2000). Intellectual and practical criteria for successful curriculum integration. In H.H. Jacobs (ed.), *Interdisciplinary curriculum: Design and implementation* (pp. 25–37). Alexandria, VA: ASCD.

Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, *102*(4), 287-302.

Allen, D., & Tanner, K. (2006). Approaches to Biology Teaching and Learning Rubrics: Tools for Making Learning Goals and Evaluation Criteria Explicit for Both Teachers and Learners. *CBE Life Sciences Education*, *5*, 197–203.

Arter, J., & McTighe, J. (2001). Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance. Corwin Press.

Baartman, L. (2008). "*Assessing the assessment*": Development and use of quality criteria for Competence Assessment Programmes. http://igitur- archive.library.uu.nl/dissertations/2008-0423-200627/UUindex.html

Ben-David, M. F. (2000). The role of assessment in expanding professional horizons. *Medical Teacher*, *22*(5), 472-477.

Boix Mansilla, V. (2005). Assessing student work at disciplinary crossroads. *Change Magazine, 37*, 1, 14-21.

Boix Mansilla, V. & Duraising, E. D. (2007). Targeted assessment of students' interdisciplinary work: An empirically grounded framework proposed. *The Journal of Higher Education, 78*(2), 215–237.

Borja, A., Tueros, I., Belzunce, M.J., Galparsoro, I., Garmendia, J.M., Revilla, M., Solaun, O., Valencia, V., (2008). Investigative monitoring within the European Water Framework Directive: a coastal blast furnace slag disposal, as an example. *Journal of Environmental Monitoring 10*, 453–462.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061–1071.

Brown, G., Bull, J., & Pendlebury, M. (1997). Peer and self-assessment. *Assessing student learning in higher education. London: Routledge*, 170-84.

Brown, G. T. L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a new zealand writing assessment system. *Assessing Writing*, *9*, 105–121.

Brough, C. J. (2012). Implementing the democratic principles and practices of student-centred curriculum integration in primary schools. *The Curriculum Journal, 23*, 345-369.

Black, P. (2000). Research and the development of educational assessment. *Oxford Review of Education*, *26*(3-4), 407-419.

Daly, K., Brown, G. & McGowan, C. (2012). *Curriculum integration in the International Baccalaureate Middle Years Programme: Literature review.* Report prepared for the International Baccalaureate organization.

Davidson, M., Howell, K. W., & Hoekema, P. (2000). Effects of ethnicity and violent content on rubric scores in writing samples. *The Journal of Educational Research*, *93*(6), 367-373.

Denner, P. R., Salzman, S. A., & Harris, L. B. (2002). Teacher work sample assessment: An accountability method that moves beyond teacher testing to the impact of teacher performance on student learning. In *Paper presented at the annual meeting of the American Association of Colleges for Teacher Education*.

Dowden, T. (2007). Relevant, challenging, integrative and exploratory curriculum design: Perspectives from theory and practice for middle level schooling in Australia. *The Australian Educational Researcher, 34*, 51-71.

Drake, S. M. (2007). *Creating Standards-based integrated curriculum*: aligning curriculum, content, assessment and instruction (2nd ed.). Thousand Oaks, CA: Corwin.

Drake, S. M. (2012). Creating Standards-based integrated curriculum: Common Core State Standards Edition (3nd ed.). Thousand Oaks, CA: Corwin.

Drake, S. M., Savage, M. J., Reid, J. L., Bernard, M. L., & Beres, J. (2015). *An Exploration of the Policy and Practice of Transdisciplinarity in the IB PYP Programme,* Research Report Brock University.

Expertgroep Protocol. (2014). *Beoordelen is mensenwerk*. Den Haag. www.vereniginghogescholen.nl

Evers, A., Lucassen, W., Meijer, R., Sijtsma, K. (2010). *Cotan beoordelingssysteem voor de kwaliteit van tests* (gewijzigde herdruk). Amsterdam: NIP/COTAN.

Field, M., & Stowe, D. (2002). Transforming interdisciplinary teaching and learning through assessment. In C. Haynes (Ed.), *Innovations in interdisciplinary teaching* (pp. 256-74). Westport, CT: American Council on Education Oryx Press.

Fogarty, R. & Pete, B.M. (2009). *How to integrate curricula* (3rd edition). Thousand Oaks, CA: Corwin.

Gerritsen-Van Leeuwenkamp, K. (2012). Het relatieve belang van vijftig kwaliteitskenmerken van toetsing voor studenttevredenheid in het hoger onderwijs. Masterscriptie. Heerlen: Open Universiteit.

Hoare, A., Cornell, S., Bertram, C., Gallagher, K. , Heslop, S., Lieven, N., MacLeod, C., Morgan, J., Pickering, A., Wells, S. & Willmore, C. (2008) Teaching against the grain: multi-disciplinary teamwork effectively delivers a successful undergraduate unit in sustainable development, *Environmental Education Research, 14*(4), 469-481.

Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education*, *13*, 121–138.

Johnson, R. L., Penny, J., & Gordon, B. (2001). Score resolution and the interrater reliabilityof holistic scores in rating essays. *Written Communication*, *18*, 229–249.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, *2*(2), 130-144.

Joughin, G. (2009). *Assessment, Learning and Judgement in Higher Education,* Australia: University of Wollongong.

Kaufman, D., Moss, D. M., & Osborn, T. A. (2003). *Beyond the boundaries: A transdisciplinary approach to learning and teaching*. Praeger Publishers.

Klein, J. T. (2000). A conceptual vocabulary of interdisciplinary science. In P. Weingart & N. Stehr (Eds.), *Practising interdisciplinarity* (pp. 3-24). Toronto: University of Toronto Press.

Klein, J. T. (2002). Assessing interdisciplinary learning K-16. In J. T. Klein (Ed.), *Inter- disciplinary education in K-12 and college* (pp. 179-96). New York: College Board Publications.

Kuhlemeier, H. (2002). *Beoordelingsschalen in praktijktoetsen : hoe ontwikkel en gebruik je ze?* (pp. 1–29). http://toetswijzer.kennisnet.nl/html/praktijktoetsen/praktijktoetsen3.pdf

Lattuca, L. R. (2001). Creating interdisciplinarity: Interdisciplinary research and teaching among college and university faculty. Nashville, TN: Vanderbilt University Press.

Lattuca, L. R., Voigt, L. J., & Fath, K. Q. (2004). Does interdisciplinarity promote learning? Theoretical support and researchable questions. *The Review of Higher Education*, *28*(1), 23-48.

Linn, R. L. (1990). Admissions testing: Recommended uses, validty, differential prediction, and coaching. *Applied Measurement in Education, 3*(4), 297-318.

Marzano, R. J. (2002). A comparison of selected methods of scoring classroom assessments. *Applied Measurement in Education*

McMillan, J. H. (2004). *Educational research: Fundamentals for the consumer*. Boston: Pearson Education Inc.

Meijers, A. W. M., & Brok, D. P. (2013). Engineers for the future: an essay on education at TU/e in 2030. Eindhoven.

Mertler, C. A. (2004). Designing Scoring Rubrics for Your Classroom. *Practical Assessment, Research & Evaluation*, *7*(25), 1–9.

Messick, S. (1995). Validity of psychological assessment. Validations of inferences from persons' responses and performances as scientific inquiry into scoring meaning. *American Psychologist, 50*, 741-749.

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical assessment, research & evaluation*, *7*(10), 71-81.

Moss, D.M., Osborn, T.A. & Kaufman, D. (Eds., 2008; 2012). *Interdisciplinary Education in the Age of Assessment.* New York: Routledge.

Newell, W. (2001). A theory of interdisciplinary studies. *Issues in Integrative Studies, 19*, 1-25.

Newell, W. (2002). Integrating the college curriculum. In J. T. Klein (Ed.), *Interdisciplinary education in K-12 and college* (pp. 119-37). New York: College Board Publications.

Nikitina, S. (2006). Three strategies for interdisciplinary teaching: Contextualizing, conceptualizing, and problem-centring. *Journal of Curriculum Studies, 38*(3), 251–271.

Oosterheert, I., Eldik, S. van, & Kral, M. (2007). *Het digitaal portfolio als instrument voor summatieve competentiebeoordeling* (pp. 0–34). Nijmegen: SURF / Hogeschool Arnhem Nijmegen. https://www.han.nl/onderzoek/kennismaken/kenniscentrum-kwaliteit- van-leren/lectoraat/leren-met- ict/publicaties/_attachments/gehele_rapport_dpf_instrument_scb.pdf

Penny, J., Johnson, R. L., & Gordon, B. (2000a). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing*, *7*, 143–164.

Penny, J., Johnson, R. L., & Gordon, B. (2000b). Using rating augmentation to expand the scale of an analytic rubric. *Journal of Experimental Education*, *68*, 269–287.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *ESRC methods programme*, *15*(1), 047-71.

Popp, S. E. O., Ryan, J. M., Thompson, M. S., & Behrens, J. T. (2003). Operationalizing the rubric: The effect of benchmark selection on the assessed quality of writing. In *Paper Presented at Annual Meeting of the American Educational Research Association*.

Ramos, K. D., Schafer, S., & Tracz, S. M. (2003). Validation of the fresno test of competence in evidence based medicine. *British Medical Journal*, *326*, 319–321.

Richards, J. C., & Bennett, S. M. (2011). Supporting upper elementary students' content literacy through a transdisciplinary framework: Crossing disciplinary boundaries in a summer camp. *Journal of Reading Education, 36*(3), 47-51.

Russell-Bowie, D. (2009). Syntegration or disintegration? Models of integrating the arts across the curriculum. *International Journal of Education & the Arts, 10*(28). 1-23.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education, *Assessment & Evaluation in Higher Education, 30*(2), 175-194.

Schilling, K. L. (2001). Interdisciplinary assessment for interdisciplinary programs. In B. L. Smith & J. McCann (Eds.), *Reinventing ourselves: Interdisciplinary education, collaborative learning and experimentation in higher education* (pp. 344-54). Bolton, MA: Anker.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, *9*(4), 1-19.

Stowe, D., & Eder, D. (2002). Interdisciplinary program assessment. *Issues in Integrative Studies, 20*, 77-101.

Thurlings, M., De Jong, Y., & Beijaard, D. (2015). *Report Peer-review at TU/e*. Project commissioned by the TU/e Bachelor College, Eindhoven.

Tillema, H., Leenknecht, M. & Segers, M. (2011). Assessing assessment quality; Criteria for quality assurance in design of (peer) assessment for learning – a review of research studies. *Studies in Educational Evaluation, 37*, 25-34.

Van Berkel, A. (2012). Kritische reflectie op competentietoetsen in het hbo. *Onderwijsinnovatie*, *2*, 17–26.

Van Berkel, H., Bax, A., Joosten-ten Brinke, D. (2014). Het toetsproces ontleed. In H. Van Berkel, A. Bax, & D. Joosten-ten Brinke (Red.), *Toetsen in het hoger onderwijs* (pp. 1-11) Houten: Bohn Stafleu van Loghum.

Van Eggen, T. J. H. M. (2009). *De kwaliteit van toetsen*. Inaugurele rede. Enschede: Universiteit Twente.

van Gennip, N. A., Segers, M. S., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, *4*(1), 41-54.

Van de Watering, G., & van der Rijt, J. (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. *Educational Research Review*, *1*(2), 133-147.

Van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: from methods to programmes. *Medical Education*, *39*, 309–317.

Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance* (Vol. 1). San Francisco, CA: Jossey-Bass.

Williams, L., & Rink, J. (2003). Teacher competency using observational scoring rubrics. *Journal of Teaching in Physical Education*, *22*, 552–572.

Wolfe, C. R., & Haynes, C. (2003a). Assessing interdisciplinary writing. *Peer Review, 6*(1), 12-15.

Wolfe, C. R., & Haynes, C. (2003b). Interdisciplinary writing assessment profiles. *Issues in Integrative Studies, 21*, 126-69.

Appendix A Summary of Key Criteria and Guiding Assessment Questions
(Source: Boix Mansila & Daws Duraising, 2007)

I. Disciplinary grounding:

Guiding Questions

Are the selected disciplines appropriate to inform the issue at hand? Are any key perspectives or disciplinary insights missing?

Are the considered disciplinary theories, examples, findings, methods, and forms of communication accurately employed, or does the work exhibit misconceptions?

II. Advancement through integration

Guiding Questions

Where is there evidence of disciplinary integration (e.g., conceptual framework, graphic representation, model, leading metaphor, complex explanation, or solution to a problem)?

Is there evidence that understanding has been enriched by the integration of different disciplinary insights?

III. Critical awareness

Guiding Questions

Does the work show a clear sense of purpose, framing the issue in ways that invite a multidisciplinary approach?

Is there evidence of reflectiveness about the choices, opportunities, compromises, and limitations involved in multidisciplinary work and about the limitations of the work as a whole?

Appendix B Guidelines for Lecturer Interviews

Gebaseerd op o.a. de kwaliteitscriteria van toetsen (van Eggen, 2009; Evers, et al., 2010; Gerritsen-van Leeuwenkamp, 2012, Linn, 1990; Messick, 1995; Tillema, Leenknecht, & Segers, 2011; Van Berkel et al., 2014).

Validiteit;

1. Meet deze toets wat het beoogd te meten?
2. Is de toets betekenisvol en bruikbaar?
3. In welke mate zijn de conclusies die uit de toets-scores worden getrokken gerechtvaardigd?

Betrouwbaarheid:

4. Is er sprake van accuratesse en consistentie van standaarden?
5. In welke mate is er sprake van helderheid en overeenstemming qua interpretatie van de geformuleerde criteria?
6. Is er sprake van accuratesse en consistentie van standaarden, criteria en besluiten tussen beoordelaars?

Bruikbaarheid:

7. Kunt u zich vinden in de bruikbaarheid van de toets?

Objectiviteit:

8. In welke mate oefent de beoordelaar invloed uit op de hoogte van de score?
9. In welke mate oefent de beoordelaar invloed uit op de uiteindelijke beoordeling?

Motiveerbaarheid:

10. Zijn de toetsinhoud en de resultaten op de toets inhoudelijk beargumenteerd?
11. Zijn de toetsinhoud en de resultaten op de toets gerechtvaardigd tegenover de student?

Transparantie:

12. Wat vindt u van de duidelijkheid en begrijpelijkheid van de toets en bijbehorende procedures voor de betrokken deelnemers?

Feedback:

13. In hoeverre geeft de toets de student persoonlijke feedback over de sterke punten en verbeterpunten?
14. In hoeverre formuleert de docent persoonlijke feedback over de sterke punten en verbeterpunten?

Ondersteuning:

15. Wat is de gemiddelde termijn die u nodig heeft om de toets te beoordelen?
16. Voldoet het aantal uur dat beschikbaar is voor de beoordeling van deze toets?
17. Wat is de termijn waaarop de student antwoord ontvangt op inhoudelijke en procedurele vragen

over toetsen?

Realiseerbaarheid:

18. Zijn de doelen en de criteria van de toets realiseerbaar voor de student wat betreft competentie niveau?

Beoordelaar's tevredenheid en vertrouwen:

19. In hoeverre bent u tevreden over de huidige inrichting van de toets?

20. In hoeverre heeft u vertrouwen over de te communiceren scores naar de student binnen uw persoonlijke expertise/ domein?

21. In hoeverre heeft u vertrouwen over de te communiceren scores naar de student buiten uw persoonlijke expertise/ domein?

22. In hoeverre bent u in staat feedback te geven over inhouden buiten uw persoonlijke expertise?

23. In hoeverre reflecteert de door u gegeven waardering voor het eindproduct de algehele waardering van het rapport?

Appendix C Guidelines for Student Interviews

Gebaseerd op o.a. de kwaliteitscriteria van toetsen (van Eggen, 2009; Evers, et al., 2010; Gerritsen-van Leeuwenkamp, 2012, Linn, 1990; Messick, 1995; Tillema, Leenknecht, & Segers, 2011; Van Berkel et al., 2014).

Betreffende de validiteit;

1. Meet deze toets wat het beoogd te meten?
2. Is de toets betekenisvol en bruikbaar?
3. In welke mate zijn de conclusies die uit de toets-scores worden getrokken gerechtvaardigd?

Betreffende de betrouwbaarheid:

4. Is er sprake van accuratesse en consistentie van standaarden?
5. In welke mate is er sprake van helderheid en overeenstemming qua interpretatie van de geformuleerde criteria?
6. Is er sprake van accuratesse en consistentie van standaarden, criteria en besluiten tussen beoordelaars?

Betreffende de bruikbaarheid:

7. Kunt u zich vinden in de bruikbaarheid van de toets?

Betreffende de motiveerbaarheid:

8. Zijn de toetsinhoud en de resultaten op de toets inhoudelijk beargumenteerd?
9. Zijn de toetsinhoud en de resultaten op de toets gerechtvaardigd tegenover de student?

Ten aanzien van transparantie:

10. Wat vindt u van de duidelijkheid en begrijpelijkheid van de toets en bijbehorende procedures voor de betrokken deelnemers?

Betreffende feedback:

11. In hoeverre geeft de toets de student persoonlijke feedback over de sterke punten en verbeterpunten?
12. In hoeverre heeft de student van de docent persoonlijke feedback over de sterke punten en verbeterpunten ontvangen?

Ten aanzien van de ondersteuning:

13. Wat is de termijn waarop de student antwoord heeft ontvangen op inhoudelijke en procedurele vragen over toetsen?

Ten aanzien van de realiseerbaarheid:

14. Zijn de doelen en de criteria van de toets realiseerbaar voor de student wat betreft competentie niveau?

Ten aanzien van student tevredenheid en vertrouwen:

15. In hoeverre is het doel van de toets duidelijk?

16. In hoeverre vereist de toets reflectie op multidisciplinaire inzichten?

17. In hoeverre vereist de toets zelfreflectie en zelf-evaluatie op eigen inzichten en leren?

18. In hoeverre beschikt de student over voldoende inhoudelijke basis wat bijdraagt aan multidisciplinaire inzichten?

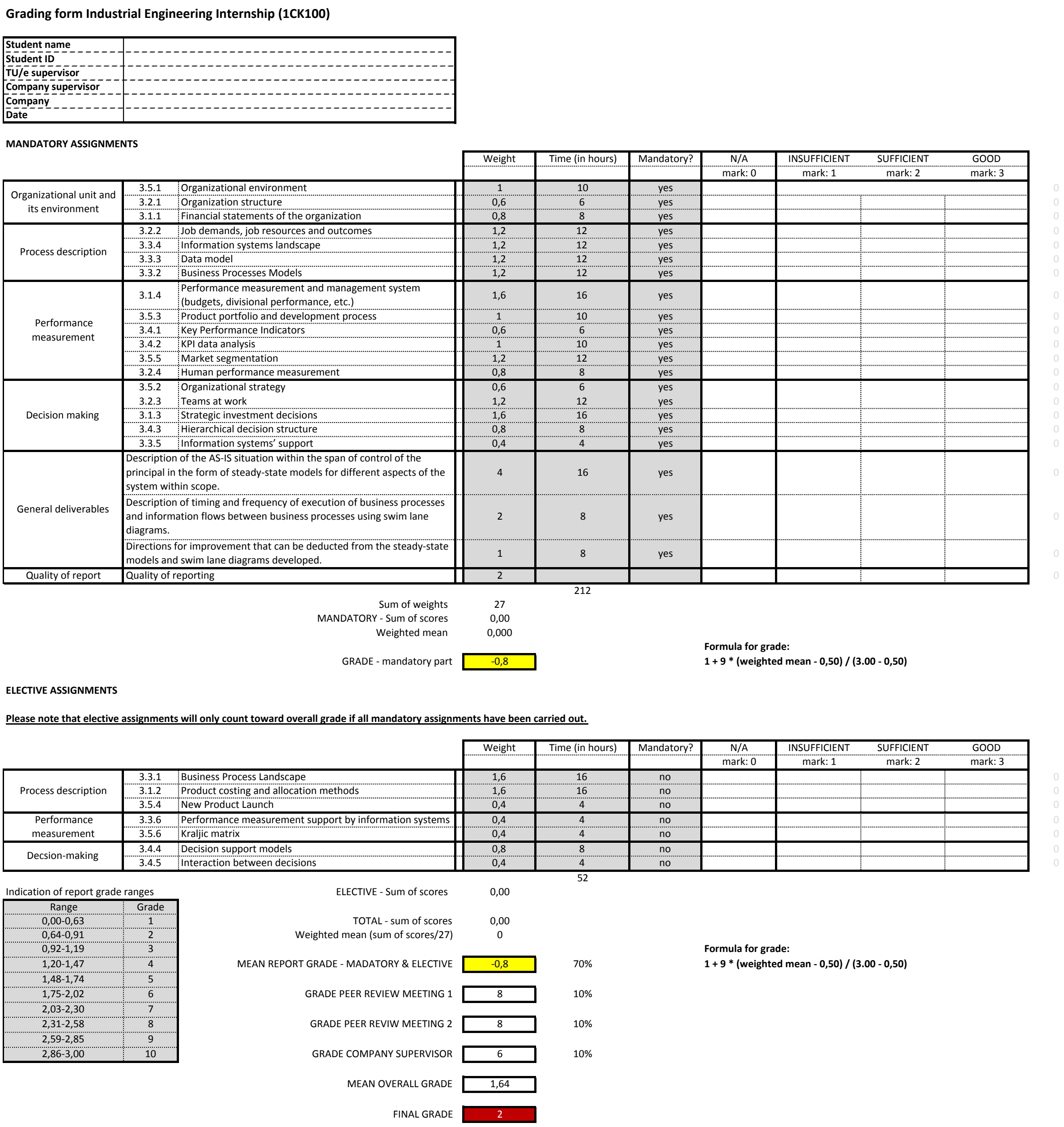
19. In hoeverre levert de toets inzichten op in de samenhang tussen verschillende disciplines?

20. In hoeverre vraagt het eindproduct om het herdefiniëren van problemen, uitwisselen van methodes, testresultaten en kwaliteitsstandaarden?

21. In hoeverre is de toets duidelijk over te behalen doelen, te beoordelen aspecten?

22. In hoeverre is de toets duidelijk over het proces van integratie en de daaropvolgende waardering?

Appendix D Excel grading sheet



Grading form Industrial Engineering Internship (1CK100)

Student name	
Student ID	
TU/e supervisor	
Company supervisor	
Company	
Date	

MANDATORY ASSIGNMENTS

			Weight	Time (in hours)	Mandatory?	N/A	INSUFFICIENT	SUFFICIENT	GOOD	
						mark: 0	mark: 1	mark: 2	mark: 3	
Organizational unit and its environment	3.5.1	Organizational environment	1	10	yes					0
	3.2.1	Organization structure	0,6	6	yes					0
	3.1.1	Financial statements of the organization	0,8	8	yes					0
Process description	3.2.2	Job demands, job resources and outcomes	1,2	12	yes					0
	3.3.4	Information systems landscape	1,2	12	yes					0
	3.3.3	Data model	1,2	12	yes					0
	3.3.2	Business Processes Models	1,2	12	yes					0
Performance measurement	3.1.4	Performance measurement and management system (budgets, divisional performance, etc.)	1,6	16	yes					0
	3.5.3	Product portfolio and development process	1	10	yes					0
	3.4.1	Key Performance Indicators	0,6	6	yes					0
	3.4.2	KPI data analysis	1	10	yes					0
	3.5.5	Market segmentation	1,2	12	yes					0
	3.2.4	Human performance measurement	0,8	8	yes					0
Decision making	3.5.2	Organizational strategy	0,6	6	yes					0
	3.2.3	Teams at work	1,2	12	yes					0
	3.1.3	Strategic investment decisions	1,6	16	yes					0
	3.4.3	Hierarchical decision structure	0,8	8	yes					0
	3.3.5	Information systems' support	0,4	4	yes					0
General deliverables	Description of the AS-IS situation within the span of control of the principal in the form of steady-state models for different aspects of the system within scope.		4	16	yes					0
	Description of timing and frequency of execution of business processes and information flows between business processes using swim lane diagrams.		2	8	yes					0
	Directions for improvement that can be deducted from the steady-state models and swim lane diagrams developed.		1	8	yes					0
Quality of report	Quality of reporting		2							0
				212						

Sum of weights	27
MANDATORY - Sum of scores	0,00
Weighted mean	0,000
GRADE - mandatory part	-0,8

Formula for grade:
1 + 9 * (weighted mean - 0,50) / (3.00 - 0,50)

ELECTIVE ASSIGNMENTS

Please note that elective assignments will only count toward overall grade if all mandatory assignments have been carried out.

			Weight	Time (in hours)	Mandatory?	N/A	INSUFFICIENT	SUFFICIENT	GOOD	
						mark: 0	mark: 1	mark: 2	mark: 3	
Process description	3.3.1	Business Process Landscape	1,6	16	no					0
	3.1.2	Product costing and allocation methods	1,6	16	no					0
	3.5.4	New Product Launch	0,4	4	no					0
Performance measurement	3.3.6	Performance measurement support by information systems	0,4	4	no					0
	3.5.6	Kraljic matrix	0,4	4	no					0
Decsion-making	3.4.4	Decision support models	0,8	8	no					0
	3.4.5	Interaction between decisions	0,4	4	no					0
				52						

Indication of report grade ranges

Range	Grade
0,00-0,63	1
0,64-0,91	2
0,92-1,19	3
1,20-1,47	4
1,48-1,74	5
1,75-2,02	6
2,03-2,30	7
2,31-2,58	8
2,59-2,85	9
2,86-3,00	10

ELECTIVE - Sum of scores	0,00	
TOTAL - sum of scores	0,00	
Weighted mean (sum of scores/27)	0	
MEAN REPORT GRADE - MADATORY & ELECTIVE	-0,8	70%
GRADE PEER REVIEW MEETING 1	8	10%
GRADE PEER REVIW MEETING 2	8	10%
GRADE COMPANY SUPERVISOR	6	10%
MEAN OVERALL GRADE	1,64	
FINAL GRADE	2	

Formula for grade:
1 + 9 * (weighted mean - 0,50) / (3.00 - 0,50)