Decoding Feedback:
Improving feedback practices for students
in introductory programming courses

Claudia Ott

A thesis submitted for the degree of
Doctor of Philosophy

at the University of Otago, Dunedin,
New Zealand

June 2014
Abstract

High failure rates in introductory programming courses testify that learning to program is challenging for many students. This problem is widely acknowledged in the field of computer science education research. In higher education, quality feedback from teachers is regarded as one of the main contributors to improve student learning. Feedback to support students’ development into self-regulated learners, who set their own goals, self-monitor their actual performance according to these goals and adjust learning strategies if necessary, is seen as an especially important aspect of contemporary feedback practice.

For this research project the notion of feedback, as described within the higher education literature, was reviewed and provides a theoretical basis to address the question of how to support first year programming students. A framework to consider effective feedback was developed and used to assess feedback practice in the context of introductory programming courses. Various opportunities to integrate research from computer science education to support feedback processes were revealed. A gap in the research became apparent when searching for ways to support students’ self-regulated learning and we realised that only those students who are informed about course demands and the impact of certain study behaviours on their final achievement are in a position to self-regulate their learning on an informed basis.

As a next step we analysed the predictive value of naturally occurring course data for students’ final performance at different stages in our course. All data sets, drawn from the records of 387 students enrolled in 2011 and 2012, correlated significantly with students’ final examination results and we were able to define risk factors as well as performance indicators. Based on those findings an infographic for students was developed communicating course characteristics and projected final performance for different achievement levels at various stages of the course.

To explore the impact of feedback on self-regulation level as well as to learn about students’ attitudes towards diagnostic course data in general, a scoping study was conducted in 2013. Over 200 students were supplied with the infographic. The results from the study suggest that students valued the information, but, despite high engagement with the information, students’ study behaviour and learning outcome remained rather unaffected for the aspects investigated. Given these multi-layered results, we suggest further exploration on the provision of feedback based on diagnostic course data – a vital step towards more transparency for students to foster their active role in the learning process.
Acknowledgements

A PhD thesis, although labelled with one main author, is always a joint venture and often a kind of joint “adventure” where supervisors guide the candidate through unknown territory taking care of his or her development and wellbeing as a researcher, but also - in an ideal case - discover something new along the way. Whereas I am not sure about the latter, I definitely approve the former: Anthony Robins was an invaluable source of inspiration and knowledge on my entire way through the field of computer science education and Kerry Shephard the vital counterpart on my expeditions into higher education. Anthony’s and Kerry’s wealth of experience and their continuous support and their trust, that I would find my own way, made this journey a privilege to be on. Thank you.

The development of the infographic required the involvement of several more people: I like to thank the interview participants for their valuable feedback before and after the rollout; my supervisor for this part of the project Russell Butson for his guidance, enthusiasm and the lively discussions; the design duo Kohlhaas and Kohlhaas for their critical revisions of the infographic; Hanna, Holger and Arthur for trialling the infographic at early stages; Alicia Berry for testing the interview; and all the demonstrators from COMP160 for giving the infographic a go before the rollout. My special thanks goes to Sandy Garner for being endlessly supportive over the entire course of this project and in particular for getting all necessary changes into the lab book at very short notice, but especially for being such a good friend.

I also want to thank all the students taking COMP160 in 2011, 2012 and 2013 for their participation in the online survey. Great response rates of over 90 per cent (!) resulted in insightful data consisting of over 10,000 single responses (!), which made this project so stimulating. One of the most enjoyable parts of the project, however, have been the 300+ hours of teaching in the labs thanks to our delightful students.

Financially this project was supported by a University of Otago scholarship. I am grateful for the opportunity to be part to the University Otago postgraduate community and I want to thank all the people contributing to the inspiring and supportive atmosphere. Additionally I want to acknowledge the financial support of the Computer Science Department as well as the Higher Education Development Centre for partly funding the completion of a postgraduate certificate in Higher Education.

Finally, Holger and Hanna, you are so much fun to be with and I am proud to call you my family. Thank you for your cheerfulness, wisdom and love.
# Table of Contents

## 1 Introduction

1.1 Personal Background 3
1.2 Motivation 3
1.3 Contributions and structure of the thesis 5

## 2 Feedback in Higher Education

2.1 Introduction 9
2.2 Principles of good feedback practice 10
  2.2.1 Focus on self-regulation 11
  2.2.2 Focus on teacher involvement 12
  2.2.3 Focus on peer-assessment 12
  2.2.4 Focus on engagement 13
  2.2.5 Focus on dialog 14
  2.2.6 Focus on students’ personality and their perceptions of feedback 14
2.3 A framework for effective feedback 15
2.4 Summary 17

## 3 Feedback in Computer Science Education

3.1 Introduction 21
3.2 Feedback on Task Level 21
  3.2.1 Practical tasks during tutorial time or as extended assignments 22
  3.2.2 Lectures 24
  3.2.3 Mid-semester and final examination 25
  3.2.4 Summary 27
3.3 Feedback on Process Level 28
  3.3.1 Programming process 29
  3.3.2 Course related processes 31
  3.3.3 Summary 32
3.4 Feedback on Self-Regulation Level 34
  3.4.1 Multi-stage assignments: Increasing task complexity with fewer instructions 34
  3.4.2 Student engagement through peer assessment and peer feedback 35
  3.4.3 Opportunities for self-assessment 36
  3.4.4 Self-monitoring course performance 37
  3.4.5 Summary 38
3.5 Summary 39
# 4 Investigating the predictive value of available course data

## 4.1 Introduction

## 4.2 Predictive factors in the literature

## 4.3 Performance indicators in COMP160

### 4.3.1 Indicator: Pre-course grades

### 4.3.2 Indicator: Attendance of lab sessions

### 4.3.3 Indicator: Number of submitted lab tasks

### 4.3.4 Indicator: Time of submission

### 4.3.5 Indicator: Time spent in labs

### 4.3.6 Indicator: Mid-semester exam result

## 4.4 Discussion of risk factors

## 4.5 Implications for the feedback process

## 4.6 Summary

# 5 Communicating course characteristics and performance indicators

## 5.1 Introduction

## 5.2 Why Infographics?

## 5.3 Infographic development

### 5.3.1 Section “Key Events”

### 5.3.2 Section “What students say …”

### 5.3.3 Section “What we say …”

### 5.3.4 Section “What the numbers say …”

## 5.4 Evaluation method

### 5.4.1 Participants

### 5.4.2 Procedure

### 5.4.3 Analysis

## 5.5 Findings and infographic redesign

### 5.5.1 General impression, Navigation and Structure

### 5.5.2 Section “Key Events”: Providing a quick course overview

### 5.5.3 Section “What students say …”: Sharing former students’ perception

### 5.5.4 Section “What we say …”: Sharing teachers’ perspectives

### 5.5.5 Section “What the numbers say …”: Explaining statistical observations

### 5.5.6 Debriefing: General impression and improvements of the infographic

## 5.6 Summary
6 Planning and conducting a scoping study to support students’ self-regulated learning

6.1 Introduction

6.2 Stages of students’ engagement in the feedback process

6.3 Research Methodology

6.4 Units of analysis (UA)

6.4.1 Students’ course performance data (UA 1)

6.4.2 Students’ self-reported confidence development (UA 2)

6.4.3 Students’ responses to the Infographic related tasks (UA 3)

6.4.4 Students’ responses to a questionnaire (UA 4)

6.4.5 Students’ responses in an interview (UA 5)

6.5 Summary

7 Evaluating stages of engagement in the feedback process

7.1 Introduction

7.2 Stage 1 & 2: Collection and immediate attention

7.3 Stage 3: Initial cognitive response

7.4 Stage 4: Prolonged attention

7.5 Stage 5: Immediate or latent reaction

7.5.1 Changes in students’ perceptions and expectations: Self-reported confidence data

7.5.2 Changes in students’ study behaviour: Naturally occurring course data

7.5.3 Students attitudes towards diagnostic course data

7.5.4 Discussion

7.6 Stage 6: Outcome

7.6.1 Assessing variations in student cohorts based on pre course results

7.6.2 Assessing mid-semester and final examination scores

7.6.3 Assessing the difference between mid-semester and final examination scores

7.6.4 Assessing performance indicators

7.6.5 Assessing risk factors

7.6.6 Discussion

7.7 Summary
## Contributions and Future Work

### 8.1 Introduction

### 8.2 Contributions – A critical reflection

- **8.2.1** A model to guide effective feedback 162
- **8.2.2** A case study of predictive factors for students’ final achievement 165
- **8.2.3** A scoping study supporting students’ self-regulated learning 168

### 8.3 Future Work

- **8.3.1** Assisting the development into self-regulated learners 172
- **8.3.3** Improving diversity and personal relevance of feedback 173
- **8.3.4** Improving student engagement with feedback through interactivity 174
- **8.3.5** Improving evaluation methodology by the use of mixed methods 175
- **8.3.6** Opening the black box – our responsibility for more transparency 176

### 8.4 Closing remarks

### References

### Appendix A – Infographic Example I (Charles Joseph Minard)

### Appendix B – Infographic Example II (Florence Nightingale)

### Appendix C – Infographic (initial version)

### Appendix D – Infographic (final version)

### Appendix E – Questionnaire

### Appendix F – Interview Protocol

### Appendix G – Mid-Semester Feedback Sheet
# List of Tables

2.1: Questions guiding effective feedback practices on task level, process level and self-regulation level based on Hattie and Timperley (2007)  

3.1: Feedback on task level  
3.2: Feedback on process level  
3.3: Feedback on self-regulation level  
4.1: Predictions for 2011  
4.2: Predictions for 2012  

7.1: Questionnaire results (UA 4) – Questions evaluating the clarity of the infographic  
7.2: Questionnaire results (UA 4) – Questions evaluating frequency of use of the infographic  
7.3: Interview results (UA 5) – Students reporting on how they used the “Numbers” section  
7.4: T-test statistics for number of attended lab sessions  
7.5: T-test statistic for year-to-year comparison of average time spend on a completed task (UA 1/C)  
7.6: T-test statistic for year-to-year comparison of number and timeliness of submissions (UA 1/B)  
7.7: T-test statistic for achievement groups in a year-to-year comparison of number and timeliness of submissions (UA 1/B)  
7.8: Groups of students dropping the course in 2011, 2012 and 2013  
7.9: Questionnaire results (UA 4) – Questions evaluating relevance and perceived usefulness of the infographic  
7.10: Questionnaire results (UA 4) – regarding diagnostic course data in general  
7.11: T-test statistic for examination scores in 2011 and 2012  
7.12: T-test statistic for examination scores in 2011 and 2013  
7.13: T-test statistic for exam scores in 2011 and 2013 for groups of higher and lower achieving students  
7.14: Correlation and T-test statistic for mid-semester and final exam in 2011 to 2013  
7.15: Comparison of “COMP150 final mark” performance indicators  
7.16: Comparison of “Number of Labs by week 5” performance indicators  
7.17: Comparison of “Mid-Semester exam result” performance indicators  
7.18: Comparison of “Lateness average by week 8” performance indicators  
7.19: Comparison of “Number of Labs by week 11” performance indicators  
7.20: Assessment of risk factors for 2011 to 2013  
7.21: Predictions for 2013
# List of Figures

5.1:  Section “Key events”  57
5.2:  Section “What students say …”  58
5.3:  Section “What we say …”  59
5.4:  Section “What the numbers say …”  61
5.5:  Analytic Induction (Gibbs, 2010) modified for infographic design evaluation  64
5.6:  Section “Key events” revised  66
5.7:  Section “What students say …” revised  67
5.8:  Section “What we say …” revised  68
5.9:  Section “What the numbers say …” revised  70
6.1:  The temporal dimension of student engagement with assessment feedback after Price et al. (2011)  77
6.2:  Stages of student engagement with the infographic based on Price et al. (2011)  78
6.3:  Sample questionnaire item addressing understanding of performance indicators  78
6.4:  Two questionnaire items addressing prolonged attention  79
6.5:  Units of analyses (UA), related data sets and aspect covered  81
6.6:  Methods of analyses employed  81
6.7:  Frequency of average time spend per submitted task  86
6.8:  Online survey – four statements for each submission  88
6.9:  Online survey – one additional statement depending on the task number  89
6.10: Proportions of students responding to the survey in 2011, 2012 and 2013  90
6.11: Infographic related tasks, printed on the back of the infographic  92
6.12: Stages in the engagement process and related units & methods of analysis  97
7.1:  Overview of evaluation of Stage 1 & Stage 2  102
7.2:  Overview of evaluation of Stage 3  104
7.3:  Proportions of correct, incorrect and missing answers to infographic related questions based on a sample of 36 task solutions.  105
7.4:  Overview of evaluation of Stage 4  108
7.5:  Overview of evaluation of Stage 5  112
7.6:  Infographic detail indicating task difficulty  117
7.7:  Weekly attendance (UA 1/A) for students completing the course in 2011, 2012 & 2013  121
7.8:  Frequency of average time spend per submitted task (UA 1/C) of students completing the course in 2011, 2012 & 2013  123
7.9:  Percent of electronically submitted tasks (UA 1/B) per week in 2011, 2012 & 2013  124
7.10: Proportion of students not sitting the final exam after completing the mid-semester exam based on their mid-semester score (>=50% and <50%) 127
7.11: Overview of evaluation of the learning outcome 138
7.12: Results for the different stages in the engagement process and outcome 155
“Huge numbers of papers have been published in computing education conferences and journals in the past 40 years, so we would expect much to have been learnt about teaching and learning in computing. However, after decades of research, we still have only a vague understanding of why it is so difficult for many students to learn programming, the basis of the discipline, and consequently of how it should be taught.”

(Malmi et al., 2010, p. 9)
1.1 Personal Background

With almost no teaching experience I wrote a proposal for a computer science education research project. Why? I guess the answer goes back 25 years when I was in my first year studying computer science in Weimar (Germany). As a trained main-frame computer operator, learning the programming language Pascal was easy as pie and I would never have thought that the same course caused substantial difficulties and major frustration for others students if my boyfriend back then hadn’t been one of those. In my memory we spent months going over the basics of programming again and again – without any considerable progress. He just couldn’t see the big picture; there was nothing he could connect the new ideas with. Suddenly, one day, I literally watched how everything in his mind fell into place. I might have not been a great teacher in terms of passing on my enthusiasm for the discipline, because, even though he earned some money by programming as a student, he never pursued a computer science career whereas I had found what I wanted to do for a living.

After working as a freelance computer programmer for many years, I came across Anthony Robins’ (2010) publication, postulating the Learning Edge Momentum as an explanation why grade distributions in introductory programming courses are typically bi-modally distributed. Rather than accounting students’ aptitude or talent, he suggests that the integrative nature of the course material leads to a spiralling effect as we learn on the edge of what we already know. If these edges are fragmented or non-existent, learning is difficult and becomes more and more difficult as the number of missing links increases. In contrast, if connections can be established, they form solid edges for the learning of higher-level concepts. Reading Anthony’s publication got me hooked, because it so much reflected my own, very first, “teaching experiences” and I wanted to learn more about the way people learn computer programming.

1.2 Motivation

Learning the first programming language is hard (e.g. Pears et al., 2007) and introductory programming courses (CS1) are infamous for high failure rates (Dehnadi and Bornat, 2006). At the Computer Science department at Otago University around 30% fail COMP160 (“General Programming”), a typical CS1 course that is offered as a first year programming course and requires no prerequisites. In 2011 the subject area “Computer Science” was amongst the eight subject areas with the highest failure rates out of 84 subject areas listed (University of Otago, 2012). Some other universities report extreme failure rates of 60% (Bennedsen and Caspersen, 2007). However it is also known, that a considerable group of students has no difficulties in learning how to program. Approximately 30% of the students pass COMP160 with A-grades. Those bi-modal grade distributions make CS1
courses an interesting research field, primarily investigating the question of how to support students in acquiring the skills and knowledge to master their first programming language.

In higher education, quality feedback is considered as a main contributor to improve student learning. Hattie (2009) ranked feedback amongst the top 10 influences on students’ achievement of the 138 influences he investigated based on a total of over 800 meta-reviews. Contemporary principles of good feedback practice acknowledge feedback not only as information provided in a detailed and timely manner to the students, but also as a complex process, incorporating students’ engagement and other influences on students' actions and behaviour. Feedback, which targets students’ development into self-regulated learners, as opposed to feedback related to a specific task or the processes necessary to solve this task, is regarded as “sustainable” feedback (Hounsell, 2007; Carless et al., 2011) and to have long lasting effects beyond university study (Nicol and Macfarlane-Dick, 2006).

The computer science education (CSEd) literature reflects the emphasis on feedback. Numerous studies investigate possibilities to support students working on a programming task. For example, matters of error handling are addressed by integrated development environments (IDEs) consisting of compliers and debuggers especially designed to improve feedback for novices in the field of programming (Pears et al., 2007). Pair programming, as a specific form of peer-feedback, is a well established research domain (Salleh, Mendes and Grundy, 2011), so is the provision of learning resources for self-assessment with instant feedback on the task (e.g. Denny et al., 2011) just to name a few. This structured investigation of the CSEd revealed good coverage of feedback targeting programming tasks and processes, but supporting students’ self-regulation on course level was not found to be a research topic.

Whereas solving programming problems is in the core of an introductory programming course and feedback on this level is vital, managing course demands is seen in COMP160 as another challenge students are faced with. Especially working consistently and staying on top of the material needs to be considered as a main factor influencing students’ failure or success in the course. Only students who practiced and understood concepts involved can manage the increasing level of complexity. Once a student falls behind, catching up becomes increasingly harder. Driven by such a momentum, two groups of students, as reflected by the bi-modal grade distribution, develop: (1) students who manage all levels and end up with a solid understanding of the material and (2) students who were not able to proceed with the learning because necessary concepts were not fully understood. The theory of the Learning Edge Momentum (Robins, 2010) emphasises such learning mechanisms over matters of talent or a “natural” ability as an explanation what influences students’ success or failure in CS1 courses.
Close investigation of naturally occurring course performance data from 2011 and 2012, involving 386 COMP160 students, demonstrated that our students’ lab task submissions, pre-course grades and mid-semester examination results are predictive for their final outcome. To support self-regulated learning, data about students’ performance need to be made available and interpreted to be turned into relevant feedback information. The integration of performance data into the feedback process to make students aware how actual study behaviour influences the course outcome, was implemented as a scoping study for COMP160 in second semester in 2013 involving roughly 200 students. In order to assess the effectiveness of the feedback intervention various units of analysis were investigated, not only to evaluate the effect on students’ study behaviour and learning outcome, but also to learn about students’ perception regarding feedback on self-regulation level in general.

1.3 Contributions and structure of the thesis

There are various contributions of this thesis, ranging from establishing a theoretical basis to consider effective feedback over an adaptation of the notion of feedback for the CS1 context to conducting an explorative study and the discussion of concrete results.

**Contribution 1 – The development of a model to guide effective feedback**

In search of an answer to the question *What constitutes effective feedback?* a framework to consider feedback practices was derived from the higher education literature. Questions guiding the assessment or planning of effective feedback practices are established for different stages and levels of the feedback process and presented in Chapter 2.

This framework was used to investigate the question of *How to translate effective feedback practice into the CS1 context?* within the practical scope of COMP160, which is considered as a typical CS1 course. Many opportunities for feedback were discussed and links to related research in CSEd were established. The consideration of feedback in the CS1 context and links into related research are presented in Chapter 3 and may inspire CS1 instructors and teachers to critically review their own feedback practices.

**Contribution 2 – The definition of predictive factors for students’ final achievement**

During this exercise we identified a lack of feedback regarding course-related processes. Especially when it comes to students’ self-regulation, we need to ask on what basis our students are able to make informed decisions about performance goals and how can they judge their own performance realistically. Consequently two of the questions addressed in Chapter 4 are *What performance indicators are available?* and *What indicates risky or successful study behaviour?* In fact we found significant correlations between available course data and students’ final performance in COMP160, which indicate later success or failure.
Here again, given the common structure of our course, we hope to initiate similar explorations for other CS1 courses.

**Contribution 3 – An explorative study to support students’ self-regulated learning**

Addressing the question of *How to effectively communicate course characteristics and performance indicators?* resulted in the development of a visual representation (infographic) to illustrate course characteristics and to indicate students’ performance in the final exam based on their actual performance at different stages of the course. The rationale, design aims, evaluation process and the final version of the infographic are documented in Chapter 5 and can be considered as a concrete example for packaging complex course data into an approachable and engaging format as one artefact delivered to all students.

The planning and conducting of a scoping study and the roll-out of the infographic to over 200 COMP160 students, were guided by the question of *What stages are important in the process of engagement with feedback?* and is the focus of the first part of Chapter 6. The chapter also addresses the question of *What units of analysis are available to investigate the degree of engagement at the different stages?* and various data sets, as collected to evaluate the impact of the infographic on students’ perceptions, study behaviour and learning outcome, are introduced. Chapter 6 not only describes the methodology and limitations of our own specific scoping study, but it also advances on the temporal dimensions of student engagement with feedback, as stated by Price, Handley and Millar (2011), as a valuable model to consider the effectiveness of a feedback intervention where students’ learning outcome is only one aspect to be evaluated.

The results of the scoping study are discussed in Chapter 7 by addressing the question of *To what degree did students engage with the feedback provided?* for the different stages of the engagement process as derived from Price et al. (2011). Despite a high degree of engagement with the infographic throughout the course we found students’ final results mainly unaffected by this feedback. This is surprising as students reported that the infographic had a supportive and motivational impact on their learning. In fact most students responding to the questionnaire agreed with the statement that diagnostic information should be shared with students and 95% of students considered such information as helpful in some respect to self-regulate their learning. These positive responses are encouraging results and we suggest the provision of diagnostic course data as a vital part of common feedback practices to guide students’ learning efforts.

In Chapter 8 the different contributions of this project are summarised and discussed. As feedback on self-regulation level is highly regarded in higher education and was well received by our CS1 students, but did not improve students’ actual learning outcome, we need to ask how future work can address potential issues with feedback on this level.
“Although a frequently used term, feedback does not have clarity of meaning. It is a generic term which disguises multiple purposes which are often not explicitly acknowledged.”

(Price et al., 2010, p. 278)

This chapter is based on sections 1 and 2 of a paper submitted to ACM Transactions on Computing Education that is currently under revision:

Ott, C., Robins, A., & Shephard, K. (2014). Translating principles of effective feedback for students into the CS1 context. Manuscript accepted for publication.
2.1 Introduction

Computer science education is an established research field but has been criticised in the past for being too focussed on disciplinary aspects at course level, rather than on higher-level pedagogy in terms of goals and the role of the teacher (Kinnunen et al., 2010). Randolph et al.’s (2008) review found an alarming number of studies either not to involve human subjects at all, or base their findings on anecdotal evidence only. Publications with a theoretical education focus were found to be rare (Joy et al., 2009) and a large number of studies were missing some kind of developed theory, model, framework or instrument (Malmi et al., 2010). Addressing Mark Guzdial’s comment that “Too much of the research in computing education ignores the hundreds of years of education, cognitive science, and learning sciences research that have gone before us” (in Alstrum et al., 2005) this chapter is intended to provide a theoretical basis to consider effective feedback as derived from higher education research. In Chapter 3 those effective feedback practices are considered in the CS1 context by pointing to topics and current developments in CSEd research.

As already mentioned in the introduction, quality feedback is regarded as one of the main contributors to improved student learning (e.g. Hattie & Timperly, 2007; Carless, 2006; Hounsell, 2003; Ramsden, 2003; Askew and Lodge, 2000). There is an immense body of research in higher education investigating the effects of feedback on student learning. Based on over 800 meta-reviews, Hattie (2009) ranked “Feedback” as the 10th strongest influence on students’ achievement based on effect sizes of 138 influences he investigated, outperforming influences such as “Problem-solving teaching” or “Mastery learning”.

In higher education, feedback from teachers to students is mainly discussed in the context of assessment with a focus on formative assessments, but also increasingly in the light of self-regulated learning and self-assessment. Over the last 50 years the perspective on feedback shifted from a behavioural approach, using for example programmed instruction with immediate feedback to shape learners’ behaviour with regards to desired responses, towards a constructivist one (e.g. Burke and Pieterick, 2010). This constructivist approach to learning emphasises the role of the learner - knowledge needs to be actively constructed and is based on prior experiences and beliefs (e.g. Biggs and Tang, 2007; Prosser and Trigwell, 1999). Therefore feedback needs to be investigated not only from an information perspective (What information is necessary at which stage of the learning process?) but also from a process perspective (How does feedback influence the learner’s cognitive and metacognitive processes?).

Large-scale studies in Australia and the UK found that students are often dissatisfied with the provision of feedback, in particular the accuracy, timeliness and consistency of feed-
back information (Carless et al., 2011). Increasingly large class sizes and a wider diversity of student backgrounds can be seen as today’s main challenges in providing students with quality feedback (Hounsell, 2007). In addition, students’ level of engagement in the feedback process highly impacts the effectiveness of feedback.

Acknowledging these challenges and the complexity of the learning process, the question of What constitutes effective feedback? has no simple answer. In the following Section 2.2 the extensive literature on feedback is summarised and influential feedback models and advice of good practice are considered. In Section 2.3 Hattie and Timperley’s (2007) suggested model of effective feedback is introduced and augmented by important aspects as found in the literature.

2.2 Principles of good feedback practice

To address the question of what constitutes effective feedback researchers have either conducted experimental studies, reviewed the literature on these, or have investigated students’ and/or teachers’ perception of effective feedback. For example Beaumont, O’Doherty and Shannon (2008) studied the perceptions of quality feedback of 37 high school students and 13 teachers and compared these with perceptions of first year university students and tutors. The school students’ experiences of quality feedback are described as the “Dialogic Feedback Cycle” (DFC) where feedback discussions are aimed to improve students’ grades and occur at three stages of an assessed coursework task: (1) preparatory guidance, (2) in-task guidance and (3) performance feedback. The authors found that these experiences highly influence students’ expectations in the first year of their university study. Given large class sizes and a focus on self-directed learning at university, it is not surprising that these expectations cannot be met. Focus groups and interviews involving 108 first year students revealed a significant misalignment between students’ expectations regarding quality feedback at the beginning of the course and their actual experiences. Students reported little preparatory guidance (e.g. lack of explanation of the task criteria or discussion of model answers) and few opportunities for formative feedback. Issues with the feedback received (mainly at the performance stage of the DFC) involved inconsistency of marking and a lack of timeliness and detail. The authors are worried about the “demotivating effect that a perceived lack of quality feedback appeared to have on a significant proportion of students” (Beaumont et al., 2008, p. 9). The authors suggest to refocus on the first two stages of the DFC to address the specific problems identified by the study.

General principles of good feedback practice are widely discussed in the literature. Sadler (1989) developed a theory of formative assessment where, in order to benefit from feedback, the learner needs to “(a) possess a concept of the standard (or goal, or reference level) being aimed for, (b) compare the actual (or current) level of performance with the standard,
and (c) engage in appropriate action which leads to some closure of the gap” (p. 121, original emphasis). Underlying is the notion of feedback as information to help students to close the gap between actual and desired performance. This definition is also used by Nicol and Macfarlane-Dick (2006) in their literature review, which concludes with seven principles of good feedback practice (p. 205):

**Good feedback practice:**

1. helps clarify what good performance is (goals, criteria, expected standards);
2. facilitates the development of self-assessment (reflection) in learning;
3. delivers high quality information to students about their learning;
4. encourages teacher and peer dialogue around learning;
5. encourages positive motivational beliefs and self-esteem;
6. provides opportunities to close the gap between current and desired performance;
7. provides information to teachers that can be used to help shape teaching.

In contrast to Sadler’s learner-focused view, Nicol and Macfarlane-Dick’s emphasis is on the tutor and what kind of feedback needs to be in place. The authors, in line with Sadler (1989), stress that good feedback practice should strengthen the students’ ability to self-regulate their learning and that self-assessment skills are a vital precondition. Feedback targeted on students’ self-regulation ability is seen to have a long-lasting effect beyond university study.

### 2.2.1 Focus on self-regulation

Ideally, self-regulated students work in cycles of setting their own goals based on a genuine interest in the topic, select appropriate strategies to reach these goals, have the ability to focus their attention, self-monitor their progress, adjust the strategies accordingly, and reflect critically at the end of the process in terms of self-evaluation as well as attribution of failure or success which, in turn, leads to improvements for the next cycle. These “self-fulfilling cycles of academic regulation” are introduced by Zimmerman (1998) with numerous references to research in the field.

Before, in 1995, Butler and Winne published an influential literature review on feedback in the context of self-regulated learning. The authors developed a model of self-regulated learning, which places the student’s ability to generate internal feedback (“monitoring”) at all stages of the learning process in the centre. Further discussion about “sustainable” feedback followed to discuss strategies and practices to help students to develop into self-regulated learners (e.g. Hounsell, 2007, Carless et al., 2011). There is a clear shift from viewing feedback as information transmitted from lecturers and tutors to students towards
a perspective were the student is seen as the active part in the feedback process becoming increasingly independent (Nicol and Macfarlane-Dick, 2006).

Carless et al. (2011) define sustainable feedback practices as “dialogic processes and activities which can support and inform the student on the current task, whilst also developing the ability to self-regulate performance on future tasks” (p. 397) and seek for an answer in an interview study, involving teaching staff, as to what constitutes those practices. Besides multi-stage assessment practices, dialogical feedback on oral presentations and the promotion of student self-evaluation, the authors found feedback on the self-regulation level neglected because of (1) students’ resistance caused by their expectation to be told what to do, (2) lectures’ anxiety to challenge this resistance and (3) a packed curriculum focussing on topic-specific content only. The authors conclude on the need to communicate the purpose of the feedback and to highlight the students’ role in the feedback process.

2.2.2 Focus on teacher involvement

Communication and teachers’ active involvement in the feedback process is also emphasised by Hattie (2009). In his book “Visible Learning” he offers an additional perspective on feedback by stating that “teaching and learning can be synchronised and powerful” (p. 173) when teachers seek feedback from students to engage with students’ gaps in knowledge, level of understanding, the errors they make or misunderstandings they hold. This kind of feedback makes the actual learning “visible” to the teacher and provides an opportunity to act as problems occur. Likewise Nicol and Macfarlane-Dick (2006) suggest frequent assessment and diagnostic tests to enable lecturers to adapt their teaching. However in times of increasing class sizes and limited resources, frequent assessment of students’ actual understanding seems to be problematic.

2.2.3 Focus on peer-assessment

Peer-assessment is seen as one answer to improve feedback processes without increasing the workload of the teaching staff. Peer-assessment involves students in the grading process and was reported to have positive effects on student learning and self-regulation. In their literature review Liu and Carless (2006) describe the following advantages – that students:

- get an active role while managing their own learning,
- improve their understanding of standards and grading criteria,
- develop objectivity in the process of assessment,
- improve their own understanding about the subject matter by reflecting on and articulating of issues involved,
- receive more timely feedback than staff can provide (especially if challenged by mass education demands),
Strong links are seen between the involvement in the process of peer-assessment and the development of self-assessment skills (Nicol and Macfarlane-Dick, 2006). Despite these advantages a large-scale survey, conducted by Lui and Carless (2006), showed that students and staff alike are not in favour of peer assessment for four reasons: (1) concerns about reliability of grading, (2) perceived limited expertise of peers, (3) issues with power relations (academics sharing their powers as well as students feeling uncomfortable having the power of grading) and (4) time factors. To address the first three problems the authors promote *peer-feedback* which is described as a dialog providing “rich, detailed comments but without formal grades” (Liu and Carless, 2006, p.280) over *peer-assessment* where students primarily grade each other’s performance or work.

### 2.2.4 Focus on engagement

Gibbs and Simpson (2004) also address quality feedback in the light of assessment. Their observation that “Assessment sometimes appears to be, at one and the same time, enormously expensive, disliked by both students and teachers, and largely ineffective in supporting learning” (p. 11) motivates a review of theory and empirical studies. In conclusion they state 10 conditions under which assessment improves student learning. The 5 conditions addressing feedback imply that students are more likely to engage with feedback which is: (1) frequent and detailed, (2) focussed on task performance rather than on students’ personality, (3) timely to be still relevant for further learning and/or assistance, (4) aligned to the purpose and success criteria of the assignment and (5) appropriate to students’ level of understanding.

Even if all advice regarding good feedback practice is integrated into the course processes, it still needs the engagement of the learner for the feedback to be effective. Gibbs (2010) emphasises this aspect and notes: “It is not inevitable that students will read and pay attention to feedback even when that feedback is lovingly crafted and promptly provided” (p. 18). He suggests steps to improve students’ engagement by providing feedback without marks and the incorporation of self-assessment and multi-stage assessments. Consecutive hurdles in the process of students’ engagement with assessment feedback were identified by Price et al. (2011) as the failure to (1) collect the feedback response and (2) immediately attend to it, (3) understand the response and (4) take action due to limitations in resources, skills, opportunities or confidence. This perspective highlights important stages in the engagement process where each stage triggers further engagement (or disengagement). Repeated unsatisfactory prior experience with the feedback process is likely to lead to disengagement in such a way that students do not even collect the assessment feedback. The
authors point out that such behaviour is not an instant response but developed over time and involves different courses and programs, when for example feedback was continuously found to be useless (e.g. too general or not understandable) or staff was perceived as too busy to discuss feedback responses.

2.2.5 Focus on dialog

As dialogue appears to be an effective way to ensure that feedback is received and understood, it is not surprising that dialogic feedback is highly praised in the literature as desirable practice. For example the already mentioned Principle 4 of good feedback practice (Nicol and Macfarlane-Dick, 2006) focuses in particular on the encouragement of teacher and peer dialogue. This was reiterated by Carless et al. (2011) based on findings of a study of actual feedback practices.

Likewise Price et al. (2011) state that dialogue is a key element of the feedback process, which should (1) communicate feedback purpose and processes upfront, (2) foster positive learner identity early on, (3) address feedback and invite students’ responses, (4) guide and encourage students’ engagement with the feedback and (5) as a result lead to a shared understanding of the complexity of the feedback process. The authors note:

Supportive dialogue is a key part of the social practice of assessment, and is dependent upon trust and the perception of a joint enterprise involving students and staff. Students recognise their need for dialogue to enable them to fully work with their feedback and to induct them into the disciplinary community but they are frequently frustrated by lack of opportunity and by the social structures that obstruct dialogue. (Price et al., 2011, p. 894)

2.2.6 Focus on students’ personality and their perceptions of feedback

The level of engagement with feedback also depends on the learners’ personality and their self-theories. Dweck (1999) found that students, who see intelligence as developmental are more learning-orientated and view challenges as opportunities for learning. In contrast, students, who believe that intelligence is fixed, were found to be somewhat performance-orientated and likely to give up when faced with difficulties. Those students may interpret negative feedback on a personal level with damaging effects on their motivation and self-esteem, because failure is seen as a lack of intelligence. For a summary of self-theories see Yorke and Knight (2004). A demoralizing effect of negative feedback was reported by Poulos and Mahony (2008) based on a study of students’ perceptions of effective feedback. Especially in the first year of their study at university, students experienced negative feedback or failure as devastating and expressed a need for more communication to aid the transition from school to university.
How students’ perception of feedback influences their engagement with the feedback process was the focus of a study by McLean (2012). Conducting a phenomenographic study, used to investigate the variations in the way how a phenomenon is experienced, the author found that students view feedback as information in four different ways of increasing complexity: (1) as “telling”, (2) as “guiding”, (3) as “developing understanding”, (d) as “opening up a different perspective”. The more complex conceptions include the simpler ones. Furthermore McLean could demonstrate that the more inclusive and expansive students’ perception of feedback is, the lower are the barriers to respond to feedback provided. She concludes regarding “better” feedback practices:

*In this instance, ‘better’ does not necessarily mean more work for teachers. Instead, ‘better’ means working with students to figure out their view of feedback and tailoring feedback information to fit. It can also mean stimulating and challenging students to reflect on their current views.* (p. 274)

### 2.3 A framework for effective feedback

Best possible feedback processes acknowledge and respond to the learner’s personality and provide detailed, timely feedback on an individual level. This is particularly important in the first year to influence learners’ successive engagement in these processes positively and to help the transition from school to tertiary study.

However there seems to be a contradiction between the requirement of feedback being detailed as well as learner- and task-specific and the aim of feedback to help students to develop into independent, self-regulated learners over time. Hattie and Timperley (2007) established a model to consider feedback, which untangles the different levels and stages of the process. Based on their review of numerous meta-analyses they identify four levels of feedback (Hattie and Timperley, 2007, p. 87, Figure 1):

1. Task level – How well tasks are understood/performed.
2. Process level – The main process needed to understand/perform tasks.
4. Self level – Personal evaluations and affect (usually positive) about the learner.

On “self level”, feedback is often in the form of praise and used to comfort or support students, but usually contains little task or process related information and is considered to have limited potential to improve learning. Despite the authors’ earlier remark that praise directed to the effort, self-regulation or personal engagement “can assist enhancing self-efficacy” (Hattie and Timperley, 2007, p. 96) the authors state that “Praise may be counter-productive and have negative consequences on students’ self-evaluation of their ability”
and point to a certain “unpredictability” of praise for different groups of learners (Hattie and Timperley, 2007, p. 97). Because of these inconclusive findings and the limited transferability into a specific learning domain, feedback on “self-level” is not part of the framework introduced in this section.

In addition to the four levels of feedback Hattie & Timperley (2007) state three major questions, which need to be addressed at each level. Only by answering these questions can a discrepancy or gap between the actual and the desired performance be highlighted and acted on.

1. **Where am I going? (“feed up”)** – Learning intention, goals, success criteria: Goals need to be specific rather than general, and sufficiently challenging.

2. **How am I doing? (“feed back”)** - Actual performance, understanding: Feedback regarding expected standard or success criteria and not in comparison with other students’ progress.

3. **Where to next? (“feed forward”)** – Progression and new goals: Information that leads to greater learning possibilities, enhanced challenges and the development of more self-regulated learning.

The authors point out that these questions are interrelated, e.g. next steps can only be considered in relation to the goals and the current progress. It is important to notice that Hattie and Timperley account the self-regulation level as the most effective one, but raise awareness that feedback on task and process level are necessary to enable students to act on self-regulation level. Their model illustrates that feedback should not be seen as a series of events but rather as an on-going process on different levels, where a shift from feedback on task-level towards the self-regulation level is desired over time.

The underlying notion of feedback as closing the gap between desired and actual performance, highlights the need to establish feedback not only around students’ actual achievement, but also to clarify goals and success criteria upfront as well as to indicate steps for improvement. The resulting three stages as stated in Hattie and Timperley’s model of effective feedback (“feed up”, “feed back”, “feed forward”) are used in Table 2.1 to structure the most important aspects of good feedback practice as found in the literature. Questions were formulated to extend Hattie and Timperley’s model. In seeking to improve existing feedback processes these questions need to be addressed on task, process and self-regulation level.
Table 2.1: Questions guiding effective feedback practices on task level, process level and self-regulation level based on Hattie and Timperley (2007).

<table>
<thead>
<tr>
<th>Goals (“feed up”)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Are goals clearly communicated and understood by the learner?</td>
<td></td>
</tr>
<tr>
<td>Are goals realistically set for the learner’s situation?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actual progress/ performance (“feed back”)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is feedback given regularly, in appropriate detail and according to the purpose of the task and the goals set?</td>
<td></td>
</tr>
<tr>
<td>Is the learner’s level of understanding incorporated in the way feedback is given?</td>
<td></td>
</tr>
<tr>
<td>Is feedback always focussed on the progress/performance (under the student’s control) rather than on the student’s personality?</td>
<td></td>
</tr>
<tr>
<td>Are frequent assessment tasks and diagnostic tests in place to uncover students’ difficulties? Is this information used and addressed by the teacher?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activities necessary for better progress/ performance (“feed forward”)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Are strategies/advice provided on how to improve to reach the goals? Is this based on the student’s actual progress/performance?</td>
<td></td>
</tr>
<tr>
<td>Is corrective advice given rather than information about strengths/weakness?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communication in general</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is feedback/feed forward timely and can it still be acted on? (e.g. resubmission or multi-stage assessments)</td>
<td></td>
</tr>
<tr>
<td>Is it ensured that feedback/feed forward is received and acted on?</td>
<td></td>
</tr>
<tr>
<td>Is student-teacher or peer dialog encouraged whenever possible?</td>
<td></td>
</tr>
<tr>
<td>Is the learner’s personality taken into account when feedback/feed forward is provided?</td>
<td></td>
</tr>
</tbody>
</table>

2.4 Summary

Literature in higher education provides many studies, frameworks and opinions on feedback. Quality feedback is generally recognised, as being timely, detailed, relevant and understandable for the students. For the process of learning, effective feedback states and clarifies the goals, informs students about actual performance and indicates ways for improvements towards these goals; and supports the development of students’ self-regulation. Ideally feedback should take the learner’s characteristics and abilities into account and make sure feedback messages are received, understood and can be acted on. Furthermore good assessment and feedback practice should provide opportunities for lecturers and tutors to gain insight into students’ difficulties and misunderstandings, which can be addressed while teaching. It is important to implement good feedback practices not only at task level but also at process and self-regulation level.

Feedback which targets students’ processes of self-regulating their learning is seen to have long lasting effects beyond university study and is also called “sustainable feedback”, as self-regulated learners work more and more independently in cycles of setting their own goals, self-monitoring the actual performance and adjusting strategies accordingly. Therefore contemporary principles of good feedback practice acknowledge feedback not only as
information provided in a detailed and timely manner, but also as a complex process, incorporating various influences on students’ actions and behaviour. Here aspects of students’ engagement and teachers’ involvement, the importance of dialog and peer-assessment as well as the learners’ personality were found to be influential for the effectiveness of the feedback process.

To structure those aspects and address the question of “What constitutes effective feedback?” a model proposed by Hattie and Timperley (2007) to consider effective feedback was used. By viewing feedback as closing the gap between desired performance and actual performance on task, process and self-regulation level, the authors propose three stages of the feedback process: (1) clarify the goals, (2) consider actual performance and (3) assist necessary activities for better progress or performance. We expanded this model by questions considered as important to assess and improve feedback practices for each stage and around general aspects of communication. However frameworks such as the one above (Table 2.1) often lack advice for practical implementation. In the following chapter we provide such practical advice by discussing the different levels and states of the feedback process and the possibilities to improve feedback in a generic CS1 context.
Chapter 3

Feedback in Computer Science Education

“The real challenge to computing education is to avoid the temptation to re-invent the wheel. Computers are a revolutionary human invention, so we might think that teaching and learning about computers requires a new kind of education. That’s completely false: The basic mechanisms of human learning haven’t changed in the last 50 years.”

(Almstrum et al., 2005, p. 191)

This chapter is based on sections 3 and 4 of a paper submitted to ACM Transactions on Computing Education that is currently under revision:

Ott, C., Robins, A., & Shephard, K. (2014). Translating principles of effective feedback for students into the CS1 context. Manuscript accepted for publication.
3.1 Introduction

In this chapter feedback is considered specifically within the context of a CS1 course, identifying practical opportunities for applying the principles of good feedback in the teaching and learning of programming. The common features of typical CS1 courses comprising lectures, practical (laboratory/tutorial) sessions and assessment (including regular practical tasks, larger assignments and exams) are discussed. This discussion is based on the CS1 course at our own institution, COMP160, which is run as a typical combination of weekly lectures and accompanying lab sessions over one semester. The content of COMP160 is based on a standard textbook (Lewis, DePasquale and Chase, 2010) and introduces the programming language Java so that students are familiar with the main principles of object-oriented programming by the end of the course. However, the discussion is intended to be generic, so that the answers to the following three questions should be relevant to most CS1 courses:

- How to consider feedback on task, process and self-regulation level in the CS1 context?
- How to address the questions guiding effective feedback practices on those three levels in a typical CS1 course?
- What topics from computer science education research need to be considered to establish or support current feedback practices?

In order to answer these questions each level of feedback (task, process and self-regulation level) is addressed separately and in the following three sections (3.2 to 3.4). Course components involved at each level are specified and the questions guiding effective feedback practice are considered for those components (e.g. tutorial session or examinations). Whenever applicable potential areas for the improvement of feedback practice are identified and entry points to the related research in CSEd to support those improvements are highlighted. A short summary at the end of each section provides a quick overview about the most important aspects addressed in the discussion.

3.2 Feedback on Task Level

Feedback at this level is concerned with a specific learning task, for instance writing an essay, or in the case of CS1, writing a small computer program. Goals and success criteria need to be clearly stated and regular feedback should clarify how well the task has been accomplished. Missing aspects or faulty interpretations need to be addressed by corrective advice. Hattie and Timperley (2007) emphasise that if feedback on this level is too detailed or specific it might distract from strategies and lead to trial-and-error behaviour. It is likely to be most beneficial if erroneous hypotheses can be rejected and cues on further direction
and strategies are provided (Hattie and Timperly, 2007). Usually this kind of feedback does not generalize to other tasks.

Depending on the particular structure of a CS1 course, feedback on task level needs to be considered for some or all of (1) the practical tasks students are expected to solve during tutorial session (including any preparation for such tasks) or extended assignments (extending beyond a single tutorial session), and (2) tasks or exercises which occur in lectures, and (3) any examinations, which can also be considered as tasks, because students have to accomplish them as part of the course.

3.2.1 Practical tasks during tutorial time or as extended assignments

In our course the two weekly lectures are accompanied by two weekly, tutor-assisted laboratory sessions of 2 hours each. No formal teaching is conducted and students work in their own pace on the practical tasks. There is one practical task designed for each of the 25 scheduled sessions. Students can “call” (via a queuing system) when problems occur or when they completed the task and want to submit the work, which requires a tutor to check the task solution for correctness. Each call is attended by one of the two or three tutors present in the lab.

The goals for task completion are typically stated as a description of the program's intended functionality, which is then taken further apart into instructions. Test cases might be defined to verify the correctness of the implementation. It is important that students fully understand the task requirements to avoid later confusion when working through the instructions. Embedded questions as part of the task description might be considered for more engagement with the intended functionality and the instructions, but we face a common problem here: as some students are challenged by understanding the task goal and instructions, other students might feel that the task goal is rather trivial and instructions are already overly prescriptive. "Extension tasks" which are not a requirement for task completion, but are a voluntary addition can be seen as one possibility to challenge those students and adjust the task goal to the learners' needs.

Following the advice for effective feedback students should get regular, detailed and timely feedback regarding their actual performance. Scheduled occasions to uncover gaps between the goals and the actual performance are arranged when students are required to discuss their solutions with a tutor before submission. In our course a tutor would ask the student to run the program and to demonstrate some test cases. The code is checked for the required concepts (e.g. use of variables vs. “hard” coding) and programming style (e.g. commenting). If everything is solved correctly this process takes one to two minutes. In the case that deficiencies are discovered the amount and quality of feedback depends on different
factors: (1) the time available, (2) the tutor's personality and experience and (3) the student's attentiveness and willingness to discuss problems. Within this spectrum marking students' lab work can produce a lively discussion about misunderstandings or alternative solutions, or a rather short advice, regarding parts of the preparation task or the program solution to be reviewed. In each case the discussion needs to be focussed on the solution at hand rather than on the student's personality, but should take the student's level of understanding into account.

The tutors' varying experience (e.g. professional teaching fellows or 3rd or 4th year students) and teaching styles can cause inconsistency in the quality of feedback, especially if students talk to different tutors regarding one task. Arranging for "personal tutors" in a way that students are preferably approached by the same tutor during one lab session would result in more consistent advice. The use of software metrics (e.g. Cardell-Oliver, 2011) and assignment rubrics (e.g. Becker, 2003) could be evaluated for shared marking criteria amongst tutors.

The students' level of activity and readiness to seek help also influences the amount of feedback they get. Active students with rather outgoing personalities get more feedback by "calling" for help frequently and at different stages of task completion. These calls are excellent opportunities for individual and detailed feedback on task level. However, problems occur if students are not attending the labs regularly or do not call for help or do not submit lab work. Here formal tutorials or a required minimum of lab sessions to be attended might be the answer.

Feedback to indicate how a student can improve is directly linked to the feedback on the actual performance. In the lab tutors would discuss students’ actual programming solutions by referring to the requirements. If the standards are not met, tutors would give advice on how to improve the solution. This iterative process might start with the advice to review certain concepts in the textbook or lecture notes and develops into more task-specific feedback later on. From our experience it is always tricky to maintain the right balance between “telling students what to do” and “helping them figure out how to get there themselves”. Hattie (2009) writes: “The art is to provide the right form of feedback at, or just above, the level where the student is working […]” (p. 177). Asking questions to guide their line of thinking seems a good approach for students who are “almost there”. However such feedback is not effective if students are still at the knowledge acquisition phase – here instructions are more helpful (Hattie, 2009).

One-to-one conversations provide in fact the most desirable form of feedback, but may appear as not sustainable in times of increasing student numbers and current staff student ratios. “Extreme Apprenticeship”, a form of continuous, personal feedback between advi-
sor and students, was introduced by Kurhila and Vihavainen (2011) for CS1 classes with about 150 to 200 students and resulted in higher pass rates and improved student satisfaction without extending the budget on teaching staff. Investigating students’ perception of feedback Pears et al. (2013) found students unsatisfied with the feedback they get. As one way to improve the feedback processes the authors suggest discussing not only assignment results on a one-to-one basis or as a group but also the expectations before the due date of the assignment. A preference for individual feedback over group feedback was voiced by teaching staff in Isomöttönen and Tirronen (2013) based on the experience that summarising feedback relevant for an entire group was found to be too time consuming. That students in fact appreciate one-to-one feedback was shown by East and Schafer (2005), who conducted an experiment using in-person grading as one of three experimental conditions. Significant differences in student grades could not be observed between the groups, but students of the in-person grading group “were far more satisfied with their feedback method” (East and Schafer, 2005, p. 381).

3.2.2 Lectures

The introduction of interactive teaching techniques is a possibility to empower the lecturer to monitor more closely students’ actual understanding by frequent assessment of newly introduced concepts even in large classes (Bruff, 2009). Classroom response systems such as Clickers1 are a promising technology to reveal students’ misconceptions by posting conceptual questions or small tasks. Students can provide an answer choice for multiple choice questions (MCQ) or a short answer by using a remote control. Depending on students’ feedback, the lecturer might choose to alter the course of the lecture to address incorrect answers immediately. Improvements of students learning outcome have been reported for example for large physics classes (Deslauriers, Schelew and Wieman, 2011).

In the context of CS1 and CS2 courses Chamillard (2011) reports on students’ enthusiasm for the new technology but not necessarily on improvement of the learning outcome, whereas Cutts and Kennedy (2005) found students’ actual engagement with the technology to be limited and report on a poor correctness of students’ responses. Just-in-Time Teaching (JiTT) is an example of an interactive teaching technique, which is often used in combination with a classroom response system. Carter (2012) reports on the introduction of JiTT in a large CS1 class setting. The traditional lecture content was delivered by screen-casts, which students were expected to watch before the lecture. At the beginning of each lecture Clickers were used to assess students’ comprehension of the material and “deliver mini-lectures on an as-needed basis” (Carter, 2012, p. 362). The remaining lecture time was used for in-class activities based on peer instruction (e.g. Porter et al., 2011; Zingaro, 2014).

1 www.h-itt.com
Bailey and Forbes (2005) implemented JiTT by providing web-assessments prior to the lecture (web component) and feedback regarding these assessments in the lecture (classroom component) in an introductory CS course. The classroom component also included the use of a personal response system similar to Clickers. Gannod, Burge and Helmick (2008) propose the “inverted classroom” to promote in-class activities over traditional lectures, which are delivered through podcasting, implementing immediate feedback on a regular basis for software engineering students. In general the interventions mentioned above were seen as to improve students’ engagement with the subject matter, but rarely are linked to students’ learning outcome. Kennedy and Cutts (2005) found the frequency and correctness of students’ responses over the semester associated with their performance in the end-of-semester assessment and report on a positive relationship between students’ use of the voting system and their learning outcomes. Zingaro (2014) presents results suggesting that peer instructions increase students’ self-efficacy, but an improvement on students’ final exam results was not significant.

Further advances towards students’ active contribution before or during the lecture time are described in a review by Hamer et al. (2010) and include “content creation” (e.g. course material or algorithm visualisations), “solution sharing” (e.g. code reviews) and “annotations” (e.g. digital ink annotations created and transmitted via a Tablet PC during the lecture). The authors raise concerns about missing evaluations or unclear definitions of “success” in the reports they reviewed.

3.2.3 Mid-semester and final examination

In COMP160 the 25 lab tasks add up to 25% towards the final grade. Written examinations at mid-semester and the end of the semester contribute another 15% and 60% respectively. Students must pass the final examination to pass the course. This passing criterion (achieve more than half marks) makes the final exam an important part of the course. However performance orientated goals such as scoring over 50% to pass the exam, are not useful to guide students’ preparation and successful completion of the examinations. Here practice examinations can be made available prior to the examinations to illustrate learning goals in terms of concepts or skills required. In our course practice examinations, which are made available online, can be studied and discussed with tutors during practical sessions.

The format of a written, paper-based examination using MCQ and/or short answer questions appears to be problematic, because it leads to a mismatch between the overall intended learning outcome of solving programming problems and the actual assessment of definitions and selected concepts. Mismatches between intended learning outcome and actual assessment tasks are suspected to encourage short-term and surface learning (e.g. Biggs and Tang, 2007, Carless, 2007). The students’ tendency to “learn what they think they will
be tested on” (Biggs and Tang, 2007, p. 169, original emphasis) can be used constructively by aligning the assessment with the intended learning outcome. Learning-oriented assessments as described by Carless “should promote the kind of learning dispositions required of graduates and should mirror real-world applications of the subject matter” (Carless, 2007, p. 59). However, it may be difficult in practice to assess possibly hundreds of students’ practical programming skills in a computer laboratory. Automated assessment tools are widely discussed (for an overview see Ihantola and Ahoniemi, 2010) and can be considered to reduce the workload of marking, but conducting a fair computer-based assessment for big classes seems still problematic.

Feedback about the actual performance in the examinations only as an achieved score, provides little information on how the learning goals have been met and which concepts need to be reviewed to improve future performance. Paul Ramsden shares a strong opinion in his book “Learning to Teach in Higher Education” about assessment feedback in form of a mark or grade only:

> It is impossible to overstate the role of effective feedback on students’ progress in any discussion of effective teaching and assessment. Students are understandably angry when they receive feedback on an assignment that consists only of a mark or grade. I believe that reporting results in this way, whatever the form of assessment, is cheating students. It is unprofessional teaching behaviour and ought not to be tolerated. (Ramsden, 2003, p. 187)

He is not the only author raising concern. Gibbs and Simpson warn: “grades without feedback may be particularly damaging” (Gibbs and Simpson, 2004, p. 18). Especially for weaker students, assessment grades or scores are a measure of their failure and are not likely to encourage learning. The authors suggest that feedback would be less personal if addressing the content and options for further actions. For the same motivational reason Hughes (2011) advocates an “ipsative” approach for assessment, which measures the student’s best (based on previous performance) rather than a score against a fixed standard. Based on findings from the literature that students have been more engaged with assessment feedback when given without a grade, Irwin et al. (2013) conducted a case study using “adaptive release” of feedback, where students across different subjects and faculties were required to reflect on the feedback before grades were released. Investigating students’ perceptions about this process, the researchers found that students felt more engaged with the feedback but also annoyed when the grade was seen as the principal outcome of the assessment process.

In terms of engagement, the purely summative nature of the final exam is unlikely to encourage learning. With no opportunity to improve the final grade, feedback is most likely
not perceived as relevant (e.g. Carless, 2011). In contrast “feed forward” regarding the mid-
semester examination could be powerful to reveal and communicate concepts for review
and practice. As the concepts in computer programming are seen as highly integrated
(Robins, 2010) it is vital that students internalise the lower-level concepts before moving
on to higher-level concepts. For that reason the mid-semester examination (or any other
examination occurring while the course is in progress) should trigger individual advice on
topics and resources (e.g. lecturer notes, lab tasks or textbook chapters) to review.

Personalised feedback could be provided by combining pre-written text blocks, a method
used by the Rubyric system as described in Auvinen (2011). The idea of interactive cover
sheets (Bloxham and Campbell, 2010) can be also explored: after finishing the examination
students indicate topics or tasks which they want to discuss in a practical session later on or
as a topic to be reviewed during a lecture.

3.2.4 Summary

Feedback on task level was reviewed on three course components: (1) the programming
tasks to complete in tutor-assisted lab sessions or as extended assignments, (2) frequent
small tasks during the lectures and (3) mid-semester and end of the semester examinations.

Good feedback practice is established in tutor-assisted lab sessions where students get
timely, detailed feedback in a personal dialog with the tutor. This way feedback is often a
cycle of discussing the requirements, revealing gaps/deficiencies of the actual solution and
giving corrective advice on how to work towards the requirements. Problems in this pro-
cess occur when (1) students are inactive or attend lab sessions infrequently or (2) tutors
are inexperienced, not attentive to students’ level of understanding or are simply too busy.
Inconsistency in the way feedback is provided when one student is attended by different
tutors, is not avoidable, but could be addressed by “tutoring guidelines” in form of model
answers and shared marking criteria. Formal feedback sessions might be introduced for
students who appear inactive.

The use of interactive teaching techniques to conduct small tasks before or during the lec-
ture and establish discussions around students’ misunderstandings in a lecture setting have
been explored. Further research seems to be necessary to develop more effective scenarios
for the use of such classroom response systems and to assess their effectiveness.

There is a serious lack of feedback if examinations do not generate feedback other than a
score. This is potentially a missed opportunity to guide students’ learning based on their
actual performance. One way would be to address common mistakes, lack of knowledge or
misconceptions in a lecture. The other would be written feedback along with the results
and/or a personal dialog during a practical session regarding the exam results. In each case
it is important to add meaning to the score and give clear advice on the topics which need to be revised and what resources are available. Table 3.1 provides an overview of topics discussed in this section:

Table 3.1: Feedback on task level

<table>
<thead>
<tr>
<th>Task level</th>
<th>Lab tasks</th>
<th>Lectures</th>
<th>Examinations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goals</strong></td>
<td>Verifying whether students understood the goals / requirements of the tasks when problems occur or before they approach the task. Incorporating ways to set goals according to the student’s individual level of understanding (common vs. expert goals).</td>
<td>Conducting small tasks before or during the lecture to uncover misconceptions or lacks in knowledge.</td>
<td>Shifting performance orientated goals (score for passing) towards learning orientated goals (concepts, e.g. illustrated in practice exam sheets). Aligning the format of the examination with the intended learning outcome as closely as possible.</td>
</tr>
<tr>
<td><strong>Actual Performance</strong></td>
<td>Assuring that all students get a minimum of regular, individual feedback. Seeking consistency in the way feedback is provided amongst tutors. Discussing matters of good feedback practice with new tutors.</td>
<td>Addressing correct and faulty answer choices during the lecture. Adapting the lecture content based on students’ answers.</td>
<td>Providing detailed, concept-orientated feedback for mid-semester examination (or other forms of formative assessment during the course) to encourage revision. Addressing gaps in knowledge or common misconceptions during the lectures.</td>
</tr>
<tr>
<td><strong>Activities for Improvements</strong></td>
<td>Discussing ways for possible improvements along with unsatisfactory solutions. Taking the students’ attentiveness and level of understanding into account (instructions vs. questions)</td>
<td>Providing corrective advice and resources for revision for prominent misconceptions.</td>
<td>Advising resources available for revision before and after the examination.</td>
</tr>
<tr>
<td><strong>Communication in general</strong></td>
<td>Aiming for one-to-one dialogues as the most suitable form for timely, detailed and individual feedback.</td>
<td>Considering interactive teaching technology especially for large classes for immediate collection and discussion of students’ answer choices.</td>
<td>Providing individual and timely feedback with the opportunity to revise and improve (e.g. re-sitting the exam).</td>
</tr>
</tbody>
</table>

3.3 Feedback on Process Level

According to Hattie and Timperley (2007) feedback on process level is addressing the processes necessary for task completion. One aspect is to help students to develop strategies for error detection and correction. Another aspect is to provide students with cues, which guide information search and the application of strategies. The authors claim that this type
of feedback is more effective than feedback on task level, because it is targeting the development of task strategies rather than the task outcome. In other words this feedback should empower students to become more independent in the process of task completion.

Translated to the CS1 context, feedback on process level is concerned with firstly the programming process rather than the program’s correctness and functionality, and secondly with the processes on course level such as meeting terms requirements, attending the lectures and lab sessions or preparing for lab tasks and examinations. These two aspects are used to structure the following two sections.

### 3.3.1 Programming process

What are the expectations and goals in terms of managing the programming process? Ideally, students understand the requirements of the task and plan their program solution before they start to code. The program is developed in small, testable chunks where test cases are planned and checked at each stage of the development. If compiler or run-time errors occur, students are confident in reading and interpreting the error messages. They can solve most common problems independently by using an (online) reference, a debugger or other strategies to localise and address the problem (e.g. commenting out statements in question, using printouts to watch variable content and program states). At the end of the development, requirements are reviewed, final tests are run and a critical reflection might lead to some ideas about improvements. It goes without saying that the code is commented, legible and follows general “clean code” rules as stated in Martin (2009).

There are many implicit goals regarding the process of programming:

- understanding the requirements,
- planning the solution,
- successive development,
- frequent testing,
- fluency in using error detection and correction,
- good programming style and
- critical reflection.

For a CS1 course it is important that these goals are clearly communicated at the beginning and reinforced during the course as they become relevant. In our course good programming style in terms of commenting code and sensible variable names are requirements from the first task onwards. Tutors are expected to discuss deficiencies and ways of improvement. In contrast, planning the solution, successive development or frequent testing are not requirements for the task completion and therefore these activities are often ne-
glected in the feedback process. If no goals are stated, the discussion about actual deficiencies and possible improvements is unfounded.

In general it is problematic to teach the advantages of good programming practice because students need to have some programming experience and need to be faced with programming problems of a certain complexity to appreciate strategies and techniques such as planning and commenting. It is a matter of good timing to introduce process-related strategies when they are most relevant for successful task completion. Exercises should be scheduled accordingly for discussion and feedback. Another way of communicating goals of performance is seen in providing “exemplars”, authentic students’ work to illustrate high-quality achievement (Sadler, 2002). Annotated model solutions might be useful to clarify some clean code principles.

Assessing the actual performance while solving a programming task is closely related to error detection and correction and the use of compliers/interpreters and debuggers. Compilers or interpreters will keep posting error messages until the last syntax error is resolved. That can be a frustrating process especially for a beginner who might interpret these errors as personal failure or lack of ability (Perkins et al., 1989). Unlike in essay writing for example, there is no way to assume the program is well crafted if it is not compiling. On the other hand, also different from writing an essay, students are potentially able to detect and address program errors in a self-guided way. However, the tools involved are complex and require a certain degree of user experience. For example error messages are often cryptic or even misleading and require tutor support to be resolved. Small exercises could help students to learn how to read a runtime error or typical compiler errors and how to resolve such errors.

Programming environments for novices have been developed to reduce the complexity of professional environments. In a literature review of teaching introductory programming, Pears et al. (2007) provide an introduction and overview of educational programming environments. These are often equipped with visual debuggers, which provide instant feedback regarding actual program behaviour. There are numerous algorithm visualisation tools, working mainly as stand-alone applications (for an overview of the scope see Shaffer, Cooper and Edwards, 2007). These tools are not linked to actual programming tasks but can be considered as learning tools supporting the programming process.

If students are required to talk to a tutor at the stage of task completion, this would be an excellent occasion not only to discuss how students met the expectations of good programming practice and how they could improve, but also to guide students towards a critical reflection on their current solution (What have you done to solve this problem? Why did you use concept X and not Y? Are there alternative solutions?). However Isomöttönen
and Tirronen (2013) found students’ engagement with reflection tasks at the point of task completion to be limited as students had no opportunity to apply the insights to the already finished tasks. That “feedback should directly enhance student outcomes on subsequent tasks” (Pears et al., 2013, p. 107) was stated as one key principle of four to improve feed-forward feedback for two introductory programming courses in Sweden and Australia. The authors note: “It is important that whatever the chosen exercises, that feedback deals with generic skill development, and is useful in improving performance on the next task in the sequence” (Pears et al., 2013, p. 111).

In summary, marking criteria for programming tasks need to address, besides the program’s functionality, also the programming process and matters of good programming style. A clear communication of those marking criteria at the start would help students to understand the expectations and act accordingly. If process related goals are clearly stated, feedback can address unsatisfactory programming approaches and assist improvement towards these goals for the next tasks.

### 3.3.2 Course related processes

For most CS1 courses students will be expected to attend all lectures and practical sessions. Furthermore students may be expected to come prepared to practical sessions. Before examinations, students will generally be expected to review lecture notes and other course materials. If questions occur, time during practical sessions can be used for clarification. These are all “common sense” rules/goals for us as teachers, but are they also for our students? What needs explicit enforcement? What needs to be followed up? When should we intervene if these standards are not met? How these questions are addressed depends on the specific characteristics of the course and on what were seen as successful strategies in the past.

If lecture attendance is not compulsory, students need to be self-motivated to attend regularly. A tight link between the lectures and the required practical work can help to emphasise the relevance of the lectures: referring directly to upcoming practical tasks and presenting similar examples with specific clues or hints, which assist task completion by addressing the most common problems, could increase students’ motivation to attend the lectures and lead to more independent work in the practical sessions. If students are struggling with the practical work and have not been to the lecture, feedback could be directed to clarify the expectations and to find ways to motivate improvements of study attitudes and behaviour on an individual basis. Anecdotally there is a strong correlation between the completion of practical work and final outcomes in CS1. Certainly this is the case in our course, where data from 2006 to 2012 show a strong correlation between the number of submitted lab tasks and the score in the final exam ($r=0.602, p<0.01, N=1111$), hence reinforcement of
regular practical session attendance and task completion should be a top priority, even though the correlation does not imply causation.

If goals (such as a number of practical sessions which must be attended) are clearly stated as course requirements it is possible to inform students early on in case the requirements are not met and to search for solutions to improve actual study behaviour. Ideally the students contacted need to get back to course coordinator to ensure the feedback was received, understood and further action can be discussed by taking students’ individual situation into account. Such a discussion can lead for example to the arrangement of private tutoring sessions as well as the decision to withdraw from the course to concentrate on other courses.

Assuming that better preparation for any mid-semester or final examination results in better understanding and a higher pass rate, providing practice examinations is a good starting point. Those practice exams could become an engaging and powerful element in the reviewing process if students (1) are clearly expected to solve these, (2) get automatic or personalised feedback on their performance and resulting scores and (3) topics and resources for further review are provided.

3.3.3 Summary

The process level of feedback was discussed for two very different processes CS1 students need to manage: (1) the process of completing programming tasks and (2) the process of meeting the course requirements. While teachers may have high expectations of how students manage these processes, they can also assist in terms of setting goals, providing structured feedback, and highlighting actual performance and opportunities for improvement.

Feedback processes need to be established around all aspects of good programming practice. There are a number of expectations regarding the programming process, but few of them are stated as standards or goals in our course. The lack of explicit goals and related exercises (e.g. using pseudocode or UML diagrams for planning a solution) leads to a lack of feedback and feed-forward. When goals regarding the programming process are only partially communicated and assessed, their relevance is diminished for students and tutors alike. However process-orientated feedback targeting general aspects of good programming practice is important for students’ development into independent, efficient software programmers. The emphasis on planning programming tasks is as vital as providing students with knowledge and resources for effective error detection and correction. Tasks could be designed to aim for the development of those strategies and could offer possibilities to discuss students’ progress in this process.
Common sense rules define good study behaviour. The question of what we see as “good enough” or as “not satisfactory” might be hard to judge for first year students. Here explicit expectations and goals need to be stated and reinforced at different stages of the course. It seems a good idea to explain why these goals are set out to make them relevant for the students (e.g. “We expect you to work continuously and to the schedule, because course history has shown that catching up later is unlikely due to growing complexity of the tasks”). Important goals need to be accompanied by regular feedback to inform students regarding their performance in meeting these goals and by guiding improvement if necessary. The feedback process is broken if scheduled occasions and information considering students’ actual performance are missing. Aspects of feedback on process level are summarised in Table 3.2:

Table 3.2: Feedback on process level

<table>
<thead>
<tr>
<th>Process level</th>
<th>Programming</th>
<th>Course related</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goals</strong></td>
<td>Introducing explicit goals and exercises to establish good programming practice.</td>
<td>Aiming for a tight link between lectures and lab work to emphasise the importance of lecture attendance.</td>
</tr>
<tr>
<td></td>
<td>Using best practice examples or model solutions to illustrate these goals.</td>
<td>Encouraging students to attend the lectures individually if problems with the subject matter become apparent.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Considering a number of compulsory lab sessions or number of submissions as course related requirements.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communicating explicit expectations or goals for the preparation of the examination.</td>
</tr>
<tr>
<td><strong>Actual Performance</strong></td>
<td>Explaining that errors are part of the process and no measure of ability or intelligence.</td>
<td>Providing feedback as early as possible if students do not meet the course requirements.</td>
</tr>
<tr>
<td></td>
<td>Providing information / guidelines / resources to empower students to address compile and run-time errors in a self-guided way (e.g. emphasis on the use of online resources such as references, tutorials or discussion boards).</td>
<td>Making sure that students received the feedback by asking for responses.</td>
</tr>
<tr>
<td><strong>Activities for Improvements</strong></td>
<td>Encouraging students to reflect on their program solution and think about possible refinements applicable also to improve subsequent tasks.</td>
<td>Ensuring that actions for improvement are discussed (e.g. additional lab sessions or private tutoring) and personal matters are taken into account (e.g. sickness or extreme workload).</td>
</tr>
<tr>
<td><strong>Communication in general</strong></td>
<td>Using an IDE specifically designed for beginners in programming.</td>
<td>If important requirements or goals are stated, feedback needs be provided on a regular basis and in a timely manner regarding students’ actual course performance.</td>
</tr>
<tr>
<td></td>
<td>Extending the discussion of an actual programming solution towards process orientated aspects (e.g. successive development or clean code principles)</td>
<td></td>
</tr>
</tbody>
</table>
3.4 Feedback on Self-Regulation Level

Feedback at self-regulation level is aimed to improve students’ self-monitoring skills and to address the way students direct and regulate their own learning. It seems that strategies of self-regulated learning play an important role when mastering the first year at university, as students are expected to learn much more independently than in high school. For many students it is also their first year away from home and managing the different demands of multiple courses and personal interests can be a big challenge. Zimmerman’s (1998) description of “naïve self-regulators”, who are performance orientated rather than learning orientated, often uninterested, monitor the outcome rather than the process and seem unable to adapt their learning strategies because failure is attributed to a lack of ability and not to a lack of effort, sounds too familiar. The question of what kind of feedback supports students’ development into self-regulated learners is addressed in this section.

In the CS1 context, where students explore the new territory of computer programming, feedback on the self-regulation level can help them to become confident and efficient software developers, who know the tools of the trade and how to approach the problems involved. If initial program designs are not successful, compiler errors frequently occur and program crashes persist, these hurdles should ideally be taken as challenges and should not cause frustration. If, in addition, students experience computer programming as exciting and rewarding and develop a genuine interest in improving their skills, than the best outcome that we could hope for is achieved. In this section some ideas for feedback on the self-regulation level which appear to be relevant in the CS1 context are introduced.

3.4.1 Multi-stage assignments: Increasing task complexity with fewer instructions

Multi-stage assignments are seen to be effective when “required standards are becoming increasingly transparent and dialogic feedback processes support students’ self-monitoring” (Carless et al., 2011, p. 398) and when single assignment tasks are spaced out during the course to ensure consistent study activity rather than a rush towards the end (Gibbs, 2006). Hernández (2012) found in a study, involving 138 undergraduate students, that the aspect of continuous assignments was indeed valued by students.

Multiple lab tasks, which are distributed over the semester, can be seen as a multi-stage assignment, where (dialogic) feedback at the completion of one task should also be aimed to guide improvements for the next task. Furthermore an increasing level of complexity forces the students to apply more and more advanced concepts and it is desirable that at the same time instructions become less prescriptive to foster students’ autonomy in understanding and implementing the requirements. However there is a dilemma: more advanced programming concepts (e.g. inheritance) require non-trivial program designs to be illustrative. Detailed instructions are needed to scaffold the rather complex program development,
but towards the end of the course minimal scaffolding should be needed to foster students’ confidence and a sense of ownership for their work. A scheme where instructions or tasks are at different levels for students to choose from (basic, intermediate, advanced) could support a sense of self-regulation. Students could be encouraged to challenge themselves by aiming for higher level tasks or instructions which require more self-directed planning and implementation.

As task instructions become increasingly less prescriptive in order to help students to develop into autonomous programmers, tutor interaction with students should also reflect this intention. Giving students answers to questions, which they could potentially work out themselves, is counter-productive in this sense. Targeted questions and hints could instead guide the process of finding the answer or solution to a problem. That is not always easy and compromises might be necessary when a busy lab situation requires quick responses to students’ programming problems or some students’ expectation to be “told what to do” collides with tutors’ encouragement to find the answer independently.

3.4.2 Student engagement through peer assessment and peer feedback

In the process of pair programming and peer feedback students need to articulate task requirements and relevant concepts, and by judging other peers performance students learn to monitor their own performance more accurately, which is an essential aspect of the self-regulated learning process (e.g. Butler and Winne, 1995).

Peer assessment and peer feedback are established ideas in CS (see Luxton-Reilly, 2009 for a systematic literature review of tools supporting peer assessment). Pair programming can be seen as one specific form of peer feedback where two programmers are in active dialogue about the task at hand. Advantages and disadvantages of pair programming for educational purposes and specifically in CS1 are widely discussed (for a short summary see Radermacher, Walia and Rummelt, 2012) and improved learning especially for weaker students has been reported (e.g. Braught, Eby and Wahls, 2007). Addressing the problem of “unbalanced pairs” in pair programming, Wood et al., (2013) published positive results based on pairing first year students with the same programming confidence level. These findings reiterate conclusions drawn from a review of 74 empirical studies (Salleh, Mendes and Grundy, 2011). Apparently sensible planning of the pairs and guidance throughout is important to get pair programming into fruitful action. For a comprehensive overview of studies using pair programming and important aspects of planning such an intervention see Salleh et al. (2011) or Mendes, Al-Fakhri and Luxton-Reilly (2006).

Peer instruction (e.g. Porter et al., 2011) and peer review (e.g. Smith et al., 2012) are also approaches to engage computer science students in the peer feedback process without the
pressure of grading. The effects reported are mainly positive. For example Chase and Okie (2000) published a study of the positive effect of combining collaborative classroom learning involving peer instructions on passing rates and retention in the context of a three semester CS1 course. The most significant improvement was reported for female students with a decrease of drop-out/failing rate from 53% to 15%. Porter, Bailey-Lee and Simon (2013) investigated course components using peer instruction regarding the impact on pass rates and found substantial improvements. An excellent review of tools to support student collaboration in CSEd is provided by Hamer et al. (2010). In this review, peer review tools were identified as being “(by far) the most prevalent systems described in the literature” (Hamer et al., 2010, p.5).

3.4.3 Opportunities for self-assessment

Self-assessment is seen as the key element in the process of self-regulated, life-long learning (e.g. Tan, 2007; Butler and Winne, 1995). If students are able to judge their own performance in relation to self-defined goals, further action can be taken towards accomplishing these goals.

There is an abundance of tools and materials to support self-directed study in computer programming: online tutorials, lectures and quizzes (e.g. provided by KahnAcademy (www.khanacademy.org) or Udacity (www.udacity.com)), code reading tasks (e.g. JavaGuide, in Hsiao, Sosnovsky and Brusilovsky, 2010) or code writing tasks (e.g. JavaBat, in Parlante, 2007). The integration of additional instant feedback to enhance sole error detection was investigated (e.g. Kumar, 2005) as well as advantages and disadvantages of instant feedback (Corbett and Anderson, 2001). If tasks are generated automatically by using parameters a virtually endless number of unique exercises can be created. Student generated MCQs and small-scale programming problems are supported by PeerWise (Denny, Luxton-Reilly and Hamer, 2008) and CodeWrite (Denny et al., 2011). The use of educational games to support introductory CS topics in schools and criteria to evaluate such games is the focus of an investigation by Gibson and Bell (2013). The authors identified 41 games for different platforms, but found the coverage of topics limited, with introductory programming concepts as one of the two main areas addressed.

A separate literature review would be necessary to classify and evaluate the tools available and their potential support of self-directed learning and self-assessment. Ihantola and Ahoniemi (2010) or Pears et al. (2007) provide a good starting point into the literature on automatic assessment tools for programming exercises. Integrating such resources for review or lab preparation might lead to higher engagement because of the interactive character and the instant feedback provided.
Providing regular opportunities to “critically evaluate the quality of their own work during, as well as after, its production” (Sambell, 2011, p. 22) is one strategy to foster self-regulation for undergraduate students. In a study by Robinson and Udall (2006) it was shown that interactive “learning conversations” for engineering students lead to a greater sense of ownership in students’ learning. As already mentioned in Section “3.3.1 Programming process”, a critical reflection on a finished programming task combined with tutor feedback could support improvements for the next task, but could as well be useful to get students into the habit of critical self-reflection. Such wrap-up might be guided by questions touching on different aspects such as functionality, robustness, aspects of good programming style or on the overall final design of the program.

3.4.4 Self-monitoring course performance

At course level students are potentially able to self-monitor their progress as the course progresses, if their attendance, exam marks and other information about completed assessment are made available. However, the implications for not meeting course demands such as attending practical sessions, finishing assessment tasks or achieving a mid-semester examination score over a certain level may be rather unclear. The process of self-monitoring performance could be enriched with information relevant to individual students and triggering self-directed action by answering questions such as: What are my prospects for the final examination based on my mid-semester exam score? How many practical tasks do I need to complete to have a good chance to pass the final examination? Is there a correlation of timeliness in submission and examination performance? When should I seek additional help based on my course performance?

Considering the impact of the Learning Edge Momentum as postulated by Robins (2010) for CS1 courses, students falling behind early on are unlikely to catch-up later in the course. The integrative nature of the course material leads to a spiraling effect on students’ consecutive performance. It can be assumed that basic concepts that are not fully understood cause trouble with comprehending higher-level concepts. Communicating such typical course characteristics we know about, even if only from anecdotal evidence, could help students to understand the factors influencing their performance and act accordingly.

Adding meaning to naturally occurring course data would inform students about their prospects in the course. Based on individual performance data, feedback could be personalised by adapting performance goals (e.g. goals to catch up or attend labs more regularly), relating students’ actual performance to what was observed as “successful” performance in the past and pointing to aspects to improve. Our search for literature relating to feedback on self-regulation level and supporting the self-monitoring of course performance data did not yield any substantive results. While the predictive value of those data might vary a lot
between the disciplines, programs and institutions, the question, if feedback on this level is effective to influence students’ study behaviour positively, remains.

### 3.4.5 Summary

Feedback on the self-regulation level is well regarded in the higher education literature as being important for students’ development towards becoming self-regulated, resourceful, life-long learners. In the previous sections matters of multi-stage assessments to engage students in critical reflection and self-regulation, opportunities for self-assessment on task, concept and course level and different forms of peer feedback have been discussed in the CS1 context. Given the importance of fostering self-regulated learning, more high quality feedback on this level is needed.

Only if students understand the course mechanics and can derive their own performance goals as well as judge their actual performance in relation to these goals, is self-regulation at this level possible. If naturally occurring course data were found to be predictive for students’ final performance in the past, those data could be used to assist the definition of course performance goals and the translation of actual course performance data into meaningful and relevant feedback information. For a short summary of related topics to improve feedback on self-regulation level discussed in this section see Table 3.3. Boundaries between goals-, performance- and improvement-specific feedback are less apparent on this level.

<table>
<thead>
<tr>
<th><strong>Table 3.3: Feedback on self-regulation level</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-regulation</strong></td>
</tr>
<tr>
<td>(1) <em>Multi-stage assignments</em>, (2) <em>Peer feedback</em>, (3) <em>Self-assessment</em>, (4) <em>Self-monitoring course performance</em></td>
</tr>
<tr>
<td><strong>Goals</strong></td>
</tr>
<tr>
<td>Using multi-stage assignments to make standards for task completion successively more transparent.</td>
</tr>
<tr>
<td>Establishing peer feedback as opportunities to communicate task criteria and standards of good programming practice as well as concepts involved.</td>
</tr>
<tr>
<td>Communicating course characteristics as seen in the past to inform students about patterns of successful and risky study behaviour upfront.</td>
</tr>
<tr>
<td><strong>Actual Performance</strong></td>
</tr>
<tr>
<td>Introduction of pair programming to engage students in a dialog about the task requirements and the actual solution.</td>
</tr>
<tr>
<td>Incorporating and promoting resources for self-assessment such as online MCQ or programming quizzes frequently and in relation to new concepts introduced.</td>
</tr>
<tr>
<td>Introducing multi-stage assignments and scheduled dialogic feedback to improve students’ self-regulation on task and process level.</td>
</tr>
<tr>
<td>Providing individual course performance data and highlighting implications based on historical course data to support the process of self-monitoring.</td>
</tr>
<tr>
<td><strong>Activities for Improvements</strong></td>
</tr>
<tr>
<td>Engaging students in self-critical conversations to foster the sense of ownership of their learning.</td>
</tr>
<tr>
<td><strong>Communication in general</strong></td>
</tr>
<tr>
<td>Using increasing task complexity and decreasing scaffolding to support self-directed task completion.</td>
</tr>
</tbody>
</table>
3.5 Summary

In this chapter practical opportunities for applying the principles of good feedback, as established in Chapter 2, were highlighted by describing common aspects of a typical CS1 course and related areas for possible improvements on task, process and self-regulation level. This discussion not only illustrated how those level and questions guiding effective feedback practices can be considered in the CS1 context, it also structured a preliminary survey of the CSEd research literature by pointing to various topics supporting good feedback practice. Further reviews of the literature would be required to plan and conduct feedback interventions for different parts and stages of a CS1 course.

Feedback on task level was reviewed for three course components: (1) programming tasks to complete in tutor-assisted lab sessions or as extended assignments, (2) small tasks during the lectures and (3) examinations. Where excellent, task-specific feedback can be provided in a tutor-assisted lab, which we consider as common in the CS1 context, we found feedback practices regarding examination results neglected in our own course and a missed opportunity to guide students’ learning. Based on students’ actual performance, clear advice on which topics need to be revised should accompany the examination score. This reflection on feedback processes already lead to significantly improved feedback on mid-semester examination results in our course where students in 2013 were provided with written, individualised feedback outlining topics and resources for review. Feedback sheets were automatically generated for nearly 200 students, based on scores for different parts of the examination. Students’ usage of the review recommendations and the impact on their final results is discussed as part of Chapter 7, Section 7.6.3.

On process level, feedback was discussed for two different processes CS1 students need to master: (1) completing programming tasks and (2) meeting the course requirements. A discrepancy was revealed between our expectations on how students should manage the programming process (in terms of self-directed error detection and correction as well as planning and testing a program solution) and missing, explicit goals as well as structured feedback to guide these processes. Model solutions, educational programming environments and reflection tasks were considered to implement process-orientated feedback targeting general aspects of good programming practice and generic skill development. Clear communication of expectations and goals to establish effective feedback was also considered as vital for processes regarding the management of course requirements such as lecture or lab attendance, task completion and passing the examinations.

Feedback to support students’ self-regulated learning plays a vital role in the feedback process. Whereas solving programming problems is in the core of an introductory programming course and feedback on self-regulation level can be assisted by multi-stage assign-
ments, peer feedback and self-assessment, managing course demands was revealed as another challenge students are faced with. Ideally students work towards the course requirements by setting personal goals of achievement levels, judging their actual performance regarding these goals and adjusting their study strategies if necessary. Prosser & Trigwell (1999) argue that students’ learning approaches are highly influenced by their prior experiences with the subject matter as well as their perceptions of the learning environment. Consequently the question is how can we support first year CS students, who are most likely unfamiliar with the demands of studying at university level in general and the subject matter of programming specifically?

We suggest that an important part of the answer is to provide students with information highlighting risk factors as well as successful study strategies, thus enabling them to self-regulate their learning on an informed basis. In the next chapter the predictive value of course data available for COMP160 is investigated and performance indicators based on various data sets are derived, which can be used to illustrate patterns of successful and risky study behaviour.
Chapter 4

Investigating the predictive value of available course data

“We believe that learning to program is problematic, and that the results achieved by students in programming courses do not correlate well with their other academic results. Our understanding of this phenomenon is patchy and poorly integrated, but it does seem clear that there are many influences at play.”

(Simon et al., 2006, p. 2)

This chapter is based on section 2 of a paper submitted to Computer Science Education that is currently under revision:

4.1 Introduction

Only if students understand the course mechanics and can judge their actual performance in relation to their course performance goals, are they in a position to self-regulate their learning. In this chapter we propose that naturally occurring course data are predictive for students final performance and that they can be used to assist the definition of course performance goals and the translation of actual course performance data into meaningful and relevant feedback information.

Based on six units of analysis, available for our course, which is considered to be a typical CS1 course, the predictive value of naturally occurring course data is investigated in order to identify patterns leading to students’ later success or failure in the course. The research questions can be formulated as:

- Which performance indicators can be identified?
- What indicates risky/successful study behaviour?
- Can students at risk be identified early in the course on naturally occurring course performance data only?

In Section 4.2 we give a brief summary about the literature regarding predictive factors of students’ failure or success in CS1. In Section 4.3 the units of analysis are introduced and their varying predictive power is discussed. The use of the performance indicators derived from the units of analysis to identify students at risk is discussed in Section 4.4 and leads to implications for the feedback process in Section 4.5.

4.2 Predictive factors in the literature

The search for predictors of students’ success or failure in programming courses is well covered in the literature. For example, factors such as math background and comfort level in class (e.g. Wilson and Shrock, 2001; Bergin and Reilley, 2005), learning approach (Simon, Lister and Fincher, 2006), final grade expectations (Rountree, Rountree and Robins, 2004) and the degree of self-regulated learning (Bergin, Reilly and Traynor, 2005) have been found to be predictive for students’ failure or success in CS1 courses. More specific to the discipline, compilation errors and students’ strategies for fixing those errors were investigated in order to estimate the mid-term exam scores (Tabanao, Rodrigo and Jadud, 2011). Even though the reappearance of certain error types and a high frequency in recompilation was found to be a negatively correlated with the mid-term score, an accurate prediction of students’ scores was not possible.
The investigation of the predictive value of other naturally occurring data sets got comparable little attention. Fenwick et al. (2009) found, while also investigating students’ compile behaviour, that students starting late with their assignment had a lower success rate. Falkner and Falkner (2012) based their study on 220,000 records of 1900 students and concluded that submission behaviour early in the course can be linked to future performance. Students who submit their first assignment late are likely to establish a consistent pattern of lateness in the following years.

The use of individual performance data requires an effective Learning Management System (LMS) to make those data accessible for students and teachers. Rößling et al. (2008) published a report on how to enhance LMSs to better support computer science education. The proposed example scenarios were primarily focussed on learning content and the integration of aspects specific to the discipline such as IDEs and algorithm visualisation. We would like to extend the notion of LMSs towards the integration of naturally occurring course data into the feedback processes given that those data are predictive for students’ final performance, as presented in the following section, and can be used as performance indicators on an individual basis. However such usage scenarios of LMSs were not found in the CSEd literature.

4.3 Performance indicators in COMP160

The course investigated is a first year Java introduction (COMP160) with about 200 students enrolled in the second semester each year. Results are typically bi-modally distributed with around a third of students finishing with A-grades and around the same number failing or dropping the course during the semester. As described briefly at the beginning of Chapter 3, the course is offered in a typical combination of lectures (50 min twice a week) and tutor assisted lab work (2 hours twice a week). Students are expected to submit one completed lab task per lab session. Usually there are 20 to 25 students in a lab session and two to three tutors are present to help students on request. Students can work at their own pace, but are advised to try and keep to schedule. If students work from home, they need to hand-in the completed programming task in person during lab time. The 25 lab tasks add up to 25% towards the final grade with no penalty for late submission. Written examinations at mid-semester and the end of the semester contribute another 15% and 60% respectively. The passing criterion to achieve 50% or more in the final examination makes this component an important part of the course.

The six units of analysis available discussed in this section are correlated with the final examination score. We will report the Pearson correlation coefficient $r$ using two-tailed probabilities considered correlations as significant on 0.01 level. The final examination score has been chosen as the best available measure of students’ actual learning outcome.
because it is the only measure taken under controlled conditions and independent of internal marks such as completed lab work or mid-semester score, which are part of the units of analysis investigated. Furthermore we consider the final exam score as the most accurate measure of students’ actual programming ability, which is naturally available.

To support this assumption an investigation was carried out to evaluate how well the final exam score reflects the tutors’ impression of students’ programming ability in terms of help required, nature of questions asked and quality of programming solutions. For the final 12 lab sessions in 2011 (second half of the course) tutors were asked to tick a 4 point-scale representing a letter grade (A, B, C, D) at the time of task submission to rate their impression of the student’s programming ability. Of the 163 students attending labs in the second half of the course and sitting the final exam, 141 students got 3 or more tasks rated by a demonstrator. The average rating was strongly positively correlated with overall final exam score (r=0.70, p<0.01, N=141). The correlation improves with the number of ratings (e.g. for students having 5 and more ratings: r=0.80, p<0.01, N=100). The inter-rater reliability was not tested because of missing data, but given the results it is possible to conclude that the final exam results in fact reflect the tutors’ judgment of students’ programming ability.

Naturally occurring course performance data were collected throughout the second semesters of 2011 and 2012 based on 386 student records (the structure and content of the course remained stable over these years). These data include pre-course grades, lab attendance, submissions and mid-semester exam scores. Login data were used to measure time spent in the labs. No information was gathered about the actual quality of submitted programming tasks, which are marked on a pass / fail basis for awarding 1% per lab (25% in total) towards the final grade.

### 4.3.1 Indicator: Pre-course grades

Around 65% of students taking our course in the second semester have attended an introductory Python course in the first semester (COMP150). This course is not a pre-requisite and students can enrol for the second semester course even though they did not take or pass the Python course. As both courses can be considered as introductory programming course we expected to see moderate to strong correlations.

For the data of year 2011 the analysis showed significant, moderate correlations between students’ final grades in the introductory Python course and their final exam score in our course (r=0.55, p<0.01, N=107). In 2012 the correlation was stronger (r=0.71, p<0.01, N=115). However this early indicator is only available for two thirds of the students and further analysis into students’ grades of other courses offered in first semester needs to be attempted in the future.
4.3.2 Indicator: Attendance of lab sessions

Attendance records are kept for the lab sessions. When looking at the attendance data of 2011 for three different groups of students based on their final mark (A-grade students, B-grade & upper C-grade students, lower C-grade and failing students) we observed a continuous attendance pattern for the top end students, while the other two groups -- especially the lower C-grade and failing students -- were more likely to drop sessions during the semester, but would show up shortly before the examinations. Given that observation, we expected to see correlations between students’ attendance and the final examination result.

By looking at the attendance data by week 4 and 5 (first 7 and 9 lab sessions) we hoped to find an early indicator for students’ final performance. However these data were only very weakly correlated. The overall attendance (25 lab sessions) was significantly, but still weakly correlated (r=0.22, p<0.01, N=165). An analysis of students’ attendance in the first 5 weeks with their attendance for the rest of the semester indicated significant, moderate correlations (r=0.489, p<0.01, N=173) and it can be concluded that students who missed sessions early on are also likely to miss sessions later.

Our data does not indicate why students have been absent and two scenarios are possible: (1) students are working at home and handing in the work at a later stage or (2) students are missing lab sessions because of a lack of motivation, a high workload or other distracting reasons. The first scenario is not a reasonable explanation for students absent from the first lab session as they need to collect a lab book before they can start. A closer look at attendance data for the first lab session revealed that students missing out on the first lab session were either not sitting or not doing as well as their classmates in the final exam. An unpaired T-test showed a significant difference between the group of students who attended the first lab session (M=60.8, SD=27.9, N=160) and the group of students who did not attend the first lab session (M=42.3, SD=23.7, N=24; t(182)=3.0725, p=0.002) in 2011. Data from 2012 produced comparable results.

The dropout rate for all students in 2011 and 2012 who did not attend the first lab, but submitted at least one lab task, was 25% (12 of 51). In contrast only 10% (33 of 334) dropped the course when they have been to the first lab session. It can be concluded that students not attending the first lab session are at higher risk to fail the course either by dropping out or because of a lower final exam result. In fact almost half of the students not attending the first lab session scored under 50% in the final exam compared with a quarter of the other group.

The reasons for these observations are speculative. It is possible that the practice and information provided in the first lab session are of some importance for later success. How-
ever, it is also possible that students, missing out the first session, lack essential motivation, the skill to organise their schedules or vital persistence to succeed in the course.

4.3.3 Indicator: Number of submitted lab tasks

An electronic submission system provides detailed information about time and task number of submission. In most cases students submit their work to the department’s server after talking to a tutor. Therefore the time of submission is in most cases the time when students finished their work. We suspected that lab tasks handed in at different stages of the course are good predictors of students’ final performance in the exam and expected to see correlations as early as by week 5, because this week challenged students with two of the hardest task of the entire course (based on students’ ratings of task difficulty).

In fact a moderate correlation was seen already by the end of week 5 ($r=0.467$, $p<0.01$, $N=165$) developing towards stronger correlation by week 11 ($r=0.608$, $p<0.01$, $N=165$) for the year 2011. Results for 2012 have been similar for week 11, but only weakly correlated for week 5 ($r=0.369$, $p<0.01$, $N=174$). The final number of submitted lab tasks per student was not a better indicator than the number by week 11, which is two weeks before the last lab session. This is due to the fact that students, who have fallen behind with the lab work, try to catch up during the last week. So the numbers by week 11 better differentiate between students who have been working consistently and the ones who have not, but might end up with the same final number of lab tasks.

The assumption that this kind of “binge” submission behaviour in the last week (individual students have submitted as many as 13 tasks in that final week) correlates negatively with success in the final exam could not be proven to be statistically significant.

4.3.4 Indicator: Time of submission

Analysing the time of submissions provides additional information. Late submission was suspected to indicate students struggling with the lab work as early as week 5. The average lateness (a negative number for late submissions defined as weeks per submitted task) correlated moderately with students’ final exam score ($r=0.56$, $p<0.01$, $N=165$) in 2011. Results were almost identical in 2012. By the end of week 5, the average lateness was significantly, but weakly correlated in 2011 ($r=0.26$, $p<0.01$, $N=166$) and turned out to be not statistically significant in 2012 ($p=0.09$).

We also tested the hand-in date of particularly hard tasks as we assumed that those tasks differentiate between students facing problems with these tasks and the students which encounter no major problems. Students rated lab task 9 in week 5 as the most difficult task of the course based on 275 student ratings for this task collected via an online survey in
2001 and 2012. A test in 2011 showed that the lateness attached to this task correlated moderately with students’ final exam results \((r=0.46, p<0.01, N=139)\). The fact that students were on time with this particular lab task suggests that those students had no difficulties with the subject matter so far. In fact only 13\% (17 out of 134) of the group who handed lab 9 in on time scored under 50\% in the final exam, as compared to 33\% (55 out of 165) of those who handed lab task 9 in late and 52\% (23 out of 41 students) who did not hand in this task at all. The results of an unpaired T-test for 2011 showed significant differences in students final exam results between the group who was late or failed to submit \((M=57.4, SD=20.0, N=107)\) and the group who handed this lab in on time \((M=77.9, SD=17.3, N=59)\); \(t(164)=6.6, p<0.001\). The T-test of the 2012 data also resulted in a significant difference between the groups.

### 4.3.5 Indicator: Time spent in labs

One might assume that spending more time in the labs leads to better results in the final exam. That is not the case in our course. Based on students’ login times we investigated the overall time students spend in the labs and the average time spent per completed task. Time information can be extremely noisy as we do not know what students actually did in the labs, but we can assume that during the daytime and especially during scheduled lab sessions, students would mainly work on their tasks. Students working at home also add to the distortion of the data set.

Both units of analysis were significantly and negatively correlated with students’ final exam results. The best indicator was the average time per completed lab tasks, which was moderately correlated in 2011 \((r=-0.49, p<0.01, N=155)\) whereas the time in total turned out to be weakly correlated \((r=-0.30, p<0.01, N=161)\).

### 4.3.6 Indicator: Mid-semester exam result

The mid-semester exam result is by far the best indicator of students’ final exam results. Strong correlations in 2012 \((r=0.78, p<0.01, N=174)\) and very strong correlations in 2011 \((r=0.85, p<0.01, N=165)\) have been observed. Going further back to the data of 2010, 2009 and 2008, similar correlations between the mid-semester score and the final exam score can be observed.

It seems that by mid-semester we already know, with some certainty, which students will score well on the final exam and which ones probably fail. This information plays an important role when considering feedback interventions that would help students to understand their actual progress in relation to their final goal.
4.4 Discussion of risk factors

Looking at the single performance indicators in 2011 we tried to quantify early warning signs by using only the indicators, which are available by week 5. Students are at higher risk to fail if they show one of the following warning signs:

1. *Finished the pre-course with a mark under 60%*
2. *Did not attend the first lab session*
3. *Missing three or more submissions (out of 9) by week 5*

In 2011, 68 students showed one or more of these early warning signs. 59% (40 students) of these students would in fact not sit the final exam (12 students) or fail the final exam (28 students). For the remaining group of 116 students not showing any warning signs, 15% (18 students) students did either not sit the final exam (6 students) or failed it (12 students).

If the early warning signs are combined with a fourth indicator, available half way through the course, predictions of failing become more accurate:

4. *Scored under 60% in the mid-semester exam*

This indicator showed up for 43 students and 29 students in this group were already indicated with an early warning sign. Only 8 students, who were unable to complete the course successfully, remained unidentified. 33 students were identified to be at risk, but completed the course successfully. An unpaired T-test showed a significant difference between students final exam results for the group of incorrectly indicated students (M=62.7, SD=10.6, N=33) and the group of students who had no indication of being at risk (M=78.1, SD=13.5, N=93; t(124)=-5.9172, p<0.001). The mean difference of 15.4 indicates that the group of incorrectly indicated students scored lower in the final exam by one letter grade (B or C) when compared with the rest of class passing the final exam.

In summary, 70% (40 out of 58) of the students scoring under 50% in the final exam or not sitting the final exam could have been already identified to be at risk by week 5 (184 students initially active, i.e. submitted one task or more). 22% (28 out of 126) of the successful students would have been incorrectly identified. By mid-semester (176 students still active) these predictions would have been improved, indicating 84% of the students not completing the course successfully (42 out of 50) although with a rate of 26% (33 out of 126) incorrectly identified. The results are summarised in Table 4.1.
Table 4.1: Predictions for 2011

<table>
<thead>
<tr>
<th>2011 Week 5</th>
<th>At Risk Sign</th>
<th>No At Risk Sign</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed Final Exam</td>
<td>28</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>Didn't Sit Final Exam</td>
<td>12</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Passed Final Exam</td>
<td>28</td>
<td>98</td>
<td>126</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>68</strong></td>
<td><strong>116</strong></td>
<td><strong>184</strong></td>
</tr>
</tbody>
</table>

χ²=37.31, p<0.001

<table>
<thead>
<tr>
<th>2011 Mid-Semester</th>
<th>At Risk Sign</th>
<th>No At Risk Sign</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed Final Exam</td>
<td>35</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>Didn't Sit Final Exam</td>
<td>7</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Passed Final Exam</td>
<td>33</td>
<td>93</td>
<td>126</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>75</strong></td>
<td><strong>101</strong></td>
<td><strong>176</strong></td>
</tr>
</tbody>
</table>

χ²=49.92, p<0.001

An application of the same model of combining the three early indicators and the mid-semester exam indicator on the data from 2012 resulted in a lower detection rate but also in a lower rate for students identified incorrectly. Please refer to Table 4.2 for a summary.

Table 4.2: Predictions for 2012

<table>
<thead>
<tr>
<th>2012 Week 5</th>
<th>At Risk Sign</th>
<th>No At Risk Sign</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed Final Exam</td>
<td>26</td>
<td>28</td>
<td>54</td>
</tr>
<tr>
<td>Didn't Sit Final Exam</td>
<td>20</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>Passed Final Exam</td>
<td>19</td>
<td>101</td>
<td>120</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>65</strong></td>
<td><strong>136</strong></td>
<td><strong>201</strong></td>
</tr>
</tbody>
</table>

χ²=42.60, p<0.001

<table>
<thead>
<tr>
<th>2012 Mid-Semester</th>
<th>At Risk Sign</th>
<th>No At Risk Sign</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed Final Exam</td>
<td>41</td>
<td>13</td>
<td>54</td>
</tr>
<tr>
<td>Didn't Sit Final Exam</td>
<td>16</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>Passed Final Exam</td>
<td>25</td>
<td>95</td>
<td>120</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>82</strong></td>
<td><strong>113</strong></td>
<td><strong>195</strong></td>
</tr>
</tbody>
</table>

χ²=57.64, p<0.001

4.5 Implications for the feedback process

One way to use this information is to intervene by offering additional tutorials or other forms of help as soon as possible. Seeing the correlation of the submission numbers by week 5 with the final exam score strengthens the theory of the Learning Edge Momentum, proposed by Robins (2010): “Once negative momentum is established it is very hard to overcome. Ideally positive momentum should be established right from the start.” (p. 40). Robins emphasises the importance of the first weeks for the final learning success. He suggests the administration of diagnostic questions by week 3 and 5, but naturally occurring course data are already predictive by week 5 and can be used as feedback for teachers.

Another way would be to communicate performance indicators as information for students to support their self-regulated learning: The self-regulated learner as described by Zim-
merman (1998) defines goals, selects strategies and monitors progress in a self-directed process, but on what basis are those decisions and judgements possible? How can we help students to understand what defines “good performance” at different stages of the course, to derive performance goals, self-monitor their actual performance in relation and adjust strategies accordingly? Here the communication of risk factors and early warning signs is vital information to enable students to judge course demands and actual progress realistically. In our course, for example, a mid-semester exam result of 8 points out of 15 towards the final mark might seem to be a good achievement because it is the equivalent of 8 completed lab tasks, but in fact it indicates a danger of failing the final exam. Such misunderstandings can be avoided if performance indicators and projected consequences are transparent for students.

Understanding the correlations of certain performance indicators with the final achievement implies accepting the fact that good or excellent final results are not achieved by chance. Looking at the performance levels of successful students can convey performance goals: 82% of the A-grade students have been on track with the lab work or at most one task behind by week 5. 86% of these students scored over 80% in the mid-semester exam and around the same percentage have been on time or only slightly late with the lab work in general. The same data is useful to illustrate the consequences of current study behaviour by showing on an individual basis how a student’s actual performance would have been correlated with final achievement in the past. Areas of improvement are revealed when these projections are unsatisfactory. For example, if the mid-semester score was under a critical level, the course material needs to be revised in order to improve the chances to reach the final performance goal. However, there is no guarantee that being on time with the submissions leads to a better performance; neither can we be sure that submitting all tasks secures a high grade. The indicators do not explain causality, but it is likely that students, more engaged with the lab work, improve their understanding of the subject matter.

4.6 Summary

The analysis of naturally occurring course data for 2011 and 2012 revealed patterns, which indicate later success or failure in our course. Most of the observations support common sense assumptions. The performance indicators identified, include the final mark of an introductory Python course in first semester (COMP150), which correlates strongly with the final exam result in our second semester course. Submission numbers by week 5 and week 11 are also predictive for students’ final outcome, but the most accurate performance indicator is the mid-semester exam score. Less obvious indicators have been identified by the time students spend in the labs per submitted task, which is negatively correlated with their final exam result. Considering the common structure of the course we expect pre-course grades, early struggle with the lab work, mid-term test results and time on task to be
general indicators for students’ final performance in CS1 courses even though the exact units of analysis and their predictive value will vary between institutions and courses.

Given the correlations of the various data sets with students’ final examination result, we conclude that success in the course is highly influenced by the persistence to stay on top of the lab work. Students missing out on the first lab session are already on higher risk to drop the course or fail the final exam. A continuous pattern of being behind with the lab work in terms of not completing lab tasks or being late with the lab work indicates lower probability of a good grade. It seems that catching up later is difficult as 82% of the A-grade students have been mainly on track with the lab work by week 5 and as the number of completed lab tasks it already predictive for students’ final outcome at this stage.

Based on those observations we can conclude that naturally occurring data can be used to identify students at risk by week 5 or shortly after the first half of the semester and reveal patterns for successful or risky study behaviour. An application of a model of risk factors on the 2011 cohort identified 70% of the students not sitting or failing the final exam correctly by week 5 based on their pre-course (COMP150) result, their attendance in the first week and their number of submitted lab tasks by week 5. By incorporating the mid-semester examination score this predictions became more accurate and 84% of students failing the course were predicted correctly.

If students are expected to self-regulate their learning, course characteristics and performance indicators need to be communicated. From this information students would be able to derive performance goals. Self-monitoring their own performance in relation to these goals can be assisted by linking students’ actual performance with a projected final achievement as seen in the past. Understanding the projected consequences of certain study behaviour would help students to identify areas of improvement. We conclude that naturally occurring course data needs to be made available to support students’ self-regulated learning and extending existing feedback practice on task and process level. The development of an approachable and engaging format to communicate course characteristics and performance indicators is in the focus of the next chapter.
Chapter 5

Communicating course characteristics and performance indicators

“The graphical method has considerable superiority for the exposition of statistical facts over the tabular. A heavy bank of figures is grievously wearisome to the eye, and the popular mind is as incapable of drawing any useful lessons from it as of extracting sunbeams from cucumbers.”

(Farquhar and Farquhar, 1891, p. 55)

This chapter is based on section 3, 4, 5 and 6 of a technical report submitted to the Higher Education Development Centre, University of Otago:

5.1 Introduction

Close investigation of naturally occurring course performance data in COMP160 over the last two years demonstrated that information such as students’ lab task submissions, pre-course grades and mid-semester examination results are predictive for the final outcome. The integration of course performance data into the feedback process to support self-regulated learning and to make students aware how actual study behaviour influences the course outcome was implemented as a scoping study for the second semester in 2013 to assess the effectiveness and students’ perception regarding feedback on this level.

This chapter describes the development of an infographic for the COMP160 lab book to communicate course characteristics and to indicate students’ final performance based on actual performance at different stages of the course. The related question can be formulated as:

- **How to effectively communicate course characteristics and performance indicators?**

The following Section 5.2 provides a short historical background and a motivation for the use of infographics to communicate complex data as graphical visualisations. Section 5.3 establishes the rationale and design objectives and introduces the infographic prototype, which was evaluated. Method and process of the evaluation are described in Section 5.4. Section 5.5 provides a summary of the evaluation results and on how design issues have been addressed in the final version of the infographic.

5.2 Why Infographics?

“Infographics” as one form of data visualisation might be a contemporary term but refers to an old idea. Already in the 1850’s there have been excellent examples on how complex data can be turned into graphical representations, which are compelling because of their clever way to illustrate various dimensions. One of the famous examples is Charles Joseph Minard’s “Carte figurative des pertes successives en hommes de l’Armée Française dans la campagne de Russie 1812-1813” from 1861 (in Tufte, 2001) capturing five changing variables in a single two-dimensional image to visualise the factors contributing to Napoleon’s downfall in Russia (Appendix A):

1. the diminishing size of the army
2. the geographical coordinates of the army as it moved
3. the direction that the army was travelling
4. the location of the army regarding certain dates
5. the weather temperature along the path of the retreat
Whereas Minard’s graphic was developed to show the order of magnitude of human suffering and losses during an historical event, Florence Nightingale (1855) used statistical data about causes of death in a "Diagram of the causes of mortality in the army in the East" (Appendix B) to persuade the British government that sanitary reforms are necessary (Brasseur, 2005). The diagram shows that more soldiers died of preventable diseases (blue areas) than of their wounds (red areas) in the Crimean War between April 1854 and March 1855 in a way which made the data easy to grasp for people not familiar with traditional statistical reports.

This example shows how visualisation can turn complex data into accessible and persuading information. Stephen Few (2009) emphasises on the immediate goal of data visualisation as understanding the data, but with the end goal of coming up with better decisions. This idea forms the rationale for the design of an infographic for the COMP160 lab book: providing students with illustrative information to understand course demands and performance indicators at different stages of the course, which might lead to better decisions regarding their commitment and learning efforts in the course. But what kind of information should we put forward and how should we present it? This question is addressed in the next section.

5.3 Infographic development

This section described the objectives, which guided the design process of the infographic. Based on teaching experience in the past and the analysis of course data from 2011 and 2012, as introduced in Chapter 4, decisions were made about the different types of information to include. The overall design objective, to present the information in an engaging, easy to comprehend layout, led to a split into four different sections, which are loosely connected but can also be considered on their own. Labels for each section were introduced to support a quick access to the different sections: (1) “Key events”, (2) “What students say”, (3) “What we say”, (4) “What the numbers say”. In the following, the design objectives for each of the four sections are introduced as well as the resulting design solution (see Appendix C for a complete printout).
5.3.1 Section “Key Events”

The first section was designed to provide a quick overview about the course structure by:

- **Identifying key events:** Major events are the two examinations at mid-semester and the end of the course. Furthermore students should be aware when practice exams are made available prior to the examinations and when personal feedback data, as relevant for the infographic, are provided.

- **Explaining the composition of the final grade:** 25% for submitted lab work (1% for each of the 25 tasks) + 15% for the mid-semester exam + 60% for the final exam.

- **Clarifying the course structure:** Two lectures and two laboratory sessions per week.

**Solution:** The first section of the infographic, labelled “Key Events” (Figure 5.1), resulted in a calendar-like layout, highlighting the key events as flags. The kind of event is represented by the colour of the flag, e.g. yellow for assignments and blue for availability of feedback data. Assignment flags include information about the internal mark towards the final grade. The two weekly lab sessions are not visible in this section, but shown in the next section which still follows the timeline layout.

![Figure 5.1: Section “Key events”](image)
5.3.2 Section “What students say …”

The second section was laid out to share former students’ perception of the course and was guided by the design objectives of:

- **Showing rapidly increasing task difficulty:** In the past it was observed that students underestimate how quickly the task difficulty increases. Especially students coming from COMP150 in the first semester find the initial labs of COMP160 easy, which leads to an underestimation of the course demands. Indicating particular hard tasks would be a good preparation and could trigger a sense of achievement.

- **Preparing students to get emotional:** Based on observations and students’ comments we know that the lab work can be perceived very differently. At times students get very frustrated but at the same time students also very much enjoy the challenges.

**Solution:** The second section, “What students say …” (Figure 5.2) follows the timeline layout of the first section and is based on students’ difficulty ratings and comments for each of the 25 lab tasks. 3551 difficulty ratings and 429 comments were collected in 2011 and 2012 via an online survey, which popped up when students submitted their tasks. A graph was composed rendering the average difficulty rating for each lab task. Selected students’ comments regarding the tasks are mapped onto the graph to illustrate that students in the past experienced the work as challenging, frustrating and time consuming, but also reported on having fun and feeling rewarded after task completion. The comments selected are intended to show the wide spectrum of possible emotions and to share former students’ experiences in an authentic voice.

![Figure 5.2: Section “What students say …”](image-url)
5.3.3 Section “What we say …”

To balance students’ perspectives the third section was introduced to share important information from the teachers’ perspective and introduce necessary explanations for parts of the infographic. Here two design objectives need to be met:

- **Providing course details and bits & pieces of advice:** Similar to an FAQ section there are details, which should be passed on to students to answer questions such as:

  - Which lab tasks are particular important for the mid-semester exam?
  - What is special about lab 17?
  - Why is it important to attend the lectures?
  - Can I fail the course only because of the final exam?

- **Allow for explanations regarding infographic elements:** Some aspects of the infographic might profit from additional explanation to strengthen the point in focus.

**Solution:** Section “What we say …” (Figure 5.3) provides a space to get students informed about important aspects from a teacher’s perspective. The section is also intended to connect the previous section of students’ perceptions with the following section by referring to different aspects in these sections, which might need further explanation. Markers (lines with red dots) are used to point to the related areas (visible in the neighbouring sections).

![Figure 5.3: Section “What we say …”](image-url)
5.3.4 Section “What the numbers say …”

The last section was reserved to communicate selected performance indicators as discussed in Chapter 4. Multiple dimensions of the performance indicators needed to be visualised in a way easy to understand for the majority of students. The design objectives guiding the development of this section can be summarised as:

- **Introducing course performance indicators for different stages of the course:** Sharing course information from 2011 and 2012 to illustrate that students’ final exam score was correlated with their:
  - Pre-course grades (COMP150),
  - Number of tasks by week 5
  - Number of tasks by week 11
  - Timeliness of their submissions
  - Mid-semester score

- **Visualising the correlation for different groups of students (per performance indicator):** For example students who handed in “8 tasks by week 5” form one achievement group, students who handed in “9 or more tasks by week 5” form another group. The correlation with the final exam score per group should be illustrated in different ways, e.g. (1) the vertical position (y-coordinate), (2) the average score of all students in the group in the final exam and (3) the actual range of the scores in the final exam.

- **Highlighting groups in danger to fail the final exam:** Achievement groups with an average score under 50% in the final exam (high chance of failing the final exam) should be clearly visible, so that risk factors are easy to grasp upfront (e.g. “under 7 lab tasks by week 5”). On the other hand groups for students with a high chance to pass the exam on a top level should also be clearly visible to illustrate successful study behaviour and support self-directed goal setting.

- **Providing a possibility to “map” actual performance:** Students should be able to “map” their own performance by identifying their actual achievement group they are in for each performance indicator at different stages of the course.
Solution: Section “What the numbers say …” (Figure 5.4) was developed as a bubble chart. 4 to 6 bubbles are used for one performance indicator and are arranged close to each other. A single bubble represents a group of students on the same actual achievement level. For each bubble the following information is visible and explained in a legend at the bottom of the infographic:

- The *label* of the bubble describes the actual achievement level (above the line, e.g. “8 Tasks”) and the average score in the final exam (percentage below the line, e.g. “66%”).

- The *box-plot* overlay for each bubble represents the actual range of final exam performance for this group.

- The *y-position* of a bubble is mapped regarding the average final exam score.

- Different *bubble colours* are used (from green to red) to indicate decreasing final performance.

- The area representing the “danger zone” (a score under 50% in the final exam) is shaded red to quickly identify groups at risk.

- The *size of a bubble* represents number of students in the actual achievement group.

Figure 5.4: Section “What the numbers say …”
5.4 Evaluation method

To evaluate to which degree the design objectives have been met, six students were individually interviewed regarding the infographic. This section describes the method used in terms of participants, procedure of data collection and analysis of the data. The Ethics Committee of the University of Otago issued ethical approval for this study on departmental level (category B).

5.4.1 Participants

Seven students, intending to take COMP160 in second semester, were approached from the cohort of 199 students taking the first semester COMP150 course. All seven students agreed to take part and six turned up for the interview. Each participant was paid $20 to compensate for time and travel. The students were chosen based on the following criteria:

- Their *approximate achievement level* to establish a mix of students achieving on high (A-grade), average (B/C-grade) and low level (failing) similar to the proportions of the entire population in 2011: 2 students to represent 33% A-grade students, 4 students of 50% B/C-grade students and 1 students of 17% failing students.
- Their *gender* to reflect roughly the proportions of the students in the course: 2 female students (15% female students) and 5 male students (85% male students) were chosen.
- Their *age* to reflect roughly the proportions of the course: 1 mature student was chosen.
- Their *personality*: all 7 students appeared to be communicative and honest during their lab sessions.

5.4.2 Procedure

Each interview followed the same structure and was broken up into five different parts. While exploring the different sections of the infographic, students were asked to “think aloud”. The “think aloud” technique is used in usability testing to understand “what is going on in a person’s head” while observing users in a controlled environment (Preece, Rogers and Sharp, 2011, p. 256). The accompanying questions for each part have been determined in advance. This semi-structured interview technique resulted in a comparable coverage of topics for each participant, but also allowed for flexibility when topics of interest arose. All interviews have been audiotaped.

The interview protocol was scheduled for the duration of 55 minutes and included the questions and aspects to be followed-up in the conversation. The general line of questioning addressed the design objectives for the different sections of the infographic and was guided by the following interview structure:
**Briefing (5 min) and Usage Scenario (5 min):**

Introducing the research topic and explaining the “think aloud” technique. The participant is asked to read the information sheet and to sign the consent sheet.

Describing the lab situation. Introducing the infographic as part of the lab book, which is received in the first COMP160 lab session.

1) **First Impression, Navigation and Structure (5 min):**

Presenting the infographic and encouraging the participant to “think aloud” in order to comment what he/she sees.

Aspects to follow up: (1) first impression as attractive/unattractive, (2) encouraging further investigation, (3) sections are recognised.

2) **Section “Key Events” and “What students say …” (10 min):**

Drawing the participant’s attention to the first & second section and encouraging to “think aloud” on what he/she sees.

Aspects to follow up: (1) time aspect, (2) parts of the final grade, (3) lab task aspect, (4) increase/decrease of task difficulty and (5) encouraging/discouraging impact of students quotes.

3) **Section “What we say …” (10 min):**

Drawing the participant’s attention to the third section and encouraging to “think aloud” on what he/she sees.

Aspects to follow up: (1) Advice understood, (2) Advice interesting, (3) Reaction to advice.

4) **Section “What the numbers say …” (10 min):**

Drawing the participant’s attention to the third section and encouraging to “think aloud” on what he/she sees.

Aspects to follow up: (1) seen as past years’ data, (2) different performance indicators recognised, (3) average final exam score per achievement group, (4) colour and size of bubbles interpreted correctly (5) mapping aspect understood. Furthermore participants were asked to “interpret” some bubble in terms of how many students passed/failed the final exam in that group.

6) **Debriefing (10 min):**

Questions about (1) what the participant expects the course to be, what kind of information was particularly (2) useful/interesting or (3) irrelevant/off-putting, (4) what other information to include. Participants were thanked and rewarded at the end of the interview.
The interview data were analysed by modifying the process of “Analytic Induction” (Flick, 2007, p.30) where pre-defined research hypotheses are adapted in a successive process to fit deviant cases. This way, research hypotheses are redefined and re-evaluated until they are fully explanatory for the cases investigated.

In the case of the infographic, the different design objectives were used for the evaluation instead of the suggested research hypotheses. During the analysis of the interview data deviant cases were revealed, which did not meet the design objectives. But rather than re-defining the design objectives, parts of the infographic were redesigned for a better fit regarding the design objectives. After the redesign a new series of interviews would have been necessary to re-evaluate whether the resulting version of the infographic meets the design objectives and does not cause other issues. This second round of interviews was not conducted. However a questionnaire and interviews were conducted as part of the scoping study to also address the degree of students’ understanding to reveal potential issues with the design. Figure 5.5 demonstrates the adapted process of the “Analytic Induction” as used for the analysis of the interview data.

![Figure 5.5: Analytic Induction (Gibbs, 2010) modified for infographic design evaluation](www.youtube.com/watch?v=SizaG3KKAp4)
5.5 Findings and infographic redesign

In this section the findings from the interview data are summarised for each statement tested with the modified version of “Analytic Induction” as described in the previous section. The following tables, which are introduced for a better overview, include (1) the design statement, (2) the related interview questions and (3) a short discussion of the findings for each statement and the deviant cases from the statement. At the end of each section the redesign is introduced addressing the deviant cases for the part of the infographic. For a complete printout of the revised, final version of the infographic please refer to Appendix D.

5.5.1 General impression, Navigation and Structure

<table>
<thead>
<tr>
<th>Statement 1:</th>
<th>Infographic is perceived as engaging - participants want to explore information in more detail.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Please describe what you see. What do you feel? Is the amount of information overwhelming, off-</td>
</tr>
<tr>
<td></td>
<td>putting or do you feel it is interesting and you want to explore in more detail?</td>
</tr>
<tr>
<td></td>
<td>There were no deviant cases. All participants were positive about the appearance of the</td>
</tr>
<tr>
<td></td>
<td>infographic and about exploring the information sheet in more detail.</td>
</tr>
<tr>
<td></td>
<td>“I like the design it is very colourful and … it looks very professional, looks fun, I am keen</td>
</tr>
<tr>
<td></td>
<td>to explore it basically.” (P6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statement 2:</th>
<th>The single sections are quickly recognised by using the side labels.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Please describe what you see. What do you think is this information</td>
</tr>
<tr>
<td></td>
<td>sheet about?</td>
</tr>
<tr>
<td></td>
<td>The “Key Events” section was easily recognised, also section “What</td>
</tr>
<tr>
<td></td>
<td>students say...”. Two participants were not instantly sure about</td>
</tr>
<tr>
<td></td>
<td>the section “What we say …” and how it is connected to the other</td>
</tr>
<tr>
<td></td>
<td>sections. The majority were attracted by the colours of the “What</td>
</tr>
<tr>
<td></td>
<td>the numbers say …” section, but were unsure about the content</td>
</tr>
<tr>
<td></td>
<td>because the years label (“2011/2012”) was not seen.</td>
</tr>
<tr>
<td></td>
<td>“It might be data from earlier years and some sort of marks or</td>
</tr>
<tr>
<td></td>
<td>other indicators of performance.” (P6)</td>
</tr>
<tr>
<td></td>
<td>About half of the participants used the left hand side labels to</td>
</tr>
<tr>
<td></td>
<td>find their way. I had the impression that the labels “What we say</td>
</tr>
<tr>
<td></td>
<td>…” and “What the numbers say …” are not descriptive enough for the</td>
</tr>
<tr>
<td></td>
<td>sections. One participant pointed out that the three dots need to be</td>
</tr>
<tr>
<td></td>
<td>used consistently.</td>
</tr>
</tbody>
</table>

There were no major issues reported, however the side labels were rearranged in order to be more descriptive for the sections (see Figure 5.7, Figure 5.8, Figure 5.9 below).
5.5.2 Section “Key Events”: Providing a quick course overview

**Statement 1:**
Key events are quickly recognised.

*Please describe what you see.*

There were no deviant cases. The Flag metaphor was well received. Comments have been positive about it.

“I really like it. It is very user-friendly … I really like the little flags actually. They are really good.” (P2)

One participant was wondering about the meaning of the colours (priorities?). Another suggested extending the lab task flag over the entire length of the timeline.

**Statement 2:**
The composition of the final grade is understood.

*What do you think of how your final mark is composed?*

The first flag “Lab Session out of 25” caused some confusion about how the 25% are earned, but participants guessed correctly. However, it is necessary to make clear that only submission counts (not attendance only) and that submitting tasks is not compulsory per lab session (different from COMP150).

For the first section “Key Events” no major changes were necessary. The explanation of the internal marks for the lab work was refined and a remark in section “What we say …” was added to explain that 50% are necessary to pass the final exam. A “timetable” was included to indicate the two weekly lectures (day and time given) and lab sessions (for students to be filled in). Based on a later remark of one participant who wanted to know more about the content of the lab sessions the background was enhanced to point to the programming concepts covered (Figure 5.6).

---

**Figure 5.6: Section “Key events” revised**

---

5 Communicating course characteristics and performance indicators · 66
5.5.3 Section “What students say …”: Sharing former students’ perception

Statement 1:
The change in students’ difficulty ratings (especially the initial increase of the perceived task difficulty) is understood and seen as a good preparation for the labs.

What do you think is this graph about? What do you think is this graph based on?

Because of the legend (easy/difficult) nobody had trouble to see what the graph was about.

“It is handy because you can tell like which labs are gone be almost - not easy - but which labs people generally find easy like the graphical stuff and then it suddenly gets straight back up where it was. You can prepare for things like that.” (P3)

However the connection to task numbers was, at least for one participant, not instantly clear and another one could not see the relation to the timeline. Most participants guessed correctly that the graph is based on students’ ratings.

Statement 2:
Students’ quotes are perceived as authentic, encouraging and displaying a wide range of emotions.

Please take some time and read through the quotes. How do you feel about it? Are these quotes encouraging/discouraging?

No deviant cases as students’ quotes were well received. Nobody was unsure about the link to the single lab tasks. The mixture was seen as informative and authentic:

“It was like a little bit worrying to see all of the sudden ‘I hate Java.’ but it also kind of means that it’s real as well, you are not glorifying the course by saying oh everyone said it was great. So you can trust that’s actual student feedback.” (P3) “[The jargon] puts a bit of comedy into it and makes you feel more involved.” (P2)

Comments were mainly perceived as encouraging.

“Find it like if I would do the course I would find it more encouraging because I would know what is going to happen and I would feel better prepared for the course.” (P5)

“I like knowing that other people had troubles as well and it wouldn’t just be me, because I often feel in the classes that it is just me having the problems.” (P2)

Only one participant would like to see more specific information about the labs.

To improve the connection with “Key Events” section the flag posts were connected to a mountain-like shape of the graph, using an up-hill (tasks are getting harder) vs. down-hill (tasks are getting easier) metaphor to illustrate changing difficulty (Figure 5.7).

Figure 5.7: Section “What students say …” revised
5.5.4 Section “What we say …”: Sharing teachers’ perspectives

**Statement 1:**
Single advice and recommendations are understood without difficulty. They are perceived as interesting and relevant.

*Please take some time to read through the advice. Please let me know if you (a) don’t understand what is said, (b) think it is not relevant or boring, (c) you think it is rather interesting or important. Do you think these advices are too teacher-like?*

There were no problems in understanding the section. Attributes used are “helpful”, “friendly”, “convenient” and “handy”.

“That is good again with the ‘mid-semester scores need to be taken seriously’, … and it also says what to do exactly ‘take action and talk to a demonstrator’ rather than just sit there and panic.” (P3)

One participant commented that the comments might not be seen when relevant, because this might be only looked at the beginning of the course.

“Type of thing I would only read once.” (P4)

Without asking it turned out that this section was perceived as not as attractive that the other sections.

“It is not something I would take the time to read.” (P2)

Some participants have overseen the markers (lines with red dots). There were suggestions of better connecting it or to make it stand out more.

“I think maybe connecting those [advice] to which lab it would be would be more useful because otherwise like as a student I probably wouldn’t bother reading that just because it’s a lot of text unless I can see some connecting thing here. Because with these ones there are all in nice colours [data] and you can just right look onto it and get the information.” (P2)

There have been comments regarding repetition about handing in the work, which is kind of “nagging”. On the other hand another participant said that this kind of repetition seems fine. None of the advice stood out to be uninteresting or useless. Most liked was the advice about the preparation for mid-semester which was once or twice followed by the wish to know about important labs.

To enhance this section visually, the background was coloured and connections to related information in other sections were emphasised. Key aspects were highlighted in bold type font. The selection of comments was reviewed to address the issue of repetition (Figure 5.8).
### Statement 1:
The representation of past years’ performance indicators is understood in terms of correlation with final exam score (average & actual achievement as box plot), group sizes and colour code.

*Please tell me what you think this graphic is about? (COMP150 indicator only)*

Four participants could not instantly see the correlations between different indicators and students’ final exam results. Group averages were seen as the averages for the indicator not for the COMP160 final exam result!

“I can’t see how this relates to COMP160” (P2)

Only two of the 6 participants could instantly read the graphics. Regarding the question, if the graphics appears to o difficult, one of the participants answered:

“No I think the way it is been displayed is probably the simplest way of doing it. I can’t think of a different way to show as much information.” (P4)

Four participants were familiar with the box plots, but perceived the graphic as too complicated.

“It is quite useful because you can work out your progress and sort of have an idea how you’ll go in the final exam, but it is quite over the top. I think these boxes here and your maximums and minimums, they are just a bit much. For a student looking on it they just go - too much I just want to see the glob and the average.” (P2)

“I would do a different type of graphics, kind of simpler. There is a little bit too much information.” (P6)

There has been no confusion about the colours. One participant commented that the colours are good because it is similar to a traffic light: red – stop and think, yellow orange – approach with caution, green – you are on track.

### Statement 2:
The numbers (labels) are understood and correctly interpreted.

*What do you think - is it easier or harder to get a good grade in COMP160 in comparison to COMP150?*

As mentioned above the labels required some explanation, but once the graphic was understood, everybody could answer the question correctly and interpreted the graph in the intended way.

“But I think the labelling of the groups is absolutely confusing and I wouldn’t be able to find it out myself.” (P6)

### Statement 3:
The meaning of the other indicators is correctly interpreted. The time aspect is recognised.

*What do you think is communicated in the next graphic? Do you think students in 2011 and 2012 had a chance of passing the final exam if their mid-semester score was under 50%? or Did all students being on time with their lab work pass the final exam?*

Basically no further explanation was necessary for the other indicators and each participant was at ease seeing the groups and related ranges of COMP160 final exam scores. Questions have been answered correctly. The order of performance indicators along time was seen as well.
Statement 4:
Participants know how they would map their own progress and why.

What do you think you are asked to do by “map your progress”?

The mapping aspect was obvious for the majority of participants. Participants seemed positive about the mapping aspect.

“I kind of got the message to keep on track and stop and think […] where would I be and - not place yourself - but think ok, maybe I should do this and keep up with it.” (P3)

“I would definitely, because I have done COMP150, I am not sure about other people, see myself and you know right away when you open the lab book and look at this you go ‘ok maybe I got 70% or something this is definitely a bit harder’ but it is not something straight away to worry about it and I can go and see where on the scale I am.” (P3)

“Keep track percentage wise so that you know where you sort of fit in from past years and where you are sitting for passing COMP160.” (P4)

“When I get my mark back … put where I was on the scale.” (P5)

“You should put yourself in one of those groups and see what your chances are.” (P6)

Without question this infographic section (“What the numbers say …”) required explanation for most participants and was perceived as too complicated by the majority. The redesign was centred on the correlation aspect (now an ordered arrangement), a substitute for the box plots to show the actual achievement for each group (now a donut chart), the average final achievement (now in the middle of the donut chart – can be used to “map” progress by shading) and the labels (now part of the indicator legend). Furthermore a bottom section was added, explaining the chart in general and including an example. Group sizes have been made explicit with an icon (person) and a number beside the icon (Figure 5.9).

Figure 5.9: Section “What the numbers say …” revised
### 5.5.6 Debriefing: General impression and improvements of the infographic

**What do you think is the course like (fun, easy, normal, difficult, frustrating)?**

Participants did recognise that students perceive the course quite differently. In general the course was seen as to be harder than COMP150. Also there were some comments that keeping up with the work is seen as the key to success and that teaching staff seems to be helpful.

"I think it will be a mixture of both (hard and easy). It does give you insight into the course." (P2)

"Yah, you just have to do your work. It is not going to be very easy, but if you work hard you will pass." (P6)

**What information of the infographic did you find particular useful or interesting?**

The 6 participants did mention the sections with the following frequency:

1. Section “Key Events” (5 participants)
   
   “I think that [events] is really useful. I like that and I like how you can instantly see ‘hey look I have an exam here’ you should start preparing and then there is your exam.” (P2)

2. Section “What the numbers say …” (4 participants)
   
   “… the correlation between actually doing the work and the end results” (P5)

3. Section “What students say …” (3 participants each)

4. Section “What we say …” (2 participants)

**What information was rather useless, not relevant, off-putting?**

The participants were in general hesitant with critique. It seems that the amount and content of information was perceived as being sufficient, helpful & encouraging.

"Wouldn’t take anything else out, maybe – no I wouldn’t - I make things a little bit more popping [in the middle part].” (P2)

**What other information do you image could be helpful?**

In the following there is a summary of what the single participants mentioned for this question:

P1: 50% pass requirement

P2: Extra tutors, Help desk (for one hour or once a week), Facebook page very useful, peer support from last years’ students available, more practical advice, structure of exams, times of labs and lectures, day of last lecture, suggests to shade in areas for your own progress. (Yes!)

P3: Nothing in particular, maybe one word about the concepts covered.

P4: Concepts are not necessarily needed because it is in the lab book anyway and as you should prepare you would know.

P5: FAQ general

P6: How others are performing in comparison, some statistics from the whole class.

The findings from the debriefing caused some minor changes which have been already discussed, e.g. to include a timetable (section “Key Events”), provide some space to map the actual level of achievement for the different indicators (shading the centre of donut chart in section “What the numbers say …”) and the revision of the advice selection (section “What we say …”). In general the debriefing was leaving me with the impression that the participants valued the information provided and could see how it would help them or to prepare and manage the course.
5.6 Summary

As self-regulation is seen as one important aspect in the learning process, we need to ask on what basis our students are expected to define their learning goals, self-monitor their performance and adjust their learning strategies if required. If courses appear as “black boxes” providing only little information about course demands, performance indicators and related passing rates, students will most likely act on the principle of hope rather than on an informed basis. The development of an infographic, as introduced in this chapter, is a step to open up the black box and offer more transparency, which allows students to judge the course requirements and their prospects in the course realistically as well as to adjust their learning efforts along the way.

For a scoping study, conducted in the second semester in 2013, an infographic was developed to inform students in an engaging way about the (1) course structure and key events, (2) former students perceptions about the course and the lab task difficulty, (3) important course characteristics and (4) performance indicators for different stages of the course based on correlations of naturally occurring course data with students’ final examination score. These design objectives resulted in 4 sections which were loosely connected by a timeline layout. An interview study, involving six students, was conducted to evaluate to what extent the design objectives were met and whether students felt attracted and inclined to attend to the information.

The interviews revealed that the students, although never exposed to this kind of information, valued the insights provided. The responses have been very positive in general, highlighting the perceived supportive character of the infographic. Not all aspects were instantly understood and the complexity of the statistical information needed additional explanation. Different design aspects, which were declared as deviant from the initial design objectives, were redesigned for the final version. This version of the infographic was not re-evaluated before it was rolled out for the scoping study and there was a potential danger that new issues contradicting the design objectives had been introduced. As part of the scoping study a questionnaire, infographic related tasks and individual interviews were conducted in order to investigate students’ acceptance, usage and perception regarding the information provided, as well as to re-evaluate the overall understanding of the information.

After the introduction of the methodology of the scoping study in the following Chapter 6, Chapter 7 presents the results of the study, including a discussion of students’ understanding informed by different units of analysis.
Planning and conducting a scoping study to support students’ self-regulated learning

“Real effectiveness can only be measured by looking at the impact. However, given the complex issues around feedback, is the impact of feedback measurable? If some form of measurement is possible how accurate can it be? How close can we get to evaluation? How hard should we try?”

(Price et al., 2010, p. 280)

This chapter is based on section 4 of a paper submitted to Computer Science Education that is currently under revision:

6.1 Introduction

The infographic as introduced in the previous chapter should provide students with answers to course related questions as well as help to establish performance goals in terms of submitted tasks at different stages or a mid-semester score to be achieved. Furthermore the infographic was designed as a tool for students to judge their own performance based on past years’ probabilities. After the development of the infographic, decisions about the form of delivery and the methods for evaluating the impact of the infographic on students’ perceptions about the course, their study behaviour and actual learning outcome need to be made. As any form of feedback can only be effective if students are engaged in the feedback process, this chapter addresses the following questions:

- What stages are important in the process of engagement with feedback?
- How to incorporate the stages of engagement in the planning and evaluation of a scoping study?
- What units of analysis are available to investigate the degree of engagement at the different stages?

A model established by Price et al. (2011) describing stages of students’ engagement with assessment feedback is introduced in the following Section 6.2 and was used to guide the conducting of the scoping study in semester 2, 2013 involving over 200 students. Section 6.3 informs about the format of the study and methods to be employed to evaluate the impact of the infographic at different stages of the course. The units of analysis are briefly introduced where relevant and described in detail in Section 6.4 regarding (1) the related research questions, (2) the method of data collection, (3) participants involved and (4) intended method of analysis.

6.2 Stages of students’ engagement in the feedback process

Attending to increase student satisfaction with assessment feedback Price et al. (2011) conducted a three-year study to investigate how to engage students more effectively in the feedback process. Interviews with students and staff were conducted to explore the stages of the engagement with assessment feedback.

(1) Collection is stated as the first “most visible indicator of students intention to engage” (Price et al., 2011, p. 888, original emphasis) where students’ motivation to collect feedback is influenced by the perceived usefulness of the feedback and the opportunity to apply it. The authors observed that, besides other reasons, students’ unsatisfactory prior experiences with feedback resulted in a diminished motivation to collect feedback. Based on a study of
360 medical students, Sinclair and Cleland (2007) reported that 46% of the students did not collect the formative feedback. Addressing the issue of students not collecting assessment feedback, Gibbs and Simpson (2004) state “Feedback is received and attended to” as one of their 10 conditions under which assessment supports students’ learning.

While students might still collect the feedback, paying (2) Attention can be discouraged if the feedback is not perceived as being usable (e.g. illegible or not to the expected detail). Other factors influencing the degree of attention include the ease of understanding as well as the credibility of and trust in the teacher providing the feedback. In contrast to Collection, the degree of Attention cannot be assessed easily. Teachers need strategies to follow-up students’ understanding and progress, e.g. by supplementing written feedback with face-to-face feedback sessions.

As the next, most critical stage, the authors state (3) Cognitive Engagement, focussing on the extent to which students can integrate the feedback in their learning, where expectations and beliefs about the purpose and nature of the feedback play an important role. Even though most staff reported to provide feedback beyond the sole justification of the mark, students rarely internalised and worked with the feedback. Price et al. (2011) observed that many students, while fully engaged in the early stages, found their expectations of specific, directive feedback not met. Extra help was needed to understand the feedback and inability to act on the feedback limited its usefulness. Students indicated that dialogue with staff would improve their degree of cognitive engagement and mentioned availability, attitude and approachability as supporting factors. The authors conclude that dialog and support are needed besides a shared understanding of feedback as a process and not as a product.

(4) Taking Action is in the core of effective feedback, but it is only one aspect of the engagement process. Price et al. (2011) warn: “Action resulting from feedback cannot be the ultimate measure of engagement with feedback, because a student may have been engaged at each stage of the feedback process but, in the end, still may not act on their feedback.” (p. 891). The authors emphasize the importance of considering the reasons for inaction. Motivation, opportunity and means are important aspects for substantive engagement. If, for example, resubmitted work can result in a higher mark, action is more likely to happen because the extrinsic motivation and the obvious opportunity for application. However taking action also depends on students’ self-efficacy. Often the gap in knowledge or skill is fully understood but further help is required to act on the feedback.

Each of these four stages needs to be considered as a potential obstacle in the engagement process, where “each stage can trigger further engagement or disengagement.” (Price et al., 2011, p. 888). The authors point out that disengagement is not a sudden reaction but de-
developed over time because previous experiences with feedback will influence the degree of later engagement (Figure 6.1).

![Figure 6.1: The temporal dimension of student engagement with assessment feedback after Price et al. (2011)](image)

These temporal dimensions of students' engagement with feedback informed and guided the decisions (1) regarding the delivery of the infographic as well as (2) what methods to employ for evaluating the impact. The rationale and research methods are described in the next section.

### 6.3 Research Methodology

Although Price et al. (2011) focus is on written assessment feedback, the model was found to be useful to untangle the different stages in any feedback process, where students are expected to collect, attend, respond and act on written feedback. Each of these stages can be seen as potential, consecutive hurdles for students’ engagement with feedback. Only if all hurdles are taken, can an impact on the learning outcome be expected. Therefore it is vital to evaluate the effectiveness of feedback at different stages and not only for the outcome to understand if and where the feedback process was potentially broken.

To do so we used and adapted the model as proposed by Price et al. (2011) for the planning and conducting of a scoping study, where students were provided with the infographic. An adaptation of the given model was necessary to evaluate not only the degree of immediate attention but also the expected frequent use of the infographic, especially when performance indications were available and students could potentially map their own performance in comparison with students’ results form the past two years. The model was extended by a stage called prolonged attention, which refers to the repeated use of the infographic after immediate attention. Figure 6.2 summarises the questions assisting the evaluation of the degree of engagement with the infographic at the different stages of the feedback process.
6 Planning and conducting a scoping study to support students’ self-regulated learning

Addressing problems of collection and immediate attention (Stage 1 & 2) the infographic was included as a size A3 foldout in the back of the COMP160 lab book (similar to Appendix D). Lab tasks 1 to 3 had a compulsory part of infographic related questions, which were printed on the back of the foldout (Appendix D). These tasks were checked for their correctness at the time of task submission. This way we could make sure that the feedback information was received and attended to.

The degree of understanding (cognitive response, Stage 3) was made visible through the correctness of the tasks. In case students provided missing or incorrect answers, tutors would address misunderstandings and explain related aspects of the infographic. In order to quantify the degree of understanding a sample of 34 infographic related task solutions were collected. A questionnaire (Appendix E) was conducted in week 12 and 13 (last two weeks of the course) to support this investigation. Five of the 19 items addressed the understanding of important aspects. An example is shown in Figure 6.3:

Furthermore six students, who have not been interviewed previously to evaluate the infographic, were interviewed at the end of the year to gain, besides other aspects, insight into the individual way students understood the infographic and reveal potentially problematic aspects.

Students’ actual course performance data were made available in the same way as in 2011 and 2012 (updated weekly and published outside the lab). After the first three lab tasks there was no obligation or encouragement for students to use the infographic. So it was interesting to see how often students used the information voluntarily at later stages of the course. To investigate this degree of self-reported prolonged attention (Stage 4) the questionnaire included four items about the frequency of usage for different sections of the infographic (Figure 6.4).
Questions of how the infographic was used after the first three lab tasks were also addressed during the interviews.

If students took action, triggered by the infographic, this can be made visible by tracking down changes in students’ actual study behaviour. Splitting the cohort of 2013 into an experimental and a control group was considered, but declined for ethical and practical reasons: Firstly, based on our genuine belief that the infographic can support students’ self-regulated learning which might in turn lead to better final results, it was considered as unethical to hold back this information for the control group. Secondly, because the infographic was part of the lab book which students need to bring to the lab sessions, we would have had no control if and to what degree the information was shared with students in the control group. For these two reasons we decided to assess changes in students’ perception, actual study behaviour and learning outcome by comparing data from 2013 with data from 2011 and 2012. The course content and management was identical for those years and we tried to minimize all factors influencing students’ perceptions and performance, which were not related to the infographic. However certain factors cannot be controlled. For example, in 2011 snow fall caused lower submission rates for those days, because students were unable to get to the lab.

The exploration of students’ immediate or prolonged reaction (Stage 5, Figure 6.2) involves a variety of units of analysis: Changes in students’ perceptions and expectations (e.g. final grade expected) are investigated on the basis of self-reported confidence data, resulting from an online survey conducted in 2011, 2012 and 2013. Course performance data collected in the same years of 2011, 2012 and 2013 are used to assess changes in students’ actual behaviour (e.g. number of submissions at different stages), which would indicate that students took action based on the information provided. If students found the information useful and influential and how much they valued the provision of diagnostic course data in general is discussed by observing questionnaire and interview responses.

Looking at students’ learning outcome is an important last stage to assess the impact of a feedback intervention. Given that students took action, changes in the learning outcome are likely to become visible. Mid-semester and final exam results are analysed to quantify changes in students’ learning outcome on a year-to-year basis. As changes in the exam re-
sults can be caused by differences in student cohorts, pre-course grades are investigated as one unit of analysis, which is independent from the infographic intervention.

The infographic was available for all 207 COMP160 students submitting at least one lab task. This number is a typical intake for COMP160 with mainly first-year university students. Of the 207 students, 137 had taken COMP150 at some stage between 2010 and 2013. Because we did not collect demographic data, actual numbers for male/female ratio as well the proportion of mature students are missing.

Ethical approvals were granted by the Ethics Committee of the University of Otago (Category A) and on departmental level (Category B) to cover the studies and related data collection methods introduced in this section:

1) Human Ethics Application: Category A (June 2012):
   - Investigation of students’ learning approaches and knowledge acquisition in COMP160 (General Programming)

2) Human Ethics Application: Category B (July 2012):
   - Study A: Investigation into reasons for quitting or failing COMP160
   - Study B: Investigation into experiences while teaching COMP160
   - Study C: Investigation into students’ confidence development in COMP160

3) Human Ethics Application: Category B (October 2013):
   - Study A: Investigation of students’ involvement and reactions to the information sheet provided in COMP160
   - Study B: Investigation into demonstrators’ experiences regarding students’ responses to course related information in COMP160
   - Study C: Investigation into students’ development and progress in COMP160 on an individual basis

The next section, 6.4, presents a detailed description of the available units of analysis and how the single data sets are used to evaluate the degree of students’ engagement with the infographic for the different stages of the engagement process.
6.4 Units of analysis (UA)

In this section the single units of analysis are introduced in more detail. The nature of the data sets is used to structure Section 6.4: Seven single data sets are described as naturally occurring course data in Section 6.4.1. Students self-reported confidence data resulting from a survey, which was conducted over the last three years, are introduced in Section 6.4.2 followed by the description of a sample of infographic related task solutions in Section 6.4.3. The last two sections (6.4.4 and 6.4.5) refer to data sets resulting from a questionnaire and a series of interviews conducted at the end of the course. For a summary of the units of analysis see Figure 6.5.

<table>
<thead>
<tr>
<th>Units of Analysis (UA)</th>
<th>Course Performance Data</th>
<th>Self-reported Confidence Data</th>
<th>Sample of IG Related Tasks</th>
<th>Questionnaire Responses</th>
<th>Interview Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Sets or Aspects covered</td>
<td>- Attendance - Submissions - Login times - Mid-semester and - Final exam scores - Pre course grades - Drop-outs</td>
<td>- Task difficulty - Progress - Confidence - Grade expectation - Frustration level - OO concepts - Profess. prospects - Progress in general</td>
<td>Task solutions for - Lab task 1 - Lab task 2 - Lab task 3</td>
<td>Infographic - Understanding - Frequency of usage - Perceived Impact Diagnostic data - Attitudes - Usefulness</td>
<td>Feedback on - Task level - Process level - Self-reg. level Infographic - Usage - Impact</td>
</tr>
</tbody>
</table>

Figure 6.5: Units of analyses (UA), related data sets and aspect covered

For the analysis of these data sets we use a small set of methods. Observation of numbers, averages and frequencies are employed to get a general impression of the nature of the data set or when research questions are not concerned with tests of statistical significance. Correlations are used to judge the existence and power of relations between two variables. Independent (or unpaired) significance tests are in the core of the analysis when differences on a year-to-year basis (between subjects) are assessed. Depending on the nature of the data set T-tests (interval) or Mann-Whitney tests (ordinal) are used. For differences within subjects dependent (paired) T-test (interval) are employed and on occasions a Wilcoxon Signed Rank test for ordinal data. Figure 6.6 provides an overview about the methods of analysis.

<table>
<thead>
<tr>
<th>Methods of Analysis</th>
<th>Numbers, Frequencies</th>
<th>Power (effect)</th>
<th>Differences (students)</th>
<th>Differences (years)</th>
<th>Differences (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>Observation</td>
<td>Correlation</td>
<td>Dependent T-Test (interval)</td>
<td>Independent T-Test (interval)</td>
<td>Mann-Whitney Test (ordinal)</td>
</tr>
</tbody>
</table>

Figure 6.6: Methods of analyses employed

In the following each unit of analysis is introduced by (1) stating the related research aspects with references to the engagement process, describing (2) the method of data collection and (3) the population or sample size as well as (4) referring to the intended method of analysing the data. At the end of each section limitations are discussed.
6.4.1 Students' course performance data (UA 1)

Students’ course performance data are collected throughout the year and form a set of naturally occurring data. Most of these data sets were already introduced in Chapter 4 and informed the investigation of the predictive value of course data to indicate students’ final performance at different stages of the course. For example, the number of submissions at week 5 and week 11 in 2011 and 2012 correlated moderately with students’ final examination score. Rather than focussing on correlations with students’ final examination results as presented in Chapter 4, these units of analysis are used to evaluate students’ engagement with the infographic and potential behavioural changes in the experimental group of 2013 when compared with the control groups in 2011 and 2012 to answer questions such as:

- Did students’ behaviour change in regard to lab attendance, time and number of submissions, time spend in labs or number and time of drop-outs?

Furthermore, students’ examination results will be used to assess changes in students’ learning outcome within the last three years. Because such changes can be influenced by differences between student cohorts, we incorporated pre-course results as an independent unit of analysis to assess potential differences between the students taking COMP160 in 2011/12 and 2013.

In preparation of this study we were cautious not to change any of the conditions for the 2011, 2012 or 2013 cohorts which might influence students performance namely (1) the composition of the final mark, (2) the regulations for attendance, terms requirements and lab usage and (3) the general structure and content of the lectures and lab tasks. UML diagrams supporting the description of the more complex lab tasks were introduced in the 2012 version of the lab book and some tasks required minor adjustments. In all three years it was assured that the same staff members presented the lectures and coordinated the lab work. Different groups of tutors helped the students in the labs, but the same two experienced staff members supported new tutors.

Differences between the experimental group in 2013 and the control groups in 2011 and 2012 are investigated by employing unpaired (or independent) T-tests to test the Null hypothesis that there are no differences within the data sets introduced in this section. On occasions paired T-tests are used to investigate within-subjects differences (e.g. comparing students mid-semester and final examination result). Employing T-tests is possible where the data sets are composed of interval scaled data. The tests are carried out on a confidence level of 95% (which means that there is a 5% chance that the differences in the means are due to chance) with p(two-tailed)<0.05 indicating significant differences. By using SPSS (version 22) the Levene’s Test for Equality of Variances (test of homogeneity of variance) is
reported as part of the *Independent Samples Test* statistic and is taken into account when reporting the results.

If not stated otherwise, the course performance data are drawn from the records of 166 students in 2011, 174 students in 2012 and 172 students in 2013 who completed the course by sitting the final exam. As drop-out time and rates vary between the years, corrections are necessary when comparing students’ course performance data on a year-to-year basis. Taking out the records of students dropping the course is necessary, because a comparison of, for example, attendance records or number of submissions is only valid for the time students have been active in the course. To avoid a distortion of the data sets by varying point in time and number of students dropping the course, these records are removed and considered separately.

Besides statistical significance, which is more likely to occur in large populations or samples (e.g. Cohen, 1992), the reports include *Cohen’s d* (d) as an indication of the effect size. There are different approaches to calculate d, depending on if (1) the differences between the means and standard deviation are used or (2) the *t*-test value and degrees of freedom. As the second approach leads to an overestimation of the effect size in the case of a paired (dependent) *T*-test (Field, 2013) we use the following approach incorporating the means (M) and the standard deviations (s) for all *T*-test statistics reporting d:

\[
\text{Cohen’s } d = \frac{M_1 - M_2}{s_{\text{pooled}}} \text{ where } s_{\text{pooled}} = \sqrt{\left(\frac{s_1^2 + s_2^2}{2}\right)}
\]

Cohen (1988) suggested some general definitions for interpreting effect size estimates: an effect size is small if d=0.20, medium if d=0.50 and large if d=0.80. However Hattie (2009) suggests, based on his experiences analyzing effect sizes in educational research, d=0.20 for small, d=0.40 for medium, and d=0.60 for large when judging educational outcomes. He continues to point out that those definition of effect size still might be in some situations “just too simple” and that small effects maybe important, especially if related interventions address vital issues or if those interventions can be implemented at low cost.

---

1 Cohen (1988) defined *d* as the difference between the means, divided by standard deviation of either group, which can be used when the variances of the two groups are homogeneous. Rosnow and Rosenthal (1996) suggest using the pooled standard deviation, *s*_{pooled}, instead.
**Attendance of lab sessions (UA 1/A)**

Attendance records are kept for the entire population of students enrolled and active in the course. These records are important, as students are required to attend a minimum of 21 lab sessions to be allowed to sit the final examination. Attendance is defined as either having completed the lab task before the end of the scheduled session (which allows students to work at home and only submit the task during their lab time) or being in the lab for the duration of the scheduled session. The attendance records are managed by the course coordinator on the basis of (1) students’ lab task submissions, which are recorded by lab tutors when students submit their work or (2) based on students’ login times if the task was not submitted. Attendance is usually granted if students were logged in for at least 90 minutes of their 120 minute lab session. We can assume high accuracy of the attendance records over the years as the same course coordinator was managing these records. The records are published, giving students the possibility to check the accuracy of their records. The attendance records for the last week are excluded from this investigation because there are data missing in all three years.

Unpaired T-tests are used to assess differences in how students attended the lab sessions (1) in general, (2) in the first 5 weeks and (3) in the 5 weeks after the mid-semester break.

**Number and time of submissions (UA 1/B)**

25% of the final mark is based on the number of submitted lab tasks. For that reason these submissions are an important part of the course for the students as well as for the course coordinator. Ideally a student would submit one finished lab task at the end of each lab session. To submit a finished lab task a student needs (1) to see a tutor to discuss the preparation task and lab task solution and (2) to electronically submit the folder with the files he/she was working on. There are two records for a submitted lab task: (1) the marking sheet provides a proof that a student has been signed off for a submitted lab task and (2) the drop box holds all files electronically submitted by the students with information about task number, date and time of submission.

As students sometimes forget to submit their lab work electronically or technical issues can interrupt the submission process, the submission numbers from the marking sheet, which can be considered as the official record, are significantly higher (M=20.79, SD=6.09) than the submission number resulting from the drop box (M=18.93, SD=5.86) as shown by a paired T-test incorporating all records from the past three years (t(587)=20.040, p<0.01). However for the investigation of changes in students’ submission behaviour, date and time of a submission are vital information. Therefore only electronically submitted lab tasks are considered if not stated otherwise.
When comparing students’ submission numbers on a year-to-year basis it is important to make sure that the existing error in the number of electronically submitted lab tasks is consistent for all three years of investigation. To do so, each student was assigned the difference between of his/her submission number resulting from the marking sheet and the number as recorded by the drop box. Three unpaired T-tests comparing this error value in submission numbers for the cohorts in 2011, 2012 and 2013 on a year-to-year basis resulted in no statistically significant differences, so we conclude that the error was on a comparable level in all three years and is unlikely to influence further investigation of this data set.

In order to address the problem that students not completing the course distort the investigation by very low submission numbers as well as that those students cannot be assigned to groups regarding their final achievement, only student records for students completing the course by sitting the final examination are compared. Student records from students dropping the course are investigated separately. The electronic submissions for lab 17 are removed from the data sets of all three years, because most students did not electronically submit this task, as it was not a programming task.

Unpaired T-test are the intended method to assess differences on a year-to-year basis regarding (1) the total number of submissions, (2) the total timeliness of submissions, (3) the number of submissions by week 5, (4) the number of submissions by week 11 and (5) the timeliness by week 8. These T-tests are repeated for groups of higher and lower achieving students based on their final examination results, as tests involving all students might obscure subtle changes for higher or lower achieving students only.

**Time spend in labs (UA 1/C)**

The actual time spent in the labs is another interesting set of data to investigate students’ working patterns. Login and logoff times for each student are recorded and are used to verify whether students were attending a lab stream. The data, drawn from these records, capture 24 hours login times from Monday to Friday. We need to be aware that students coming to the labs might do other computer work (especially in the evenings) or that they work at home and come only to the labs to submit their tasks. These scenarios lead to extreme long or short login times. Technical issues, such as missing records or students not correctly logged out, add to the distortion in the data. Therefore the login times might have limited validity or can be completely misleading in some cases.

To put the total time in relation to the achievement we calculated the average time a student spent on a task by dividing the total time by the number of lab tasks submitted by this student as recorded on the marking sheet. The following histogram (Figure 6.7) shows that a plausible range of 30 to 150 minutes per submitted lab tasks applies to 84.3% of the stu-
students, if we assume that the students spending only 30 minutes to 60 minutes per task most likely prepared for the lab work at home.

The average time spent in the labs per submitted task per student will be used to assess differences between the years using unpaired T-tests to investigate if students in 2013 worked differently than in 2011 and 2012. Correlations will be used to connect the time spent in the labs with students’ final examination score for the three years.

Scores for mid-semester (UA 1/D) and final examination (UA 1/E)

There are two examinations to assess students learning outcome under controlled conditions: (1) the mid-semester examination is held shortly before the mid-semester break and (2) the final examination at the end of the semester. Both examinations have the same format of a multiple choice question part (MCQ) and a short answer part (SA). The range of original scores varies per year depending on the content of the examinations and was converted to a value between 0% and 100%. Still it is very problematic to compare examination results, which are not based on the same set of questions, as those questions might vary in difficulty and causes for differences in the results are almost impossible to isolate.

To control for those differences an almost identical (minor changes may have been applied) final examination was conducted in 2011 and 2013, which is possible because the final examination is embargoed and not handed out to students. However, as a change between 2011/12 and 2013, more detailed feedback regarding the mid-semester examination was provided in 2013 with the possibility to review the examination instead of completing lab task number 17. If and how this additional feedback influenced students’ performance in the final exam is discussed in Chapter 7.

Population sizes for students with a mid-semester examination result are 176 in 2011, 195 in 2012 and 199 students in 2013 and are slightly larger as the populations of students
completing the course by sitting the final exam. Both data sets (mid-semester and final exam) are investigated by applying unpaired T-tests for an assessment of differences in student examination results. Correlations and paired T-test will be used to observe differences between students mid-semester and final examination result for the single cohorts of students.

**Scores for pre-course (COMP150) assessment parts (UA 1/F)**

Differences in student cohorts are investigated by comparing pre-course grades as one data set, which is not influenced by the intervention. Pre-course grades are available for a not compulsory, introductory Python course in first semester (COMP150), which is attended by 60% to 65% of the students taking COMP160. Not all students take COMP160 directly after taking COMP150. Sometimes there is a gap of several years with the possibility that students repeated COMP150. Therefore the matching of pre-course results needs to take into account all years after COMP150 was introduced, but only considering the latest result before the COMP160 enrolment.

COMP150 final marks are composed by two mastery tests, a mid-semester and a final examination. All scores are converted into a value between 0% and 100% regardless of their value towards the final grade. As passing COMP150 is not a precondition for COMP160, it is possible, for example, that some students would not have a final exam result, but would have performed the mastery tests. Consequently all students with a COMP150 final mark are included even if some assessment components were missing. This was applicable for 105 students out of the 2011 cohort of 166 students completing COMP160, 116 students out of 174 in 2012 and 123 students from 174 in the 2013 cohort.

The investigation of differences between student cohorts based on their pre-course grades is conducted by using unpaired T-tests to test the Null hypothesis of no differences for the COMP150 results between groups of students taking COMP160 in 2011, 2012 and 2013. Correlation between the different components of the COMP150 final mark and the COMP160 final mark are used to decide which of the components are most suitable to be included in the investigation.

**Number of students dropping the course (UA 1/G)**

The group of students dropping the course is of interest when investigating potential changes in students study behaviour. However students who dropped the course at some stage were excluded from the data sets discussed so far. Submission numbers and lab attendance are difficult to assess if we don’t know how many students are still active. Difficulties also arise when tests on achievement groups are intended which are based on students’ final exam results.
To investigate changes between the years in the group of students dropping the course, all students submitting at least one lab task electronically are included in this data set. The time of dropout is classified as before and after the mid-semester examination based on the availability of a mid-semester examination result. If a mid-semester result is available it will be distinguished between students scoring under 50% and over 50% in order to get an impression if the students dropping the course have been in danger of failing the final examination. This data set will be observed without applying any statistical tests. Averages of submitted lab tasks and mid-semester examination results (if applicable) will inform the discussion of these observations.

6.4.2 Students’ self-reported confidence development (UA 2)

Eliasson, Westin and Nordström (2006) investigated students’ confidence regarding different aspects of computer programming before and after an introductory programming course. The authors found a significant decline in students’ confidence and, when comparing students’ belief about their final grade with their actual grade, an even worse actual outcome. The authors report:

If we look at the total dataset, the students have significantly lower faith in their ability to program after the course than before. It is our belief that some of these students actually had too high confidence in their ability before the course and now have a more realistic view of what it is to program. (p. 26)

Motivated by this study, an online survey for COMP160 was initiated in 2011 in order to monitor students’ confidence development by prompting 4 statements (Figure 6.8) each time a lab task was electronically submitted and to be answered on a Likert-like scale:

![Online survey](image)

Figure 6.8: Online survey – four statements for each submission

The statements were chosen to monitor the confidence development based on different aspects: (1) perceived difficulty of task, (2) perceived learning progress, (3) perceived confidence and (4) confidence to replicate the task without help. Depending on the lab number students entered as part of the survey, a specific question would show up. This way, 5 diff-
Different statements were repeated every 5 lab tasks in order to pick up on trends and developments during the course (so called “weather station” questions). These reoccurring statements were concerned with the following aspects: (5) students’ expectation about their final grade, (6) their understanding of OO-programming, (7) their expectation of becoming a good Java programmer, (8) their level of frustration and (9) their level of satisfaction with the progress made (Figure 6.9).

![Online survey](image)

Figure 6.9: Online survey – one additional statement depending on the task number

With each submission (maximum of 25 per student) students entered a draw of three $20 movie vouchers. The survey in 2011 and 2012 provided us with valuable insight in students’ confidence development. Students’ difficulty ratings and optional comments informed the infographic section “What students say” (Appendix D). Most of the items correlated significantly on a weak to moderate level with students’ final exam result. Similar to Eliasson et al. (2006) we observed that students, and especially lower achieving students, lose confidence to achieve a good grade and that their grade expectations become more realistic towards the end of the course.

The survey was repeated in 2013 and the results will be used in this study to investigate changes in students’ confidence development. As the survey was unchanged and administered in the same way, the 2013 cohort forms the experimental group where as students from 2011 and 2012 will act as the control group to answer the following question:
Did students’ perceptions and expectations change in regards to task difficulty, confidence, progress made and final grade expected?

This question is interesting in the context of students’ immediate and latent reactions (Stage 5) to the infographic. For this investigation 125 items are compared consisting of five statements for each of the 25 lab tasks. Response numbers for the individual lab tasks differ widely depending on how many students submitted the task and responded to the survey. For further investigation all students not completing the course or providing less than three responses are removed from the data set. This leads to the following proportions of students participating in the confidence study: 91% (151 of 166) in 2011, 95% (165 of 174) in 2012 and 96% (166 of 172) in 2013. The following graph (Figure 6.10) summarises the response numbers per lab task (please ignore lab task 17 as the majority of students did not submit this lab electronically and therefore were not prompted for the survey) for all three years:

![Proportion of students responding to the survey in 2011, 2012, 2013](image)

**Figure 6.10: Proportions of students responding to the survey in 2011, 2012 and 2013**

In most cases the number of students responding to the survey varies not much between the years with an average difference of 6% between the highest and the lowest proportion for the three years. What is more obvious is the extreme drop in students’ responses towards the end of the course. Until lab task 15 the average proportion of responses for all three years was still over 80% and stayed above 75% of students responding up to task 21. With lab task 22 the response rate dropped under 70% and reached a minimum of 37% for lab task 25, which is caused by the fact that not all students manage to hand in these last tasks. However this drop is consistent for all three years and it should not influence the comparison of students’ responses between the years.

This observation underlines the main drawback in the survey design, namely that students not able to complete and submit certain lab tasks, cannot respond to the survey, even though they might have worked on the task and could rate the different aspects of their confidence. Especially for the “weather station questions”, which are not task related, we
miss out on valuable information. Towards the end of the year the responses are more and more dominated by higher achieving students, distorting the rating when compared for different tasks. However such comparisons are not part of the evaluation.

For the statistical analysis, all 125 items (5 per 25 lab tasks) will be compared separately in order to pick up differences in students responses on a fine-grained level. Working with averages over a number of items is not advisable because of the ordinal nature of the data (e.g. Field, 2013). Consequently a non-parametric test for independent samples needs to be employed. We use the Mann-Whitney U test, which is part of the SPSS (version 22) package. When comparing ratings within subjects a Wilcoxon Signed Rank test (also part of SPSS, version 22) is used. In order to support the discussion of statistically significant changes on a level of $p<0.05$ the effect size estimate $r$ is reported with the test statistic. Following Rosenthal (1991) $r$ is calculated as follows:

$$ r = \frac{z}{\sqrt{N}} \text{ where } N = N_1 + N_2 $$

Different from Cohen’s $d$ the effect size estimate $r$ needs to be interpreted as $\text{abs}(r) = 0.1$ is a small effect size, $\text{abs}(r) = 0.3$ indicates a medium effect and $\text{abs}(r) = 0.5$ is the threshold for a large effect (Field, 2013).

### 6.4.3 Students’ responses to the Infographic related tasks (UA 3)

As described already, students in 2013 needed to attempt a set of infographic related tasks (Figure 6.11), printed on the back of the infographic, for task completion of lab task 1, 2, and 3. A random sample of 34 infographic related task solutions was drawn towards the end of the course in week 12 in order to investigate the correctness of those answers. A high proportion of correct answers would indicate that students understood the related part in the infographic. In contrast large proportion of incorrect or missing answers would indicate that students either misunderstood or were unsure about the information or related questions. The discussion of students’ actual degree of understanding the infographic is part of the investigation of the initial cognitive response (Stage 1).

The data collection was carried out in week 12. As students needed to be signed off for their current task, the author would ask if he/she would be happy to share the written answers in the lab book for lab task 1 to 3 for a random, anonymous sample. No student declined, but one or two students could not provide the answers because they had a new lab book. The author copied all original answers, which were later compared with the correct answers. The rate of missing answers was low for lab task 1 and 2 and between 5% and 17% for lab task 3. For questions without a definite answer such as the goal related question (Figure 6.11, section 3, question 4) thresholds for correct answers were defined.
Section 1: Key Events & Advice [Questions for Lab 1, Part 3]
Section Key Events provides some information about what is happening in COMP160, parts that count towards your final mark and the concepts introduced each week. In section What we say you’ll find bits & pieces of advice.

1) How is your final mark composed? Please fill in the numbers:
   Lab Tasks ___ % + Mid-Semester Exam ___ % + Final Exam ___ % = 100% Final Mark

2) How do you get each 1% internal mark towards your final grade? You need to ...
   A) attend the scheduled lab session   C) submit a finished lab task on time
   B) submit a finished lab task at any time   D) submit a partly finished task

3) Why do students fail the course even though their final mark is 50 (C-) or more? Because they ...
   A) did not attend all lab sessions   C) scored under 50% in their final examination
   B) submitted less than 21 lab tasks   D) were late with half of the lab tasks

Section 2: Task Difficulty and Students’ Comments [Questions for Lab 2, Part 4]
In section What students say you’ll find a selection of students’ comments from 2011 & 2012 and a graph to visualise students’ ratings of the lab task difficulty.

1) If you consider the quotes, do you think students experienced the lab tasks in a similar way?
   A) yes   B) no   C) not sure

2) What did students rate to be the second hardest lab task? Lab number

3) What lab number is set for the first half of week 7? Lab number

4) Why do you think it is important to do this lab on time? Because it ...
   A) is more difficult than the previous lab   C) needs to be submitted for terms requirements
   B) is a catch-up session   D) it is a good preparation for the mid-semester examination

Section 3: Performance Indicators from 2011 and 2012 [Questions for Lab 3, Part 3]
By closely looking at data from 2011 & 2012 we realised that there are some indicators for how students perform in their final examination in COMP160. For example the COMP150 grade was moderately (in 2011) to strongly (in 2012) correlated with students exam results in COMP160. Another strong correlation was seen by the total number of lab tasks handed in. The strongest indicator, however, was the mid-semester examination result.

The charts in section What the numbers in 2011/12 said look quite complicated and there is a lot of information here. For five separate indicators we see:
- How many students were in each group for an indicator.
- The average score of the group's COMP160 final exam in the centre of the circle.
- The groups' actual distribution for the final exam results in terms of A-level (dark green), B-level (green), C-level (orange) and failing students (red).

Please have a look at the Mid-Semester exam result indicator.

1) For the group of students achieving 9 or 10 points out of 15 as their mid-semester exam result, what was their average final exam score? %

2) How many students were in that group? 

3) Can you estimate how many from this group could actually pass the final exam? Approximately %

Let’s say your goal is to complete COMP160 with a B-grade. Preferably you need to score above 65% in your final exam. (For an A-grade it would be over 80% and for a C-grade over 50%).

4) Based on the data from 2011 & 2012, what do you need to achieve to be confident to actually reach that goal?
   Number of labs by week 5  …  Mid-semeter result out of 15  
   Number of labs by week 11 …  Average lateness per submitted lab task by week 8 …
As the task solutions are anonymous it is not possible to assess to which degree the randomly sampled students reflect the entire population. Given that the sample was collected in week 12 it can be assumed, that students in the sample were most likely to sit the final exam to complete the course and that students dropping the course are not included. This limitation in mind we assume that the sample represents a cross-section of students completing the course, covering 20% (34 out of 172 students completing the course) of the population. For practical reasons the author herself collected the data, resulting in a risk of investigator bias for the data set, but care was taken to reduce this risk.

For the investigation into students’ degree of understanding the infographic, the data set is observed by comparing proportions of correct and incorrect answers for each of the 14 items: 3 items for lab task 1 (Figure 6.11, section 1), 4 items for lab task 2 (Figure 6.11, section 2) and 7 items for lab task 3 (Figure 6.11, section 3). Average values for all items will be reported to assess the degree of understanding of the related aspects of the infographic.

6.4.4 Students’ responses to a questionnaire (UA 4)

The investigation of (1) students’ understanding of the infographic (Stage 3), (2) the degree to which students referred back to the infographic (Stage 4) as well as (3) students’ degree of self-regulation and their attitudes towards diagnostic course data (Stage 5) was supported by a questionnaire administered in the last two weeks of the course. The questionnaire consisted of 19 items in five sections (Appendix E). The first four sections were addressing the four sections of the infographic (“Key events”, “What students say”, “What we say”, “What the numbers in 2011/12 said”). The last questionnaire section was concerned with students’ general attitudes towards diagnostic course data. 12 of the 19 items were statements assessed by a 5-point Likert-like scale with two anchor points of “Strongly disagree” (left) and “Strongly agree” (right). Another 3 items were related to the frequency of usage by ticking one out of four options regarding the statement “I looked up this section …”: “Often”, “Occasionally”, “Seldom” or “Never (after lab 3)”. Two items were concerned with the infographic’s influence on students’ confidence and self-regulation by ticking one aspect out of four. Ticking multiple aspects was allowed for the remaining two items rating (1) the usage of the performance indicators in the infographic section “What the numbers in 2011/12 said” and (2) the usefulness of the diagnostic course information in general. Please refer to Appendix E for a printout of the questionnaire.

The questionnaires with an attached “Information Sheet for Participants” and “Consent Form for Participants” were distributed in the labs in the last one and a half weeks of the course. A notice on the whiteboard informed about the questionnaire survey and the possibility to enter a draw of $100 by participating. Students were sometimes verbally encouraged to participate and a bag of candies was placed beside the box for the responses as a
symbolic gesture of our appreciation for responding to the survey. Offering rewards increases students’ participation but also increases the chance that students would answer quickly, only because of the reward. In fact one student submitted three responses to increase the chance of entering the draw.

In total 79 responses were collected and 75 were used for the investigation (the 2 duplicates were removed and two other responses were obviously filled in randomly) resulting in a response rate of 44% (75 of 172 students). 14 of the 19 items were answered by all 75 students responding to the questionnaire, 3 items by 74 students and one item by 73 students. Due to the fact that the survey was conducted at the end of the course, most likely the sample includes only responses from students completing the course. For 73 out of 75 responses we could connect the final examination results, which provided us with the possibility to test the Null hypothesis that the group of students responding to the survey performed equally in the final examination as the group not responding to the survey. The Null hypothesis was rejected ($t(167)=4.004$, $p<0.001$) by employing an unpaired T-test indicating that students responding to the survey ($M=71.77$, $SD=18.01$, $N=73$) had an higher examination score than students not responding ($M=59.51$, $SD=20.91$, $N=96$). The effect size of $d=0.62$ indicates a medium effect size. We conclude that the effect of self-selection resulted in a sample of higher achieving students.

The questionnaire responses will inform the discussion for different stages of the engagement process. Students’ answer choices are visualised in frequency charts without applying further statistical analysis. An attempt to quantify students’ self-reported usage of the infographic is made and connected with students’ final exam results to investigate correlations and differences for the prolonged attention of higher and lower achieving students by employing a Mann-Whitney U test for independent samples. The self-selection effect as addressed above need to be taken into account when discussing the results.

### 6.4.5 Students’ responses in an interview (UA 5)

Interviews at the end of the course (but before the final exam) were conducted to investigate students’ perception of feedback on different levels as suggested in Section 2.3. The main part of the semi-structured interview was concerned with (1) feedback regarding lab tasks, (2) the programming processes and (3) the course management, which included questions of how students understood, attended and reacted to the infographic. For a detailed interview plan refer to Appendix F.
The author conducted six interviews, each of which took between 35 and 40 minutes. Students were awarded with a $20 voucher to compensate for time and travel. The semi-structured interview technique resulted in a comparable coverage of topics for each participant, but also allowed for flexibility when topics of interest arose. All interviews were audiotaped.

Eight students were approached from the COMP160 cohort in 2013 and agreed to take part, but only six turned up for the interview. The students were chosen based on the following criteria:

- Their approximate achievement level to establish a mix of students achieving on high (A-grade), average (B/C-grade) and low level (failing) similar to the proportions of the population. Based on the mid-semester exam 2 students in the A-grade range, 4 students in the B/C-grade range and 2 students scoring under 50% were approached. Given their final mark the group of students, turning up to the interview, consisted of 2 A-grade students, 3 B/C-grade students and 1 failing student, which is an approximate mix of the entire population of students completing the course with 33% A-grade students, 50% B-grade students and 17% failing students in 2011.

- Their gender to reflect roughly the proportions of the students in the course: 2 female students and 6 male students were approached whereas 2 female students and 4 male students took part eventually.

- Their personality: all 8 students appeared as communicative and honest, based on the interaction in the labs.

Although an in-depth analysis of the interview data (e.g. by applying Analytic Induction as described in Section 5.4.3) was not carried out, the interviews form a valuable part of the data sets introduced here. Quotes are drawn from the interviews to illustrate the way the participants interacted with the infographic at different stages of the engagement process and how they felt about the diagnostic course data in general. Such anecdotal evidence provides a beneficial counterbalance to the quantitative tests of statistical significance and power.
6.5 Summary

After the development of the infographic, a research plan for a scoping study needed to be established describing the rollout of the infographic as well as the evaluation procedures. A model of “the temporal dimensions of student engagement with assessment feedback” as proposed by Price et al. (2011) was adapted to guide the scoping study. Stages in the engagement process informed decisions regarding the delivery of the infographic, the data sets to collect and the methods of analysis.

For ethical and practical reasons the infographic was rolled-out for the entire population of 207 students initially active in the course and was part of the lab book. Infographic related tasks were compulsory for the completion of the first three lab tasks. This way we addressed common problems (e.g. Sinclair and Cleland, 2007) of students not collecting or not paying attention to written feedback (Stage 1 & 2). Furthermore the correctness of the task solutions made students’ degree of understanding the infographic “visible” (Stage 3). A questionnaire was developed and interviews at the end of the year were held to inform the evaluation of students’ prolonged attention (Stage 4) as well as students’ immediate and latent reaction (Stage 5). Related to this last stage is the question of if and how students reacted to the infographic by changing their perceptions, expectations or behaviour, which can be answered by analysing the results of a confidence survey as well as course performance data. For this analysis student cohorts from 2011 and 2012 act as a control group and will be compared with the experimental group from 2013. The same approach is used to evaluate to which degree the intervention impacted students’ learning outcome based on students’ mid-semesters and final examination results from 2011, 2012 and 2013.

To which degree the self-reported data sets, namely the online survey (UA 2), the questionnaire (UA 4) and the interview responses (UA 5), are threatened by one or more “common method biases” (Podsakoff et al., 2003), effects causing a variance in the data set that is connected with the measurement method rather the subjects involved, we did not further investigate. Podsakoff et al. (2003) lists, amongst other potential sources of common method biases, social desirability which “is generally viewed as the tendency on the part of individuals to present themselves in a favorable light, regardless of their true feelings about an issue or topic” (p. 881) and acquiescence (yea-saying and nay-saying) connected with the tendency to agree (or disagree) with questionnaire items independent of their content. As we did not control for such effects, we need to be aware that conclusions are threatened by measurement errors caused by the self-reported nature of the data sets.
In Figure 6.12 the stages in the engagement process, as evaluated in the scoping study, are connected with the relevant units of analysis used to assess the degree of engagement at each stage. Related methods of analysis are indicated for each unit of analysis.

Figure 6.12: Stages in the engagement process and related units & methods of analysis

This framework will guide the discussion of the results in the next chapter by addressing each stage in the engagement process and the related questions of investigation.
Planning and conducting a scoping study to support students’ self-regulated learning
Chapter 7

Evaluating stages of engagement in the feedback process

“Previous studies of student programming ability have raised questions about students’ ability to problem solve, read and analyze code, and understand introductory computing concepts. However, it is unclear whether these results are the product of failures of student comprehension or our inability to accurately measure their performance.”

(Tew and Guzdial, 2010, p. 97)

This chapter is based on section 4 of a paper submitted to Computer Science Education that is currently under revision:

7.1 Introduction

In semester 2 in 2013 the infographic was rolled-out as part of the lab book for 207 students who were initially active in the course and submitted at least one lab task. One of the central questions in this chapter to ask is if students could improve their learning outcome (as assessed by the two examinations) based on the information they received about course characteristics and indications about their final performance. Such an assessment is not straightforward as differences in the examination scores of student cohorts are influenced by many factors. Conclusions regarding the effectiveness of a feedback intervention based on students’ “measurable” learning outcome only, can be misleading and are of limited explanatory power.

In order to get a more comprehensive view on the impact of the infographic, each stage in the engagement process, as introduced in the previous chapter, is assessed separately to answer the following question:

- To what degree did students engage with the feedback provided?

Different perspectives will enrich the discussion. For example for “Stage 5: Immediate and latent reaction” we will discuss not only changes in students’ study behaviour as “visible” actions, but also changes in students’ perceptions and expectations as well as their attitudes towards diagnostic course data in general as a reaction to the infographic. Such manifold assessment provides more depth and we will see that students’ indeed valued the information provided, but that “visible” actions were limited.

The following five sections (7.2 to 7.6) are structured according to the five stages in the engagement process. The questions stated for each stage are addressed by observing and analysing data from various units of analysis. A short summary is provided for each section. The last section 7.7 provides a discussion of the results by reflecting on the entire engagement process.
7.2 Stage 1 & 2: Collection and immediate attention

Collection and immediate attention are described as the first two stages in the engagement process with assessment feedback (Price et al., 2011). Only if students are able and willing to collect and attend to feedback provided, is further investigation into the degree of engagement or impact of the feedback sensible. For this reason the following question is of vital importance and will be approached by observing course performance data, namely the number of submissions for lab tasks 1, 2 and 3 (Figure 7.1).

- Did students collect and pay attention to the infographic?

Price et al. (2011) summarise the problems, which might lead to students’ lack of motivation to collect and attend to feedback provided. For example feedback was not seen as relevant by students either when provided too late to be acted on or if there was no opportunity to apply the feedback for further assignments. Furthermore earlier experience with feedback also plays a role in how students rate the importance of feedback. If feedback was consistently experienced as unhelpful, students’ engagement with feedback would decrease to the point that it is not even collected.

In the case of the infographic we provided a new, rather unfamiliar format of feedback. For one it is unlikely that students have been confronted with information about course characteristics in the past. Secondly the infographic, when provided at the first lab session, yielded very little information about students’ own performance (apart for the pre-course grade for 65% of the students) as the earliest performance indicator for all students is the number of submissions by week 5. In order to avoid that students ignored the infographic because they neither expected this kind of preparatory (“feed up”) information or they saw no personal relevance, we needed to find a way to engage students with the infographic right from the beginning.
The decision to make the infographic part of the lab book and to include compulsory tasks related to the infographic, ensured that all students collected and attended to the information provided. If students were marked for lab task 1, 2 or 3 we can assume that those students (1) received the lab book with the infographic foldout at the back (see Appendix D) and (2) attempted the infographic related tasks (see back of Appendix D or Figure 6.11). All of the 207 students active in the course (submitting at least one lab task) have been marked for lab task 1, 204 were marked for lab task 2 and 201 students got a mark for lab task 3. 206 students of the 207 active students (99.5%) were marked for at least two of the three tasks, which involved the attention to the infographic.

Given that demonstrators would only award lab marks for complete task solutions, including the infographic related questions for the first three lab tasks, we can conclude that 99.5% of all students in the course did collect the infographic and attended to the information provided at least twice. However there is a chance that students copied task solutions from other students’ lab books and would have been marked without attending to the information. In retrospect it is impossible to say for how many students that would have been the case. Anecdotally, cheating with task solutions is only a problem later in the course when students are running out of time to finish the lab tasks before the end of the semester or feel unable to manage the increasing task complexity, but even then it is not a common problem.
7.3 Stage 3: Initial cognitive response

According to Price et al. (2011) this stage is referring to the degree of students understanding and internalising the feedback provided. Students might pay sufficient attention and carefully read the information, but this does not necessarily imply correct interpretation. Therefore, related to Stage 3 of the engagement process, various units of analysis are employed (Figure 7.2) to answer the following question:

- Did students understand the information provided?

The parts of the lab tasks 1, 2 and 3 related to the infographic were not only designed to ascertain that students attend to all sections of the infographic, but also as a device to assess the degree of understanding of important aspects. A sample of 36 task solutions (UA 3) was collected to evaluate the correctness of student answers regarding the infographic related questions. The infographic related tasks included 14 questions in total (Figure 6.11), which were printed on the back of the infographic.

On average 82% of the students in the sample could answer the questions correctly, 12% answered incorrectly and 6% did not provide an answer. Incorrect answers might not only be explained by a lack of understanding the infographic but also by misunderstanding the questions. For example section 2, question 4 (“Why do you think it is important to do this lab on time?”) students misinterpreted the wording of “this” as the current lab number 2 and not as related to the previous question referring to lab number 12. One way or the other, demonstrators would point out and discuss incorrect or missing answers with the student during the marking process. This way it was ensured that misunderstandings or confusion regarding the infographic were addressed instantly for the 18% of students providing missing or incorrect answers.
Evaluating stages of engagement in the feedback process

Figure 7.3 summarises the proportions of correct, incorrect and missing answers for the 14 Infographic related questions (please refer to Figure 6.11 for the printout). The two questions regarding the composition of the final grade and the marking process (Section 1, Question 1 and 2) were answered correctly by 97% of students in the sample. The lowest proportion of correct answers with 69% was seen for section 3, question 3 where students were asked to consider the mid-semester performance indicator and estimate how many students passed the final exam from the group of students scoring 9 or 10 points. Admittedly this was a more complex question as students needed to fully understand the numbers section to (1) find the correct indicator, (2) find the correct group and (3) interpret the donut chart correctly. As mentioned before demonstrators were advised to discuss incorrect answers, so we believe that most of the students who could not answer correctly received an explanation on how to interpret the performance indicators.

The questionnaire (UA 4), administered at the end of the year, included five items to assess to what extent the design objectives as described in Chapter 5, Section 5.3 were met. Students were asked to rate how well aspects regarding (1) the course structure, (2) increasing task difficulty, (3) the bits and pieces of advice, (4) observed correlations with the final exam score and (5) the importance of consistent work were communicated. By shifting the focus of the statements away from students’ personal ability or inability to understand the information provided, towards an assessment of the quality of the design itself (e.g. “This section clearly showed …”) it was hoped to result in more honest and critical responses. Disagreeing with a statement meant not personal “failure” but rather flaws in the design to communicate certain aspects, but also indicates that students could not understand the information. In Table 7.1 the response for the five items are summarised.
Table 7.1: Questionnaire results (UA 4) – Questions evaluating the clarity of the infographic

<table>
<thead>
<tr>
<th>Aspect: Statement</th>
<th>Histogram</th>
<th>% pos.</th>
<th>% neu.</th>
<th>% neg.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Course Structure:</strong></td>
<td><img src="chart1.jpg" alt="Section 1 - Question 1" /></td>
<td>80.0</td>
<td>17.3</td>
<td>2.7</td>
</tr>
<tr>
<td>This section (“Key events”) made it easy to understand the structure of the course.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Increasing Difficulty:</strong></td>
<td><img src="chart2.jpg" alt="Section 2 - Question 1" /></td>
<td>78.7</td>
<td>18.7</td>
<td>2.7</td>
</tr>
<tr>
<td>This section (“What students say”) clearly showed that the tasks’ difficulty increases steadily over the first 5 weeks.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Advice:</strong></td>
<td><img src="chart3.jpg" alt="Section 3 - Question 1" /></td>
<td>73.3</td>
<td>25.3</td>
<td>1.3</td>
</tr>
<tr>
<td>The bits and pieces of advice in this section (“What we say”) were easy to understand.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Correlations:</strong></td>
<td><img src="chart4.jpg" alt="Section 4 - Question 1" /></td>
<td>71.6</td>
<td>16.2</td>
<td>12.2</td>
</tr>
<tr>
<td>Section (“What numbers …”) clearly showed how performance before and during the course correlates with the final exam result.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Consistent work:</strong></td>
<td><img src="chart5.jpg" alt="Section 4 - Question 2" /></td>
<td>86.7</td>
<td>9.3</td>
<td>4.0</td>
</tr>
<tr>
<td>It was clear from this section that working consistently pays off at the end.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Responses have been clustered into (1) positive (“strongly agree” and “agree”), (2) neutral (mid-point) and (3) negative (“disagree” and “strongly disagree”) and show that on average 78% of the students were confident that they understood the aspect in the infographic. 17% have been unsure about the information given and 5% of students disagreed that the aspects were clearly communicated. Despite the fact that the highest proportion of students disagreed (12%) about the clarity of the correlation aspect (section 4, question 1), 87% of the students reported being clear about the aspect that working consistently is important for success in this course (section 4, question 2). Again it is not entirely clear how the wording of the statements influenced students’ responses. Most likely the short statement of “It was clear from this section that working consistently pays off at the end” was easier to understand than the previous question assessing a similar aspect but worded differently: “Section (“What numbers …”) clearly showed how performance before and during the course correlates with the final exam result”. The fact that a quarter of the students responded neutrally to the statement that the advice section was “easy to understand” (section 3, question 1) is surprising as we did not expect this section to cause any issues.

Addressing if students’ responding to the statements on a positive scale or on a neutral/negative scale achieved differently in the final exam, an unpaired T-test was conducted. All five items as listed in Table 7.1 were tested separately and no significant differences were found. This was reiterated by a follow-up test, using the sum of all rating per student for the five items and comparing those summed-up ratings for the group lower achieving students with the group of higher achieving, which resulted in no significant difference.

Regardless of the small variations between the single items we can conclude that the majority of students (around 80%), responding to the questionnaire, felt confident that the infographic clearly communicated the aspects which were intended. This was reiterated by the six interviews at the end of the course (UA 5). Section 3c of the interview protocol (Appendix F) was set up to assess how the participants felt about understanding certain aspects of the infographic. Only one participant reported that she was initially confused about the infographic section “What numbers in 2011/12 said”. This confusion was resolved after she had completed the related tasks (Figure 6.11, section 3, questions 1 to 4). More objectively, the sample of infographic related task solutions (discussed at the beginning of this section) proved that a similar proportion of students (around 80%) could answer the related questions correctly and understood the information provided.
7.4 Stage 4: Prolonged attention

In contrast to assessment feedback as described by Price et al. (2011) the infographic was designed to be attended to at different stages of the course. We hoped that students would refer back to the infographic whenever performance data were available, but there was no obligation to do so. The performance indicators presented in section “What the numbers say …” were laid out in a way that students had the possibility to map their individual performance in relation to an expected final examination result, which was based on the data from 2011 and 2012. Referring back to the infographic also indicates that students found the information useful enough to voluntarily re-engage with the feed-forward (e.g. upcoming task difficulty) or feedback (e.g. individual mid-semester score as performance indicator) information. The investigation described in this section, concerned with the degree of prolonged attention, is based on students’ questionnaire and interview responses (Figure 7.4) and addresses two questions:

- Did students refer back to the infographic during the semester?
- Did students map their own performance?

The questionnaire (UA 4) included four items to investigate the question of how often students referred back to different sections of the infographic and to quantify their engagement with the performance indicators. Students’ responses regarding prolonged attention are summarised in the following Table 7.2:
Table 7.2: Questionnaire results (UA 4) – Questions evaluating frequency of use of the infographic

<table>
<thead>
<tr>
<th>Infographic Section:</th>
<th>Statement</th>
<th>Histogram</th>
<th>% high</th>
<th>% low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section “Key Events”:</td>
<td>I looked up this section ...</td>
<td><img src="image" alt="Section 1 - Question 2" /></td>
<td>43.2</td>
<td>56.8</td>
</tr>
<tr>
<td></td>
<td>N=74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section “What students say”:</td>
<td>I looked up this section ...</td>
<td><img src="image" alt="Section 2 - Question 4" /></td>
<td>67.6</td>
<td>32.4</td>
</tr>
<tr>
<td></td>
<td>N=74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section “What we say”:</td>
<td>I looked up this section ...</td>
<td><img src="image" alt="Section 3 - Question 3" /></td>
<td>36.0</td>
<td>64.0</td>
</tr>
<tr>
<td></td>
<td>N=75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section “What numbers in 2011/12 said”:</td>
<td>I looked at the “donut charts” for my own results: (please tick all that apply)</td>
<td><img src="image" alt="Section 4 - Question 4" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N=75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Column “% high” represents the percentage of students reporting high usage by answering that they used the section “often” or “occasionally”. Column “% low” clusters the responses of “seldom” or “never”. The lowest latent attention was reported for section “What we say” (advice) with 36% of students using this section “often” or “occasionally”. Section “What students say” (task difficulty and students’ comments) was used most often and 67.6% of students reported that they used this section “often” or “occasionally” and 12.2% of students responded that they never looked at this section after lab task 3.
Of particular interest is section “What the numbers in 2011/12 said” (performance indicators) because this section was designed to give students an indication about their own performance in relation to an expected final exam result. We assume that students using this section have been concerned about their prospects in the course. 20% of the students responding to the questionnaire reported that they never used this section for their own results. The 80% of students using this section did so for approximately three indicators on average (2.8 of 5 indicators possible). Almost everybody (54 out of 60 students) used the mid-semester chart for an indication about his or her projected final performance.

When combining the responses for the 4 questionnaire items for each student, as presented in Table 7.2, we found that only 4 out of 75 students responding to the questionnaire, reported that they never used the infographic after lab task 3. In other words, 95% (71 of 75 students) used the infographic voluntarily at some stage. We excluded the usage of indicator “COMP150 final mark”, because it is not necessarily indicating a prolonged attention as students could have used this indicator when initially attending to the infographic.

Furthermore the interviews at the end of the semester (UA 5) consisted of a part to investigate the prolonged attention as well. The six participants were asked about their use of the infographic section “What the numbers in 2011/12 said” (see Appendix F, section 3c, question 5). Their responses revealed that two of the three lower achieving students of the six participants (final grades: F, B-, B-) never used the section of the performance indicators. The other student used the section to look up his COMP150 result only. The three higher achieving students (final grades: B, A+, A+) reported to have been using this section on a regular basis to look up their own results (Table 7.3).

Table 7.3: Interview results (UA 5) - Students reporting on how they used the “Numbers” section

<table>
<thead>
<tr>
<th>Participant</th>
<th>Mid-sem. score %</th>
<th>Final exam score %</th>
<th>Final Grade</th>
<th>Usage of Performance Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>42</td>
<td>41</td>
<td>F</td>
<td>Participant never looked at the section after lab 3.</td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td>58</td>
<td>B-</td>
<td>Participant only used the COMP150 indicator.</td>
</tr>
<tr>
<td>1</td>
<td>51</td>
<td>61</td>
<td>B-</td>
<td>Participant never used the infographic after lab 3.</td>
</tr>
<tr>
<td>6</td>
<td>63</td>
<td>67</td>
<td>B</td>
<td>Participant used the section.</td>
</tr>
<tr>
<td>4</td>
<td>80</td>
<td>89</td>
<td>A+</td>
<td>Participant used the section.</td>
</tr>
<tr>
<td>5</td>
<td>84</td>
<td>88</td>
<td>A+</td>
<td>Participant used the section.</td>
</tr>
</tbody>
</table>

This observation suggests a possible correlation between the level of achievement and the usage of the infographic, probably revealing that higher achieving students felt more inclined to use the performance indicators than lower achieving students. However an analysis of the Pearson correlation coefficient between students’ reported number of indicators used throughout the course (based on responses as shown in Table 7.2, last row, but without the indicator “COMP150 final mark” because it was not applicable for all students) and students’ final exam score resulted in a significant weak negative correlation ($r=-0.26$, $p < 0.05$).
p<0.05, N=73) indicating that the observation from the interviews contradict with the actual trend.

In order to follow-up on the observation that lower achieving students reported a higher usage of the infographic, the group of 73 students was split into two groups based on their performance in the final exam. Given the ordinal nature of the data ("often", “occasionally”, “seldom”, “never”) a Mann-Whitney U test was employed to test the Null Hypothesis that there are no significant differences in how students responded to the four statements (Table 7.2). 11 items were tested regarding students’ prolonged attention:

- 3 items with one answer choice regarding the frequency of usage (see first 3 rows in Table 7.2)
- 6 items for the usage of the performance indicators (see last row in Table 7.2)
- 2 aggregated items consisting of (1) the number of indicators ticked (COMP150 not included) and (2) the sum of the rating regarding the frequency of the usage for infographic section 1, 2, and 3.

The number of the indicators ticked, turned out to be statistically significant. The group of lower achieving students (N=37, mean rank=42.14) reported on a higher use of the performance indicator than the group of higher achieving students (N=36, mean rank=31.72) with U=476.00, z=-2.147, p=0.032, r=-0.251. An unpaired T-test was also applied, as the number of indicators ticked is interval scaled data, with very similar results also indicating a statistical significant difference with a medium effect size of d=-0.506.

Based on the questionnaire responses we can conclude that the vast majority of students (95%) indeed referred back to the infographic at least once. The performance indicators were used by 80% of students. During the interviews (UA 5) we observed that the three higher achieving students reported on more frequent use of the performance indicators than the three lower achieving students. In contrast the statistical tests based on the questionnaire responses (UA 4) showed that the group of lower achieving students reported statistically significant greater prolonged attention for the performance indicators than the group of higher achieving students with a medium effect size of r=-0.251. Given that there were no statistically significant differences for the use of the other infographic sections we cannot conclude that lower achieving students have been more engaged with the infographic in general, but certainly with the information about their projected performance as provided by the performance indicators.
7.5 Stage 5: Immediate or latent reaction

Figure 7.5: Overview of evaluation of Stage 5

The evaluation of the previous three stages indicated a high degree of engagement. According to Price et al. (2011) it is not unlikely that students, despite their engagement in the feedback process, do not act on the feedback due to a lack of motivation or because further assistance would be necessary. This section is set out not only to investigate if students acted on the infographic by adapting their actual study behaviour but also to track down changes in students’ perceptions and expectations. As we did not, for ethical and practical reasons, employ a control group, we can only compare year-to-year. The questions to be addressed in this section are focussed on three different aspects and incorporate four of the five units of analysis available (Figure 7.5). Firstly we expected to see some change in 2013 in how students felt about the tasks and their prospects in the course at different stages, which was monitored by an online confidence survey in 2011, 2012 and 2013 (UA 2):

- Did students’ perceptions and expectations change in regard to task difficulty, confidence, progress made and final grade expected?

Secondly we analyse naturally occurring course data (UA 1) such as attendance records and submission numbers, which were gathered over the last three years to answer the questions about students’ subtle behavioural changes in 2013:

- Did students’ behaviour change in regards to lab attendance, time and number of submission, time spend in labs, number and time of drop-outs?

Finally the interviews (UA 5) and the questionnaire (UA 4) administered at the end of the course included some questions to assess how students felt about the diagnostic information provided. The two questions to be addressed here are:
- Did students find the information provided useful and influential?
- Did students’ value the provision of diagnostic course data in general?

In the following three sections (7.5.1 to 7.5.3) we present the results of the analysis regarding the three topics and related data sets as outlined above. Each section closes with a short summary. Section 7.5.4 provides a discussion incorporating all aspects used to explore students’ reaction to the infographic.

### 7.5.1 Changes in students’ perceptions and expectations: Self-reported confidence data

The confidence survey (UA 2) was composed of four aspects which were repeated for each submitted lab task: (1) perceived task difficulty, (2) perceived value for learning the Java programming language, (3) confidence while solving the task, (4) confidence to repeat the task without help. In addition to the task specific questions the survey included one specific question per task, which was concerned with the following aspects at different stages of the course: (5) students expectation about their final grade, (6) their understanding of OO-programming, (7) their expectation of becoming a good Java programmer, (8) their level of frustration and (9) their level of satisfaction with the progress made. These “weather station questions” recurred in the survey at every fifth lab tasks to track changes in students’ perceptions over time.

Considering the information provided by the infographic, we expected to see changes in how students reported on these items. For example, mastering a task, which was indicated as being difficult (infographic section “What students say”) might boost students’ confidence and therefore influence the confidence ratings (aspect (3) and (4)) for this task. Changes are also possible for students’ ratings regarding their expectations about their final grade (aspect (5)), as it is possible that students, when attending to the performance indicators, would more accurately judge their expected final outcome based on their actual performance. In this section the question to what degree students responded differently to the survey in 2013 than in the preceding two years is explored.

As a first step, before comparing the results of the online survey from 2011 and 2012 with the results in 2013, we conducted a comparison between 2011 and 2012 to get a feel for the robustness of the survey responses on a year-to-year basis for the different cohorts of students before the use of the infographic. A Mann-Whitney U test for independent samples was carried out for all 125 items (5 items for each of the 25 lab tasks) to compare the responses in 2011 with the responses in 2012. A rejection of the null hypotheses was indicated for three items out of 125 analysed:
1) Lab 1, aspect 5 (grade expectation): Group 2011 (N=142, mean rank=137.16) rated their expectations for a good grade significantly lower than group 2012 (N=150, mean rank=155.34) U=11,976.00, z=1.961, p=0.05, r=0.12.

2) Lab 5, aspect 4 (confidence to repeat task without help): Group 2011 (N=137, mean rank=129.14) rated their confidence to repeat the task significantly lower than group 2012 (N=142, mean rank=150.48) U=11,215.50, z=2.333, p=0.02, r=0.14.

3) Lab 7, aspect 3 (learning progress): Group 2011 (N=136, mean rank=129.81) rated their learning progress for this task significantly lower than group 2012 (N=141, mean rank=147.86) U=10,837.50, z=1.963, p=0.05, r=0.12.

Two of these items turned out to be marginally statistically significant with a p-value of 0.05. The effect sizes (r) were small for all three items with r<0.2. These test result led to the conclusion that students responded in a very similar way in 2011 and 2012, which (1) indicated robust survey responses between the years and (2) allowed us to combine the data sets from 2011 and 2012 for further analysis.

As a second step the general soundness of the survey is demonstrated by describing major effects as observed from the data in 2011 and 2012. For testing the Null Hypothesis between groups of students (e.g. lower achieving vs. higher achieving students) a Mann-Whitney U test for independent samples was used. Wilcoxon Signed Rank test was employed when testing for differences between ratings for the same group of students (e.g. comparing grade expectations as indicated when submitting lab task 1 with lab task 16). In the following these effects are summarised for the nine aspects covered by the survey:

1) Students’ perception of the task difficulty (aspect 1) varies widely within subjects and between the tasks. For example students rated lab task 3 as less difficult than lab task 9 (W(247)=1,719.50, z=-10.494, p<0.01) with a large effect size of r=-0.47. Furthermore we observed significant differences between the group of lower and the group of higher achieving students with a tendency of lower achieving students to rate tasks as more difficult. This effect was consistent for all lab tasks and turned out to be statistically significant when employing a Mann-Whitney U test with p<0.05 for 18 of the 25 lab tasks.

2) Students’ ratings of the task difficulty (aspect 1) are negatively correlated with their ratings of the learning progress (aspect 2). In other words students rate easier tasks as less valuable for the learning progress. Significant, but weak correlations could be observed especially at the beginning of the course for lab task 1 to 4 (e.g. lab task 3: r=-0.246, p<0.01, N=293).
3) Students’ confidence ratings (aspect 3) are strongly, positively correlated with the rating for how easy the task was perceived (aspect 1), indicating that students feel less confident when working on harder tasks. This correlation was statistically significant (p<0.01) for all 25 lab tasks with an average value of r=0.713.

4) The rating of the confidence to repeat a similar task (aspect 4) is strongly, positively correlated with the rating of the perceived confidence while working at the task. This correlation was statistically significant (p<0.01) for all 25 lab tasks with an average value of r=0.753.

5) Students’ grade expectations (aspect 5) decrease overtime, developing from a positive very weak correlation (1st rating: r=0.187, p<0.01, N=292) into a moderate correlation, with students’ actual final mark (4th rating: r=0.476, p<0.01, N=253 and 5th rating: r=0.428, p<0.01, N=236). Interestingly, when employing a Wilcoxon Signed Rank test to compare the 1st rating with the 4th rating within subjects, the decrease is only significant for the group of lower achieving students (W(99)=219.00, z=-4.759, p<0.01) with a medium effect size of r=-0.34.

6) Only higher achieving students gain confidence to explain what object oriented program design means (aspect 6). A within subject comparison showed significant differences between the 1st, the 3rd, and 5th rating for the higher achieving students (e.g. 1st to 5th rating: W(127)=3,004.00, z=5.279, p<0.01, r=0.33), but no statistical significant differences for the group of lower achieving students.

7) No changes were observed within subjects as to how students rated their expectation of becoming a good JAVA programmer (aspect 7). However when testing higher and lower achieving students separately we observed a statistically significant increase in confidence for higher achieving students (1st to 5th rating: W(123)=1,261.00, z=2.701, p<0.01, r=0.17) and a statistically significant decrease for lower achieving students (1st to 4th rating: W(99)=330,50, z=-2.232, p=0.026, r=0.16) with small effect sizes.

8) In comparison with higher achieving students the groups of lower achieving students reported to be usually more frustrated when working at the tasks (aspect 8). This difference was significant for all 5 ratings. For example for the 2nd rating, lower achieving students (N=115, mean rank=104.07) felt more frustrated than higher achieving students (N=143, mean rank=149.95) U=11,968.50, z=-5.017, p<0.01 with a medium effect size of r=-0.31. There are significant, positive correlations between how relaxed students feel and their final exam result for all 5 ratings (e.g. 2nd rating: r=0.354, p<0.01, N=258).
9) Higher achieving students are satisfied on a higher level with their learning progress (aspect 9) than lower achieving students. This was statistically significant for the 2nd to 5th rating. For example a medium to large effect size of r=-0.41 for the 3rd rating was calculated when comparing the group of higher achieving students (N=141, mean rank=157.75) with lower achieving students (N=119, mean rank=98.21) U=4,547.50, z=-6.635, p<0.01. There are significant, positive moderate correlations for all 5 ratings between how satisfied students are with their progress and their final exam result (e.g. 3rd rating: r=0.435, p<0.01, N=260).

Most of these observations are not surprising and support common sense assumptions, e.g. that higher achieving students are more confident about their prospects in the course (observation 5). Other examples include that lower achieving students rate the lab tasks as more difficult than their higher achieving classmates (observation 1) and lose confidence about becoming a good programmer where higher achieving student gain confidence (observation 7). Based on these observations we can conclude that students responded to the confidence survey in a rather systematic and predictable way.

In a last step we used the Mann-Whitney U test for independent samples to compare students’ responses from 2011 and 2012 with the responses from 2013 to explore changes in students’ perceptions possibly caused by the infographic provided in 2013. The rejection of the Null hypothesis was indicated for 10 items:

1) Lab 4, aspect 1 (task difficulty): Group 2011/12 (N=295, mean rank=229.14) rated this task significantly easier than group 2013 (N=146, mean rank=204.55) U=19,133.50, z=-1.970, p=0.049, r=0.09.

2) Lab 4, aspect 2 (learning progress): Group 2011/12 (N=295, mean rank=211.19) rated their learning progress for this task significantly lower than group 2013 (N=146, mean rank=240.81) U=24,427.50, z=2.380, p=0.017, r=0.11.

3) Lab 4, aspect 4 (confidence to repeat task without help): Group 2011/12 (N=293, mean rank=220.87) rated their confidence to repeat the task significantly higher than group 2013 (N=146, mean rank=196.57) U=17,968.00, z=-1.998, p=0.046, r=-0.10.
6) Lab 8, aspect 1 (task difficulty): Group 2011/12 (N=267, mean rank=212.29) rated this task significantly easier than group 2013 (N=140, mean rank=188.19) U=16,476.50, z=-2.007, p=0.045, r=-0.10.

7) Lab 9, aspect 1 (task difficulty): Group 2011/12 (N=260, mean rank=214.43) rated this task significantly easier than group 2013 (N=143, mean rank=179.39) U=15,357.00, z=-2.952, p=0.003, r=-0.15.

8) Lab 9, aspect 4 (confidence to repeat task without help): Group 2011/12 (N=260, mean rank=210.97) rated their confidence to repeat the task significantly higher than group 2013 (N=143, mean rank=185.70) U=16,259.00, z=-2.133, p=0.033, r=-0.11.

9) Lab 12, aspect 2 (learning progress): Group 2011/12 (N=268, mean rank=195.92) rated their learning progress for this task significantly lower than group 2013 (N=142, mean rank=223.57) U=21,594.50, z=2.366, p=0.018, r=0.12.

10) Lab 13, aspect 9 (satisfaction with learning progress): Group 2011/12 (N=253, mean rank=197.78) rated their satisfaction with the learning progress in the course significantly higher than group 2013 (N=125, mean rank=172.73) U=13,716.50, z=-2.182, p=0.029, r=-0.11.

Differences between students’ ratings in 2011/12 and 2013 were indicated for more than one item for lab tasks 4, 7 and 9. The perception of the task difficulty was affected for all three tasks in the way that students in 2013 rated the tasks as being more difficult, which was accompanied by lower ratings for the confidence while working on the task (lab task 7) or to repeat the task without help (lab task 4 & 9). Given that these lab tasks were not significantly changed between 2011 and 2013 we might interpret these changes as an effect of students taking into account the information about the increasing task difficulty at the beginning of the course. Especially lab task 7, 8 and 9 in the infographic section “What students say” were labelled with student comments underlining that these tasks were particular hard (Figure 7.6). In fact all of these tasks were rated in 2013 as being more difficult than in 2011/12 at a significance level of p<0.05.

Figure 7.6: Infographic detail indicating task difficulty
However, effect sizes for all 10 items are small with $r<0.2$ and indicate that the rejection of the null hypotheses was likely caused by large population sizes than by substantial effects. On one hand, applying a 95% confidence interval for statistical significance leads theoretically to 5% false rejections. On the other hand, taking into account that the Null hypothesis was rejected for more than one item for lab tasks 4, 7 and 9 adds to the validity of the rejections for those tasks.

**Summary and Discussion**

Data resulting from an online confidence survey (UA 2) were analysed to (1) test the robustness of the survey year-to-year by comparing the responses from 2011 and 2012, to (2) demonstrate the validity of the survey by describing major effects observed in student responses in 2011 and 2012 and to (3) explore the responses of the 2011/2012 cohort in comparison with student responses in 2013 to reveal potential changes in students’ perceptions regarding the nine aspects addressed by the survey.

125 items were analysed by testing the Null hypothesis that there were no changes between 2011 and 2012 in the way the two groups of students responded to the survey. For 122 items the Null hypothesis was retained, two items turned out to be statistically significantly different on a margin on $p=0.05$ and for one item $p$ was smaller than 0.05. Effect sizes for all three items were small with $r<0.2$. These results suggest that the survey produced robust results on a year-to-year basis.

The soundness of the survey instrument was demonstrated by reporting on significant differences in students’ ratings over time (within subjects comparison) as well as between groups of different achievement levels (between subjects comparison) based on the combined 2011 and 2012 data set. Moderate significant correlations with students’ final exam result could be observed for some aspects, e.g. students grade expectations, the level of frustration and their satisfaction with their learning progress. The observed differences as well as the correlations supported common sense assumptions, which in turn strengthened our belief in the general validity of students’ responses.

Finally students’ responses from 2011/12 for the 125 survey items were compared with the response in 2013 to test the Null hypotheses that there are no changes between the years. We could reject the Null hypothesis for 10 items and conclude that students perceived lab tasks 4, 7, 8 and 9 as being more difficult as well as feeling less confident while working on lab task 7 or to repeat lab task 4 and 9 without help. These changes are conclusive for all 7 items concerned with task difficulty and confidence. The remaining three items concerned with the learning progress are inconclusive as two indicated that students in 2013 felt more
positive about the learning progress they made by working on lab task 4 and 12, but less satisfied with the learning progress in general.

When considering the infographic it turned out that lab task 7, 8 and 9 in fact have been labelled as particular hard by the choice of student comments and the peak in the difficulty graph. This observation leads to the assumption that the infographic might have had a “priming” effect on students’ perception of task difficulty and their confidence: telling students that these tasks are difficult leads to a perception of increased difficulty and less confidence. Even though the changes in students’ perceptions of task difficulty and confidence are conclusive and statistically significant, effect sizes are small and indicate subtle changes rather than strong effects.

The difficulty graph in the infographic was intended to show students the increasing task difficulty as well as trigger a sense of achievement for particular hard tasks. That some students in fact felt positive when comparing their own experience with the information given, can be illustrated by the following comments students submitted with their survey responses (punctuation and typos adjusted in some cases):

"Wasn’t even near as bad as what others were saying, pretty easy.” (lab task 8)

"They must have toned this lab down from last year because this definitely wasn’t the hardest one so far.” (lab task 9)

"This labs difficulty wasn’t as bad as I expected. Stayed an extra hour (first time I’ve had to do it) so I think more than anything it’s a very large (and complex) lab!” (lab task 9)

"Was pretty hard but not as hard as everyone was making it out to be.” (lab task 9)

"I enjoyed this lab, felt good to be able to complete it with out to much trouble.” (lab task 9)

"This lab boosted my confidence as it was large and required lots of thinking and application of knowledge we have gained thus far. Managed to get through it without having to ask for assistance.” (lab task 9)

In conclusion the online survey results indicated few significant changes in the way students reported in 2011/12 and 2013 their confidence, as only 10 of 125 items were statistically significantly different with small effect sizes. This exploration showed that, in contrast to our beliefs, the infographic did not cause strong effects on students’ perception and expectations in regards to task difficulty, confidence, progress made and final grade expected.
7.5.2 Changes in students’ study behaviour: Naturally occurring course data

In this section we take a closer look into available course data to answer the question about actual change in students’ study behaviour, which can be interpreted as an immediate or latent reaction to the feedback provided. Unpaired T-tests are used to test the Null Hypothesis stating no change between the experimental group in 2013 and the control groups from 2011 and 2012. We expect to see that some data sets indicate changes, which can be interpreted as students’ reaction to the information provided. To demonstrate the degree of robustness of the data sets involved, the comparison of the cohorts in 2011 and 2012 is included and no changes are expected. Only consistent changes between the data set of 2011 and 2013 as well as 2012 and 2013 indicate that students possibly reacted on the infographic and changed their study behaviour accordingly.

Indicators for students study behaviour are based on the following data sets available as naturally occurring course data: (1) attendance of lab sessions (UA 1/A), (2) number and timeliness of submissions (UA 1/B), (3) the time students spend in labs (UA 1/C) and (4) the number of students who dropped the course (UA 1/G).

(1) Attendance of lab sessions (UA 1/A)

In search of subtle behavioural changes the attendance records of 2011, 2012 and 2013 (UA 1/A) are compared. We would expect that students, based on the information that consistent work is important for the success in the course, would attend the lab sessions more frequently than in the years before. Especially the illustration of increasing task difficulty might lead to a higher initial motivation to attend the lab sessions.

Attendance records are kept for the lab sessions as students need to attend a minimum of 21 lab sessions to be allowed to sit the final examination (terms requirement). Attendance is defined as either having completed the lab task before the end of the scheduled session (which allows students to work at home and only submit the task during their lab time) or being in the lab for the duration of the scheduled session. Reasons for absence can be discussed with the lab coordinator and are in general accepted if students intend to catch up.

Figure 7.7 compares students’ attendance for 2011 and 2013 on a weekly basis. Week 13 is excluded because no reliable attendance records were kept for the last week. Changes between the years appear insignificant, however in 6 of the 12 weeks, students in 2013 attended the lab sessions on a slightly higher level. In 2011 and 2012 an average of 86.6% of students sitting the final exam attended the lab sessions. The average is slightly higher in 2013 with 88.8%.
Figure 7.7: Weekly attendance (UA 1/A) for students completing the course in 2011, 2012 & 2013

An unpaired T-test was used to test the significance of the observed differences between the years for the number of attended lab sessions per student for three aspects: (1) number of attended lab session during week 1 to week 11, (2) number of attended lab session week 1 to week 5 to investigate the initial attendance and (3) number of attended lab sessions in week 8 to week 11 to investigate the attendance in the second half of the semester. The results are summarised in Table 7.4.

The Null hypothesis that the number of attended lab sessions was similar for all three years was rejected for the attendance in the second half of the semester between 2012 and 2013. Students in 2013 attended on average 6.87 lab session out of 8 possible as compared with 6.44 in 2012. The effect size of this change is small with d=−0.250. By splitting student cohorts into higher and lower achieving halves of students based on their final exam score we could see that this change was significant for the group of lower achieving students.

<table>
<thead>
<tr>
<th></th>
<th>Week 1 to week 11</th>
<th>Week 1 to week 5</th>
<th>Week 8 to week 11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2011/2012</strong></td>
<td>no significance</td>
<td>no significance</td>
<td>no significance</td>
</tr>
<tr>
<td><strong>2011/2013</strong></td>
<td>no significance</td>
<td>no significance</td>
<td>no significance</td>
</tr>
<tr>
<td><strong>2012/2013</strong></td>
<td>no significance</td>
<td>no significance</td>
<td>no significance</td>
</tr>
<tr>
<td><strong>N=174/172</strong></td>
<td>M=6.44/6.87</td>
<td>SD=1.99/1.40</td>
<td>t(344)=−2.326</td>
</tr>
<tr>
<td><strong>p=0.021</strong></td>
<td>d=−0.250</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>lower half</strong></td>
<td>no significance</td>
<td>no significance</td>
<td>no significance</td>
</tr>
<tr>
<td><strong>2012/2013</strong></td>
<td>no significance</td>
<td>no significance</td>
<td>no significance</td>
</tr>
<tr>
<td><strong>N=87/87</strong></td>
<td>M=5.87/6.53</td>
<td>SD=2.16/1.55</td>
<td>t(172)=−2.300</td>
</tr>
<tr>
<td><strong>p=0.023</strong></td>
<td>d=−0.351</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As this change is not consistent for both years in the control group when compared with the experimental group of 2013, we cannot conclude that it was caused by the infographic.

An additional comparison of the combined cohorts of 2011 and 2012 with the 2013 cohort did not result in any significant differences between the two groups. Taking into account
that 80% to 94% of students sitting the final exam attend their scheduled lab session, the room for improvements is limited and strong positive effects cannot be expected. However the investigation suggested that the intervention had no negative impact on students’ attendance either.

(2) Time spend in Labs (UA 1/C)

In the previous section it was shown that students attend the lab sessions on a regular basis, a behaviour which is encouraged by the terms requirements of at least attending 21 of 25 lab sessions. In contrast, the time students spend in the labs is not regulated to such a degree. Students are free to leave the lab when they are finished with the work as well as to stay for a longer time if they need too. It is important to note that students do not necessarily work on their tasks for the duration of the time they are logged on. For this reason and occasional technical problems (e.g. students not being logged out) time data need to be treated with caution.

In this section time data (UA 1/C) are explored to quantify students’ activity in the lab and detect changes in students behaviour of using the lab. However spending more time on average on a task does not necessarily indicate higher success in the final exam. There is a significant negative correlation between students’ average time on task (calculated by dividing the total time spend in the lab with the number of submitted tasks) and their final exam result considering all 511 students sitting the final exam in 2011, 2012 and 2013 ($r=-0.394$, $p<0.01$, $N=511$). When comparing the two groups of higher ($M=76.01$, $SD=40.87$, $N=256$) and lower ($M=107.01$, $SD=51.59$, $N=255$) achieving students we see a significant differences ($t(509)=-7.53$, $p<0.001$) of lower achieving students spending on average 30 minutes more per submitted lab task. In this regard students showing a higher activity by spending more time in labs does not indicate potential higher success, but can be seen as a measure of dedication and trying harder.

Figure 7.8 shows the proportion of students regarding the time they spent on average on one submitted lab task. In all three years over 50% of students spent between 60 and 150 minutes per task. Each year there are approximately 7% of students on the extreme ends of spending under 30 minutes or over 210 minutes. Students with extremely short times per task were most likely working at home and only submitting their tasks during their lab time. Extremely long task times indicate that the computers in the labs were used for other activities as well.
Figure 7.8: Frequency of average time spend per submitted task (UA 1/C) of students completing the course in 2011, 2012 & 2013

From the frequency chart it seems that students in 2012 spent less time on their task than in 2011 and 2013. This observation was confirmed by an unpaired T-test, which resulted in significant differences between 2011 and 2012 as well as 2012 and 2013 indicating that students in 2012 spent significantly less time per submitted lab task than in 2011 and 2012. The comparison of 2011 and 2013 resulted in no significant differences (Table 7.5).

<table>
<thead>
<tr>
<th></th>
<th>Average time spend in lab per submitted lab task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Table 7.5: T-test statistic for year-to-year comparison of average time spend on a completed task (UA 1/C)</td>
</tr>
</tbody>
</table>
| 2011/2012| N=165/174  
M=97.58/82.79  
SD=56.39/41.38  
t(337)=2.741  
p=0.006  
d=0.299 | |
| 2011/2013| no significance | |
| 2012/2013| N=174/172  
M=82.79/94.42  
SD=41.38/47.51  
t(344)=-2.428  
p=0.016  
d=-0.261 | |

Alternatively we observed the correlation coefficients of students’ average time per submitted task and their final examination result, showing a weak negative correlation for the cohort in 2012 (r=-0.278, p<0.01, N=174) as compared with moderate negative correlations in 2011 (p=-0.420, p<0.01, N=165) and 2013 (p=-0.545, p<0.01, N=172). The effect of higher achieving students spending less time in the labs per submitted lab task than lower achieving students was consistent for all years, but the strongest in 2013.

Given these results we cannot claim that students reacted on the feedback provided in 2013 in a way that they spent more or less time on average to complete a lab task.
(3) Number and timeliness of submissions (UA 1/B)

Performance indicators in the infographic were concerned with the correlation of the final exam result with (1) number of submission by week 5, (2) number of submissions by week 11 and (3) an average value of the timeliness per submission by week 8. All three performance indicators were set out to illustrate the positive relationship of tasks completed with the final exam results as well that it is important not to fall behind with the lab work. If students not only attended to and understood this information but also reacted to it, we would expect to see some change of how students submitted their lab work.

![Percent of expected submissions per week in 2011, 2012 and 2013](Image)

**Figure 7.9: Percent of electronically submitted tasks (UA 1/B) per week in 2011, 2012 & 2013**

By comparing the percent of electronically submitted tasks per week of the expected tasks (expected is the submission of two tasks per week for each student who is still active in the course) on a year-to-year basis, we see some variation between the years (Figure 7.9). It seems that students in 2012 submitted their lab work at a higher rate than in 2011 and in 2013. However there is an overall pattern of high submission rates in the first three to four week followed by a drop in submissions in week 5 (caused by the increasing task difficulty of lab task 8 and lab task 9) as well as week 9 (caused by lab task 17 which is not electronically submitted by most students as it is not a programming task) and a final high submission rates for week 12 and week 13 when students caught up by submitting missing lab tasks.

To investigate the significance of the observed variations between the years an unpaired T-test was carried out to test the Null hypothesis, that submission numbers and timeliness were the same in 2011, 2012 and 2013. In case the illustration of the performance indicators influenced how students submitted their lab work, we would expect to see consistent changes between 2011 and 2013 as well as between 2012 and 2013. For the exact dating of the submission the electronically submitted lab tasks (UA 1/B) were used.
Table 7.6 presents the results of the unpaired T-tests for the total number of submission as well the numbers by week 5 and week 11 (2\textsuperscript{nd} and 5\textsuperscript{th} infographic performance indicator) for all 3 years. In addition the overall timeliness and the timeliness of submissions by week 8 (4\textsuperscript{th} infographic performance indicator) are included. Students who did not complete the course were removed from this analysis.

Table 7.6: T-test statistic for year-to-year comparison of number and timeliness of submissions (UA 1/B)

<table>
<thead>
<tr>
<th></th>
<th>Overall Performance Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. submissions Total</td>
</tr>
<tr>
<td>2011-2012</td>
<td>N=166/174 M=20.25/21.18 SD=4.07/3.72 t(338)=-2.203 p=0.028 d=-0.238</td>
</tr>
<tr>
<td>2011-2013</td>
<td>no significance</td>
</tr>
<tr>
<td>2012-2013</td>
<td>N=174/172 M=21.18/20.17 SD=3.72/4.19 t(344)=2.369 p=0.018 d=0.255</td>
</tr>
</tbody>
</table>

Statistical significant differences can be observed for the total number of submissions per student. There was an increase between 2011 and 2012 of 0.93 lab tasks. No change was apparent for the comparison of 2011 and 2013. A decrease of 1.01 lab tasks per student was significant between 2012 and 2013. These changes indicate that the 2012 cohort submitted approximately one task more on average than cohorts in 2011 and 2013. From this analysis it cannot be concluded that the use of the infographic influenced students submission behaviour.

In a second step the cohorts of students were divided, based on their final exam score, into upper and a lower achievement groups to test if there are changes on how those groups submitted their lab work (Table 7.7).

The analysis showed consistent changes for the year of intervention for the total number of submitted lab tasks as well as the number of submissions by week 11. In both cases a decrease of submitted lab tasks for the higher achievement group can be observed (indicated in bold font). Given the consistency of the change we could conclude that the communication of the performance indicators caused higher achieving students to feel more relaxed about the submission of their lab work. This effect was most apparent between the 2012 and 2013 cohort with a medium effect size of $d=0.468$ for the total number of submissions.
Table 7.7: T-test statistic for achievement groups in a year-to-year comparison of number and timeliness of submissions (UA 1/B)

<table>
<thead>
<tr>
<th>Overall</th>
<th>Performance Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. submissions Total</td>
<td>Timeliness Total</td>
</tr>
<tr>
<td>upper half 2011 2012</td>
<td>no significance</td>
</tr>
<tr>
<td>lower half 2011 2012</td>
<td>N=83/87</td>
</tr>
<tr>
<td>upper half 2011 2013</td>
<td>no significance</td>
</tr>
<tr>
<td>lower half 2011 2013</td>
<td>no significance</td>
</tr>
<tr>
<td>upper half 2012 2013</td>
<td>no significance</td>
</tr>
<tr>
<td>lower half 2012 2013</td>
<td>no significance</td>
</tr>
</tbody>
</table>

(4) Number of students dropping the course

Each year a number of students decide not to complete the course for various reasons. In this section the numbers of students dropping the course at different stages are compared on a year-to-year basis to investigate if 2013 was any different from the other two preceding years. Table 7.8 is a summary of the number of students who (1) started with the course (submitted at least one lab task) and did not sit the mid-semester exam or (2) dropped the course after the mid-semester test. Students who dropped the course after mid-semester are split into two groups: (2a) scoring 50% and over in the exam and (2b) scoring under 50%. Students who were enrolled but never submitted any lab work are excluded from this investigation, as we need to assume that those students have never been actively attending a lab session and were not exposed to the feedback provided.
Table 7.8: Groups of students dropping the course in 2011, 2012 and 2013

<table>
<thead>
<tr>
<th>Year</th>
<th>Year compl/enrol. FE&gt;50/enrol.</th>
<th>(1) Submitted lab work only</th>
<th>(2) Sitting mid-sem. but not final exam</th>
<th>(2a) Mid-sem. Result &gt;=50%</th>
<th>(2b) Mid-sem. Result &lt;50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>166/185=90% 126/185=68%</td>
<td>N=9 (4.9%) AV Sub. Tasks = 4.4</td>
<td>N=10 (5.4%) AV Sub. Tasks = 8.5 AV MS score = 56.3</td>
<td>N=6 AV Sub. Tasks = 9.3 AV MS score = 67.7</td>
<td>N=4 AV Sub. Tasks = 7.2 AV MS score = 38.7</td>
</tr>
<tr>
<td>2012</td>
<td>174/202=86% 120/202=59%</td>
<td>N=7 (3.5%) AV Sub. Tasks = 4.3</td>
<td>N=21 (10.4%) AV Sub. Tasks = 11.8 AV MS score = 49.7</td>
<td>N=10 AV Sub. Tasks = 14.5 AV MS score = 66.6</td>
<td>N=11 AV Sub. Tasks = 9.3 AV MS score = 34.3</td>
</tr>
<tr>
<td>2013</td>
<td>172/207=83% 127/207=61%</td>
<td>N=8 (3.7%) AV Sub. Tasks = 3.5</td>
<td>N=27 (13.0%) AV Sub. Tasks = 7.9 AV MS score = 39.6</td>
<td>N=6 AV Sub. Tasks = 10.2 AV MS score = 67.7</td>
<td>N=21 AV Sub. Tasks = 7.2 AV MS score = 31.5</td>
</tr>
</tbody>
</table>

During the 3 years 594 students (185 in 2011, 202 in 2012 and 207 in 2013) submitted at least one lab task. 82 of those students did not complete the course because they did not sit the final exam. The lowest completion rate (first column) was recorded for 2013 with 83% compared with 86% in 2012 and 90% in 2011. The lowest rate for students scoring 50% and over in the final exam was recorded for 2012 with 59%. The number of students deciding before the mid-semester exam to drop the course (second column) is similarly low between the years with 3.5% to 4.9%. Bigger variations are apparent for the number of students who did not sit the final exam after sitting the mid-semester exam (third column) with 10.3% for 2012 and 13% for 2013 as compared with 5.4% in 2011.

When taking a closer look how those groups of students are composed, 2013 shows the highest proportion of students scoring under 50% and dropping the course (Figure 7.10).

Considering that the infographic communicated the fact that over 75% of students scoring in the mid-semester exam under 9 points out of 15 (60%) failed the final exam, we might conclude that students in 2013 reacted to this information. In fact in 2011 and 2012 only 9.6% (5 out of 52) of students scoring under 50% in the mid-semester could score 50% and over in the final exam, which is the requirement for passing the course. In 2013, 26% (6 out of 23) of students scoring under 50% in the mid-semester exam passed the final exam with 50% and over. We can assume that students in 2013 were more inclined to drop the course when they realised they are in danger of failing the final exam based on their mid-semester exam score and the performance indicator. However we cannot prove these assumptions.
Despite the lowest completion rate in 2013 with 83%, the proportion of students scoring over 50% in the final exam was still larger than in 2012 (61% vs. 59%, Table 7.8, first column). So a lower completion rate does not necessarily lead to lower success rate if students, who would fail eventually, decide not to complete the course. Those students are better advised to drop the course and concentrate on other courses when catching up or working harder is not an option. Of course we cannot know if the 21 students would have failed the final exam, but if so, 11.4% (6 out of 44: 23 students really sitting the exam and 21 dropping the course) scoring under 50% in the mid-semester exam would have passed the final exam in 2013, which is comparable with the proportion of 9.6% (5 out of 52) in 2011/12.

In conclusion the drop-out rate in 2013 was highest of the three years compared with an almost double as high proportion of students deciding not to sit the final exam who scored under 50% in the mid-semester exam. It is possible, but still speculative, that those students reacted to the infographic, which showed that a low mid-semester exam indicates a high danger of failing the final exam.

**Summary**

Naturally occurring course data were explored in search of subtle changes in students’ study behaviour, which could be interpreted as a reaction to the information provided in 2013. The following data sets from 2011 and 2012 were compared with the data set from 2013: (1) attendance of lab sessions, (2) number and timeliness of submissions, (3) the time students spend in labs and (4) the number of students who dropped the course. Any consistent, significant changes apparent between 2011 and 2013 as well as 2012 and 2013 were of particular interest.

The analysis of the attendance data resulted in no consistent changes as described above. The Null hypothesis was rejected for the attendance in week 8 to 11 in 2012 and 2013 indicating slightly higher attendance for the half of lower achieving students in 2013. As measures of attendance have been not significantly different between 2011 and 2013 we cannot conclude that the feedback intervention influenced the way students attended the lab sessions. The observation of the data sets also revealed that on average 86% (in 2011) to 89% (in 2013) of students attended the lab sessions and we concluded that on such a high level significant improvements can hardly be expected.

By analysing the time students spent in the labs (based on their login times per submitted lab task) we found that students in 2012 spent significantly more time on their tasks than in 2011 or in 2013. No significant differences could be observed between 2011 and 2013, which led to the conclusion that the infographic was not influencing the time students spent in labs.
A consistent change was observed when analysing the number of submissions per student. The group of higher achieving students submitted fewer lab tasks in 2013 than in 2011 and in 2012. This change was consistent for the total number of lab tasks as well as for the number of lab tasks by week 11. If and how this change can be interpreted either as a reaction to the infographic or due to other factors influencing the 2013 cohort is speculative, but one possible explanation was that higher achieving students felt more confident to achieve a good grade based on the performance indicators given, than in the previous years which resulted in a lower number of finished lab tasks.

The last investigation of students who dropped the course showed a higher number of students dropping the course after mid-semester exam in 2013 than in 2011 and 2012. The majority of the students dropping the course at this stage in 2013 had a mid-semester exam result of under 50% (21 out of 27). As the infographic showed that over 75% of students with a mid-semester result under 9 points (out of 15) failed the final exam in 2011 and 2012 we concluded that it is possible that students reacted to this information and decided not to sit the final exam, which resulted in a higher pass rate for the group of students scoring under 50% in the mid-semester exam and sitting the final exam.

In conclusion the only consistent changes in students’ study behaviour in 2013 based on naturally occurring course data from 2011 to 2013 appeared to be the number of submissions of the group of higher achieving students. The attendance of lab sessions and the time spent per lab task was not affected. An interesting observation was seen in the fact that in 2013 the highest number of students dropped the course in the second semester after sitting the mid-semester exam and that 21 of those 27 students had a mid-semester score under 50% connected with a high danger to fail the final exam.

### 7.5.3 Students attitudes towards diagnostic course data

The investigation of students’ self-reported confidence data (7.5.1) and naturally occurring course data (7.5.2) showed that changes in students’ perceptions and actual study behaviour are either difficult to measure or basically not existent. The results from the year-to-year analysis appeared robust indicating little or no change. Rather than looking for change, in this section students’ attitudes towards diagnostic course data are discussed. This discussion is based on the questionnaire responses (UA 4) as well as on the interviews, which were held at the end of the semester (UA 5). The questionnaire at the end of the year included some questions to assess how useful and influential students perceived the information provided and to get some information about how students valued the diagnostic course data in general.
The fact that 94% of the students responding to the questionnaire reported that they referred back to the infographic at least once and that 80% used the performance indicators to check their own results in relation to the diagnostic information (see Section 7.4 Stage 4: Prolonged attention), is an indicator that those students found the information relevant enough to do so. Besides questions about the frequency of use of the infographic, there were also questions addressing the relevance and perceived usefulness of the single infographic sections. Table 7.9 provides an overview about how students answered those seven aspects of the questionnaire.

Table 7.9: Questionnaire results (UA 4) – Questions evaluating relevance and perceived usefulness of the infographic

<table>
<thead>
<tr>
<th>Section “What students say”</th>
<th>% pos.</th>
<th>% neu.</th>
<th>% neg.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aspect 1:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Seeing other students’ difficulty ratings and comments was a good preparation on what to expect in this course.</em></td>
<td><img src="chart1" alt="Bar chart" /> 76% 16% 8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Aspect 2:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>I wish other courses would provide similar information.</em></td>
<td><img src="chart2" alt="Bar chart" /> 82% 16% 1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section “What we say”</th>
<th>% pos.</th>
<th>% neu.</th>
<th>% neg.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aspect 3:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Most of the information in this section was relevant to me.</em></td>
<td><img src="chart3" alt="Bar chart" /> 52% 37% 11%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The section “What students say” was well valued and 76% agreed that the information was a good preparation to what to expect in the course (aspect 1) and 8% of students disagreed on this item. Similar reactions were reported during the interviews:

“Even early on, it [refers to “What students say” section] showed to me that it wasn’t to be a super breezy course.” … “I didn’t find it super frustrating, because I think I knew that this wasn’t going to be an easy lab [lab 9] … so I found it helpful.” (P6)
Students also commented on how the information influenced the way they prepared for the labs. Regarding the question if and how the infographic influenced their approach to and their study in the course (Appendix F, section 3c, question 4) one participant answered:

“I think so, definitely, with that part [refers to “What students say” section] I would look at it and it influenced the way I went into certain labs if I went in casually or … if I would do all the readings because this looks hard.” (P5)

“This made me more aware especially this … [refers to “What students say” section] … is the week with the two highest labs and I might need to do a bit of work beforehand.” (P3)

So it comes to no surprise that the 82% of students agreed with the follow up item that they would like similar information in other courses as well (aspect 2) and only 1% disagreed with this statement.

Most information in section “What we say” was perceived as relevant (aspect 3) by half of the students (52%). A rather high proportion of 37% rated the relevance in a neutral range and 11% disagreed on the relevance. In fact none of the participants in the interviews commented on this section in particular.

The last section “What the numbers in 2011/2012 said” was addressed by four items regarding the usefulness as well as the impact. The item “It was clear form this section that working consistently pays off at the end” was followed by the item “This kind of information was interesting” (aspect 4) intended to assess the relevance of the entire section of diagnostic information. It is possible that students interpreted this item as referring to the consistency of work only. 36% of students agreed that the information was interesting, but also 25% disagreed on this. In contrast to the low rating of how interesting students thought the information in this section was, 70% agreed that the information provided them with a good feeling on how they are going in the course (aspect 5) and 7% disagreed. The two follow-up items about the impact of the information in terms of feeling more or less confident (aspect 6) as well as working harder or less hard (aspect 7) provides a mixed picture: Even though almost half of the students said they felt more confident, nobody responded that he or she worked less. 61% of students reported that they worked “as normal” which is not necessarily a contradiction. 32% of students reported that they worked harder even though 40% of those students (14 of 35) reported they also felt more confident. It seems that the infographic had a motivational impact, as some students perceived that they worked harder when they felt more confident. The Null hypothesis, that there is no difference in the way higher and lower achieving students responded to the 7 aspects in Table
7.9, could be rejected for aspect 7 indicating that lower achieving students (N=37, mean rank=31.73) reported to a higher degree that they worked harder than higher achieving students (N=36, mean rank=42.42) by applying a Mann-Whitney U test: U=861.00, z=2.511, p=0.012, r=0.29.

During the interviews three students also mentioned the motivational aspect. When asked if and how the performance indicators helped to understand their own progress (Appendix F, section 3c, question 5) two answered:

“I got quite a few labs behind at one point. I think I was like six labs behind … and I just kept looking at this and I thought I have to catch up … that was the motivation because showing how many labs are … and I am in the red zone … oh my gosh.” (P4)

“I would say so.” [refers to indicator “Number of labs by week 11”] “There was a time when I was a little behind and I just … obviously you have a look and you like see if you do a bit of more work …” (P3)

Participant P4 also reported that as a consequence she went to additional tutorials, because the infographic showed her that catching up would mean a difference of a B or C grade. She found this information important, because otherwise she would have thought that the labs did not really matter. Regarding the questions of the use of the performance indicator section (Appendix F, section 3c, question 5) one students answered:

“This sort of scared me a little bit because I did fall a bit behind before the mid-semester, so I was worried, but it felt I have to go up against these odds and I can do this, I can beat this. Where is if I didn’t have this sort of thing I might have felt oh it doesn’t matter too much, it won’t affect it so much.” (P6)

In more general terms all participants commented on the usefulness of the infographic to prepare them for the course, e.g.:

“It was like a real good an eye-opener to what the course is gonna be like it is quite realistic to the labs too … so did sort of keep me on track a bit … it did got me to worry a little when I saw the results.” (P1)

“[It] helped to setting the goal for me … like completing all the work, try not fall behind and make sure that it is submitted it and done, and give me a kind of idea where I am sitting and what I need to do.” (P2)
“In the first couple of weeks I found it quite useful, because … you don’t know really what to expect.” (P3)

“I think it was really good … in a course like this it is important seeing the pass rates and stuff or this [not sure what section is referred to] … because like it lets you to know what you have to do. … I think it is kind of important … that you have shown us kind of like how many people like failed or like passed.” (P4)

How students value the provision of diagnostic data in general was addressed by 3 items, which referred to students’ awareness and the use of diagnostic information (Table 7.10).

**Table 7.10: Questionnaire results (UA 4) – regarding diagnostic course data in general**

<table>
<thead>
<tr>
<th>General</th>
<th>% pos.</th>
<th>% neu.</th>
<th>% neg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I was surprised that there is information indicating students’ later success or failure in a course.</td>
<td>44%</td>
<td>37%</td>
<td>19%</td>
</tr>
<tr>
<td>Such diagnostic information should be shared with students.</td>
<td>84%</td>
<td>12%</td>
<td>4%</td>
</tr>
<tr>
<td>Diagnostic information about my final performance would help me: (please tick all that apply) to balance my workload better to take action when necessary to reach my goals for a course to stay focused to take more responsibility to feel more in control of my learning. other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I don’t think it would help.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I don’t think it would help.
Almost half of the students (44%) were surprised that there is information to predict later success or failure in a course. 19% have been apparently aware that such information exists and disagreed with the statement. The next item referred to the importance of sharing such information: 84% answered that such information should be shared with the students and 4% disagreed with the statement. For a follow up item, about how students think diagnostic information would help them, the most popular choice was “to balance my workload better” (68%) followed by “to take action when necessary” (61%) and “to stay focussed” (60%). Around half of the students also ticked “to feel more in control of my learning” (51%) and “to take more responsibility” (49%). Only 5% answered that they don’t think it would help.

When testing the Null hypothesis that there is no difference in students’ final exam results for students selecting an option and the group of students not selecting an option on how diagnostic course data would help, only the item “to take more responsibility” turned out to be statistically significant. Students selecting this option did not as well as students not selecting this option. In other words, lower achieving students were more inclined to select this option than higher achieving students.

Summary

A discussion of students’ attitudes towards diagnostic course data was based on questionnaire responses and six interviews administered at the end of the course. Students’ positive responses towards the relevance and usefulness of the infographic, especially for section “What students say” and “What numbers in 2011/12 said”, was backed up by statements from the interviews.

76% of the students responding to the questionnaire agreed that the difficulty graph was a good preparation for what to expect in the course and 82% wished that other courses would provide similar information. Students interviewed reported that seeing the difficulty of the tasks was helpful and influenced the way they prepared for the labs. The numbers section was seen as providing a good indication for students’ own progress by 70% of the students responding to the questionnaire and 47% reported that they felt more confident as a result. An additional motivational aspect was revealed by the interviews as three of the six students reported that the correlations between finished lab tasks and final exam score motivated them to hand in more work when they had fallen behind. All six participants reported that the infographic was useful to prepare them for the course in general.

This general sense of usefulness was reflected by 84% of students responding to the questionnaire that diagnostic course data should be shared with students and most students responded that such information would help them to balance their work load better (68%) or “to take action when necessary” (61.3%). Answer choices indicating an interest in more
self-regulation were also selected by approximately half of the students: “to feel more in control of my learning” (51%), “to take more responsibility” (49%) and to reach my goals for the course” (48%). Only 5% responded that they don’t think such information would help them.

### 7.5.4 Discussion

Students’ immediate and latent reactions to the feedback provided were interpreted in this section as changes in their perception and study behaviour. To assess these changes we investigated self-reported confidence data (UA 2) as well as naturally occurring course data (UA 1) gathered in 2013 and compared those on a year-to-year basis with data from the last two years before the infographic was provided.

The analysis of the self-reported confidence data indicated a few small differences in how students experienced the difficulty, their learning progress and their confidence regarding individual lab tasks. 10 out of 125 items indicated change, mainly concerned with higher difficulty ratings and lower confidence ratings. Interestingly three of the five lab tasks indicating differences, were illustrated in the infographic as being particularly difficult, which suggests that knowing the tasks are going to be difficult leads to a perceived higher difficulty. However some students also commented on the fact that they found it not as difficult as indicated or felt rewarded to have managed the task without help. Small effect sizes with \(\text{abs}(r)\leq0.15\) for all significant changes indicate rather small differences between the years, which should be interpreted with caution.

The communication of the task difficulty did not discourage students from attempting the most difficult tasks 8 and 9. In fact the attendance of the lab sessions scheduled for these tasks in week 5 had the highest attendance in 2013 with 91% of students attending as compared with 86% in 2011 and 88% in 2012. However an improved attendance was only significant when 2013 was compared with 2012 and only for week 8 to week 11. This increase was caused by the group of lower achieving students, but did not lead to significantly increased submission rates. In fact a decline was significant for the number of submissions in 2013 when compared with 2011 and 2012 for the group of higher achieving students.

In conclusion the analysis of the self-reported confidence data as well as the naturally occurring course data showed that these units of analysis are composed of robust data sets, which indicate no or only small differences on a year-to-year basis. Therefore the questions stated at the beginning of section “7.5 Stage 5: Immediate or latent reaction” of

- Did students’ perceptions and expectations change in regards to task difficulty, confidence, progress made and final grade expected?
and

- Did students’ behaviour change in regard to lab attendance, time and number of submission, time spend in labs, number and time of drop-outs?

need to be considered in the light of very subtle changes and of small effects sizes. Some movement was seen regarding the drop-out rates were a higher proportion of students decided not to sit the final exam after being confronted with a low mid-semester score of under 50%.

As already discussed for the attendance data set, one reason for the general robustness of the data sets might be that most students are already operating on a high level which leaves little room for improvements. This “capping effect” was described by one interviewee (UA 5) when asked if and how he was influenced by the infographic:

“Not really … because … I started just off saying like you know I gonna go to lectures, I'll do my labs, hopefully get them in on time and then you know keep working.” (P3)

This “I do my best anyway” attitude would also explain why higher achieving students appeared to be less concerned with the performance indicators. It seems plausible that if students are convinced that there is not much room for improvement, feedback targeting self-regulation is not relevant, because there is no opportunity to apply it.

In contrast to the small measurable changes of students’ perception and actual study behaviour, students responding to the questionnaire (UA 4) reported quite enthusiastically how they felt that former students’ difficulty ratings were a good preparation for the course (76%), that they wished that other courses would provide similar information (82%) and how the performance indicators provided a good indication about their progress in the course (70%). Almost half of the students (47%) reported that they felt more confident when they were looking up their own results and a third (32%) claimed that they worked harder as a reaction. That students indeed valued the diagnostic information provided was shown convincingly by the 84% of students agreeing to the statement that such information should be shared with students. The most popular answer choice on the question of how students think such information would be helpful was “to balance my workload better” (68%) followed by “to take action when necessary” (61%). Both of these rather pragmatic answer choices indicate a degree of struggle to oversee the complex nature of first year university courses and that improvement of performance might not be students’ main concern.
7.6 Stage 6: Outcome

The question of whether or not students’ learning outcome improved in 2013 is addressed in this section. Despite students’ reports of high acceptance and perceived usefulness of the infographic as concluded from questionnaire responses and interviews in previous sections, the investigation of self-reported confidence data and naturally occurring course data indicated very subtle changes regarding students’ perception of task difficulty as well as their confidence and students’ actual study behaviour in terms of lab task submissions and drop-out rates. Therefore analysing students’ actual performance in the mid-semester and final examination might shed some light on the question of whether the infographic can be considered as influential for students’ learning outcome as assessed by these examinations (Figure 7.11).

Differences between student cohorts are problematic when comparing examination results on a year-to-year basis. Potential variations, which might influence the interpretation of students’ exam results in COMP160 are addressed by comparing students’ pre-course grades. The first question to be addressed in this section can be formulated as:

- Are there differences between student cohorts in 2011, 2012 & 2013?

By comparing students mid-semester and final examination results year-to-year the second, central question of this section is approached:

- Did students’ mid-semester and final examination results in 2013 differ from previous years?

This question is investigated in more detail when looking at changes between students’ mid-semester and final examination results within subjects, elaborating on the following question:
Could students improve their final exam score to a higher degree than in former years?

Of particular interest is a group of students, which reacted to the possibility of revising for the mid-semester examination. This possibility was introduced in 2013 for the first time when students were provided with a detailed revision sheet for their mid-semester exam and the opportunity to do this revision as a catch-up option for Lab 17.

As an alternative way to investigate students' final examination results, students actual performance is observed in comparison with the indicated performance as provided by the numbers section of the infographic. For this investigation graphical charts are compared and observed visual differences are tested for their statistical significance. For example the final exam results of the group of 56 students who handed in fewer than 7 lab tasks of 9 by week 5 in 2011 and 2012 is compared with the final exam results of 30 students of the same group in 2013. Assessing the accuracy of the performance indicators in this way provides an indication of the robustness of students' final exam results. In other words, significant differences between the indication and the actual results in 2013 point to changes of the final exam results between the years.

What differences can be observed when comparing the performance indicators as provided by the infographic with students' actual final examination results?

Following the same idea of assessing changes in students' final examination results in 2013 by evaluating the accuracy of predictions, which are based on correlations observed in former years, the risk factors as stated in Section 4.4 are applied to the 2013 cohort. The related question finishing the assessment of students' learning outcome is:

How accurate was the prediction of students at risk to fail the final examination based on the model from 2011?

A discussion at the end of this chapter explores the findings from this section aiming for a firm conclusion on how students' learning outcome was influenced by the infographic.

7.6.1 Assessing variations in student cohorts based on pre-course results

When comparing students' course results on a year-to-year basis the questions of to which degree variations in the results are caused, besides other factors, by variations in the student cohorts, needs to be considered. In this section students' pre-course grades are compared in search of variations, which, if found, could act as rough baseline to interpret students results from the last three years. Unfortunately, the only data set available for such an
exploration are results from COMP150, the Python programming course run by the department in the first semester, which was taken by 61% to 67% of students (UA 1/F).

Strong correlations between students’ COMP150 and COMP160 final marks from 2011 to 2013 (r=0.651, p<0.01, N=344) suggest that students’ COMP150 results are predictive for their COMP160 results and that similar skillsets are required in both courses. Therefore COMP150 results are interesting in order to compare student cohorts independent from the intervention. COMP150 is assessed by two mastery tests, a mid-semester exam and a final exam. The two examinations vary between the years, but the mastery tests are unchanged and published in the lab book. Over the four years of 2010 to 2013, 704 students completed COMP150 by sitting the final exam. There is a strong correlation between students’ results of the first and the second mastery test (r=0.636, p<0.01, N=704) and a moderate correlation between the average of the first and the second mastery test with the final COMP150 exam (r=0.478, p<0.01, N=704). In COMP150, students’ mid-semester exam score correlated weakly with the results from the mastery tests or the final exam. Based on the fact that the mastery test (1) did not change between the years and (2) results correlated moderately with the COMP160 final exam score (r=0.409, p<0.01, N=355), we will use this unit of analysis for further investigation.

To assess potential variations between the cohorts of students who took COMP150 before enrolling for COMP160 we tested the Null hypothesis that there are no differences for the mastery test results on a year-to-year basis. In 2011, 105 students of 166 completing COMP160 did complete COMP150 previously. In 2012 there were 116 out of 174 with a COMP150 result. An unpaired T-test comparing the COMP150 mastery test results for the COMP160 students from 2011 (M=88.0, SD=19.1, N=105) with the results in 2012 (M=91.4, SD=14.1, N=116) resulted in no statistically significant difference (t(219)=-1.50, p=0.135). Given that there were no differences, the cohorts of 2011 and 2012 students were combined and compared with the cohort of students completing COMP160 in 2013. An unpaired T-Test comparing the mastery test results for students taking COMP160 in 2011 and 2012 (M=89.81, SD=16.72, N=221) with students from 2013 (M=85.34, SD=21.23, N=123) resulted in a statistically significant difference (t(342)=2.149, p=0.046), indicating that students in 2011 and 2012 performed better.

When exploring students’ examination results in the following section we should keep in mind that variations in the student cohorts were found, based on approximately 65% of students enrolled in COMP160 between 2011 and 2013, indicating that the 2013 cohort did not do as well as the students in 2011 and 2012 in the pre-course mastery tests. However the small effect size of d=0.232 does not allow to consider these variations as greatly influential for the further exploration of students results.
### 7.6.2 Assessing mid-semester and final examination scores

Examination results vary when compared on a year-to-year basis. Besides variations in student cohorts, as discussed in the previous section, one main factor is the variation in questions and tasks composing the examinations. Another factor can be seen in the date the examination is scheduled for. An examination at the end of the examination period leaves students with more time to prepare as an examination scheduled for the beginning. Speculations are also possible on how the time in the day or the location might influence students’ examination results. Unfortunately, most of these factors cannot be controlled and may result in differences in exam results. Despite the fact that 65% of the students from the 2011 and 2012 cohort performed equally in the COMP150 mastery tests, the COMP160 exam results were significantly different in those years (Table 7.11).

#### Table 7.11: T-test statistic for examination scores in 2011 and 2012

<table>
<thead>
<tr>
<th></th>
<th>Mid-Semester</th>
<th>Final Exam (FE)</th>
<th>FE MCQ only</th>
<th>FE SA only</th>
<th>Final Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>no significance</td>
<td>N=166/174</td>
<td>M=64.71/58.77</td>
<td>SD=21.46/20.34</td>
<td>t(338)=2.62</td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td>M=69.68/58.23</td>
<td>SD=16.80/17.60</td>
<td>t(338)=6.126</td>
<td>p=0.000</td>
</tr>
</tbody>
</table>

There were no statistical differences for the mid-semester exam results and the results for the short answer (SA) part of the final exam, however the results for the multiple choice question (MCQ) part differed greatly with an effect size of $d=0.666$, which is considered as a medium to large effect. Interestingly the MCQ part is the only component where students in 2012 performed significantly worse. Given that this part was not testing a different skill-set than the other exam parts, we need to assume that the variations in students’ final exam results were caused primarily by more difficult MCQ in the 2012 final examination.

To control for variations in the examination questions and to allow for more accurate comparisons of the examination results, the final exam questions in 2013 were almost identical to 2011 and the set of mid-semester questions was very similar for both years.

#### Table 7.12: T-test statistic for examination scores in 2011 and 2013

<table>
<thead>
<tr>
<th></th>
<th>Mid-Semester</th>
<th>Final Exam (FE)</th>
<th>FE MCQ only</th>
<th>FE SA only</th>
<th>Final Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>no significance</td>
<td>N=166/172</td>
<td>M=64.71/64.75</td>
<td>SD=21.46/20.46</td>
<td>t(336)=0.016</td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td>M=69.68/72.67</td>
<td>SD=16.80/17.34</td>
<td>t(336)=1.612</td>
<td>p=0.108</td>
</tr>
</tbody>
</table>

There were no statistical differences for the mid-semester exam results and the results for the short answer (SA) part of the final exam, however the results for the multiple choice question (MCQ) part differed greatly with an effect size of $d=0.666$, which is considered as a medium to large effect. Interestingly the MCQ part is the only component where students in 2012 performed significantly worse. Given that this part was not testing a different skill-set than the other exam parts, we need to assume that the variations in students’ final exam results were caused primarily by more difficult MCQ in the 2012 final examination.
Table 7.12 shows that there have been no statistically significant differences in the way students performed in the single examination components for the cohort of 2011 and the cohort in 2013. The test statistic is reported to show that differences between the means are marginally and that the direction of change is inconsistent.

Very low effect sizes show that besides the insignificance of the differences, the distribution of the means of the two groups is in fact almost identical with the lowest effect sizes for the final exam result and final mark. In order not to miss changes for groups of students the population was split into halves of higher and lower achieving students to be compared on a year-to-year basis (Table 7.13).

<table>
<thead>
<tr>
<th></th>
<th>Mid-Semester</th>
<th>Final Exam (FE)</th>
<th>FE MCQ only</th>
<th>FE SA only</th>
<th>Final Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upper half</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>N=83/85</td>
<td>no significance</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>2013</td>
<td>M=86.60/82.49</td>
<td>significance</td>
<td>significance</td>
<td>significance</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>SD=8.96/10.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>t(166)=2.757</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p=0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d=0.425</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lower half</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>2013</td>
<td>significance</td>
<td>significance</td>
<td>significance</td>
<td>significance</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| The group of higher achieving students did not as well in the 2013 mid-semester exam as the group in 2011 with a medium effect size of d=0.425. In contrast the group of lower achieving students showed no statistically significant differences in the exam results. Besides the fact that the mid-semester examinations in 2011 and 2013 have been similar but not identical, this observation of change coincides with the changes observed for the number of lab task submissions (Section 7.5.2 Naturally occurring course data, (3) Number and timeliness of submissions). Based on the submission rates there were statistically significant differences for the half of higher achieving students between 2011 and 2013 as well as for 2012 and 2013, indicating lower submission rates for the 2013 cohort. As one possible explanation of the cause we concluded that, based on the infographic performance indicators, higher achieving students might have been more relaxed about finishing their lab work. In case of the mid-semester examination higher achieving students might have not taken the preparation as seriously as in the former years where students had no indication about their progress in the course.
7.6.3 Assessing the difference between mid-semester and final examination scores

Although there was no difference indicated for the final exam results between cohorts from 2011 and 2013, in this section we focus on changes of students mid-semester results compared with their final exam results (within subjects). Given that the examinations were different in 2011/13 and 2012 results from this observation need to be treated with caution:

Table 7.14: Correlation and T-test statistic for mid-semester and final exam in 2011 to 2013

<table>
<thead>
<tr>
<th></th>
<th>Correlation: Mid-semester – Final Exam</th>
<th>Paired T-test: Mid-semester – Final Exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>N=166</td>
<td>N=166</td>
</tr>
<tr>
<td></td>
<td>r=0.851</td>
<td>M=72.62/64.71, M paired difference=7.90</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>SD=19.84/21.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t(165)=8.941</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d=0.383</td>
</tr>
<tr>
<td>2012</td>
<td>N=174</td>
<td>N=174</td>
</tr>
<tr>
<td></td>
<td>r=0.783</td>
<td>M=69.10/58.77, M paired difference=10.32</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>SD=18.46/20.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t(173)=10.550</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d=0.532</td>
</tr>
<tr>
<td>2013</td>
<td>N=172</td>
<td>N=172</td>
</tr>
<tr>
<td></td>
<td>r=0.770</td>
<td>M=70.82/64.75, M paired difference=6.07</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>SD=17.53/20.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t(171)=5.999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d=0.317</td>
</tr>
</tbody>
</table>

Strong and very strong correlations can be observed for all three years (Table 7.14, 2nd column) with similar strong correlation in 2012 and 2013. The tendency of students to perform not as well in the final examination as compared to the mid-semester examination is consistent for all years with small to medium effect sizes (Table 7.14, 3rd column). For 2012 the biggest effect size can be observed where students lost around 10% points between mid-semester and final examination. The highest average score for the mid-semester appeared in 2011 with M=72.62, but this higher performance was not maintained as students performed in the final exam very similarly to the 2013 cohort. In conclusion the students in 2013 lost points in the final exam at the lowest level with a difference between the means of 6.07, but the final examination results remain on the same level as for the 2011 cohort.

The investigation of the difference between mid-semester and final examination result is also motivated by a change we introduced to provide students with feedback on their mid-semester exam. In the years before 2013, students were allowed to collect their mid-semester exam sheet for revision, but got no detailed feedback. In 2013 we produced an individual summary (see Appendix G for an example) about the sections of the exam to be revised. This summary was handed out with the exam sheet during lab time. This way most of the students received their exam result along with a list of sections, which needed review.
Lab task 17 is designed as a catch-up lab covering a topic, which is not assessed in the final exam. Students before 2013 were advised to skip the lab task if they needed to catch-up with their other lab work. In order to motivate students to revise, we amended lab task 17 in 2013 with the possibility of doing the mid-semester exam revision instead. Revision sheets were available with similar tasks as in the exam, which students needed to solve to get their lab mark. 71% (122 students of 172) of the students completing the course were prompted for revisions. Of this group 46% (56 students of 122) did the revisions. 26% (32 students) did the lab 17 as stated in the lab book and the remaining 28% (34 students) did not submit lab 17. For further analysis we only consider students submitting lab task 17 either by doing the revisions or by completing the task from the lab book to answer the two questions:

1. *Was the mid-semester exam result significantly different for the students who chose to revise and the students who did not? (Who decided to do the revisions?)*

To address question 1 an unpaired T-test was administered and there was no significant difference found between the group doing the revisions (M=65.50, SD=15.49, N=56) and the group not choosing the revision option (M=66.51, SD=13.04, N=32) with very similar means for both groups.

2. *Did the group of students doing the revisions do significantly better in the final exam than the group which not chose to do the revisions? (Was there any impact on students’ final results?)*

As there was no difference for the mid-semester results we simply repeat the T-test for the final exam results comparing the group of deciding for the revisions (M=58.19, SD=15.67, N=56) with the group not deciding for the revisions (M=62.94, SD=17.86, N=32). Despite a mean difference of -3.74, which indicates that students doing the revisions performed not as well as students not doing the revisions, this difference is not statistically significant (t(86)=-1.00, p=0.32, d=-0.283).

The comparison of student mid-semester and final exam result showed that the decrease in scores had the lowest effect size in 2013, which was mainly caused by lower mid-semester scores and therefore should be interpreted with caution. An intervention, providing students with the possibility to work on suggested revisions for their mid-semester exam turned out to be not effective. There was no statistically significant difference for the final examination results between the group of students doing the revisions and the students not doing the revisions.
7.6.4 Assessing performance indicators

Exploring the accuracy of the performance indicators, as given by the infographic, provides an alternative perspective to answer the question in which way the learning outcome (as assessed by the mid-semester and final examination) might have changed between the cohort of 2011 and 2012 and the cohort exposed to the feedback intervention in 2013. High accuracy indicates little or no change in the way students performed in the final exam based on different performance indicators, low accuracy indicates a change in one or the other direction. Furthermore such an investigation would reveal to which degree the performance indicators need to be adjusted for further use.

Each performance indicator is split into four achievement groups indicating the predicted performance in the final exam. An indicator for one group combines three different types of information (see Section 5.5.5):

1. Size (proportion) of achievement group (number at the bottom of the donut chart)
2. Average final exam score of this achievement group (number in the middle of the donut chart)
3. Frequency of the results in the final exam for four different categories (the donut chart)

In this section each performance indicator will be observed and discussed regarding these three aspects. Unpaired T-tests are used to assess the significance of observed differences and to test if the Null hypothesis can be rejected stating that there are no differences between the final exam results for the achievement group in question in 2011/12 and 2013. It is important to keep in mind that low final exam results in 2012 (see Section 7.6.2) influenced the performance indicators based on the 2011/12 data to indicate lower final exam results than the actual results in 2013. Such findings may be considered as an improvement, but should be treated with caution.

Indicator 1: COMP150 final mark

The following table compares indicator 1 as given by the infographic (left hand side) based on students performance in 2011 and 2012 with an adaption of the indicator based on students actual performance in 2013 (Table 7.15). Apart from a bigger group of lower achieving students in COMP150 (<60pts out of 100pts) there are few differences observable. In 2013 the average score in the final exam was higher for all four groups, because the final exam results in 2012 were lower probably due to a harder MCQ part (as discussed in Section 7.6.2). In line with the observation of the charts, unpaired T-test for each achievement group, comparing the final exam results from 2011/12 with 2013, resulted in no statistical differences.
Table 7.15: Comparison of “COMP150 final mark” performance indicators

<table>
<thead>
<tr>
<th>Indicator 1 - As provided by the Infographic</th>
<th>Based on students’ actual performance in 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>70%</td>
<td>70%</td>
</tr>
<tr>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Indicator 2: Number of submitted lab tasks by week 5

For indicator 2 (Table 7.16) the group sizes are similar between the years. Again higher or equal average scores in the final exam are observable. Distributions of the final exam scores are comparable for all four groups apart from group “7 Labs” which can be observed with a higher proportion of A-grade students and an increase of the group average score by 11%.

Table 7.16: Comparison of “Number of Labs by week 5” performance indicators

<table>
<thead>
<tr>
<th>Indicator 2 - As provided by the Infographic</th>
<th>Based on students’ actual performance in 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>90% COMP160 final exam</td>
<td>90% COMP160 final exam</td>
</tr>
<tr>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>70%</td>
<td>70%</td>
</tr>
<tr>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>
Accordingly the unpaired T-test indicated a rejection of the Null hypothesis (t(89)=-2.715, p=0.008) when final exam results for this achievement group (“7 Labs”) in 2011/12 (M=54.62, SD=18.10, N=63) and 2013 (M=65.66, SD=17.38, N=28) were compared. For all other groups there were no statistically significant differences.

**Indicator 3: Mid-semester examination result**

When comparing the performance indicator based on the mid-semester results as illustrated by the infographic (here given as 0 to 15 points towards the final grade) with students’ actual performance in 2013 (Table 7.17) an overall higher performance in the final exam can be observed. A slightly smaller group of students scoring 13 points and over in the mid-semester indicates fewer students performing at a top level in mid-semester. This observation coincides with the finding that the upper half of students in 2013 performed not as well in the mid-semester exam as in 2011 (see Table 7.13, Section 7.6.2). This smaller group of students scoring 13 points and over in the mid-semester exam in 2013 (M=89.13, SD=7.55, N=32) performed significantly better in the final exam (t(111)=-2.430, p=0.17) than the group of students in 2011/12 (M=84.05, S=10.82, N=81). The distribution reveals that all students were scoring over 65%. The observable increase for the average final exam result for the lowest achievement group scoring under 9 points of 36% in 2011/12 (M=36.31, SD=14.20, N=81) to 42% in 2013 (M=42.25, SD=13.43, N=41) is also statistically significant (t(130)=-2.222, p=0.028). For the remaining two groups there were no statistically significant differences.

Table 7.17: Comparison of “Mid-semester exam result” performance indicators

<table>
<thead>
<tr>
<th>Indicator 3 - As provided by the Infographic</th>
<th>Based on students’ actual performance in 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph 1" /></td>
<td><img src="image2.png" alt="Graph 2" /></td>
</tr>
</tbody>
</table>

7 Evaluating stages of engagement in the feedback process · 147
Indicator 4: Average lateness by week 8

Group sizes for indicator 4 (Table 7.18) appear to be similar but with an observable smaller group for students being on average over 1 lab task late with their submissions. The increase of the average final performance of this group is not significant. The only statistically significant difference (t(175)=-2.629, p=0.009) was seen for the group of students being slightly late, which showed an increase in the final exam result from 60% in 2011/12 (M=60.30, SD=18.97, N=117) up to 69% in 2013 (M=68.22, SD=19.01, N=60).

Table 7.18: Comparison of “Lateness average by week 8” performance indicators

<table>
<thead>
<tr>
<th>Indicator 4 - As provided by the Infographic</th>
<th>Based on students’ actual performance in 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Graph of lateness distribution" /></td>
<td><img src="image2" alt="Graph of lateness distribution" /></td>
</tr>
</tbody>
</table>

Similar distributions and average scores show that there is no clear discrimination in the final exam results between the group of students on time and slightly late with there lab work as well as between students late one task and more than one task. In fact students in the group being approximately one lab task late on average display a lower average score of 53% than students being more than one lab task late with and average score of 55%.

Indicator 5: Number of submitted lab tasks by week 11

The comparison in Table 7.19 shows that, fewer students submitted over 19 lab tasks by week 11 in 2013. Where in 2011/12 122 of 340 students (36%) were in this achievement group, in 2013 only 46 of 172 students (27%) had submitted more than 19 lab tasks. This coincides with a significant drop of submitted lab tasks in 2013 for the group of higher achieving students (Table 7.7, Section 7.5.2). Overall students performed better in the final exam with higher average results for all four achievement groups, but none of these observed differences was statistically significant.
Table 7.19: Comparison of “Number of Labs by week 11” performance indicators

<table>
<thead>
<tr>
<th>Indicator 5 - As provided by the Infographic</th>
<th>Based on students’ actual performance in 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
</tbody>
</table>

Summary

In this section the quality of the performance indicators as provided by the infographic was assessed by comparing the indicated performance based on the data form 2011/12 with students’ actual performance in 2013. For this assessment visual differences in the charts, which communicate achievement group sizes, average performance in the final exam as well as the frequency, were discussed.

In most cases the sizes of the achievement groups appeared to be similar, which indicates that students performed comparably at different stages of the course. The significance of the observed differences regarding the final examination results per achievement group was assessed with unpaired T-tests, testing the Null hypothesis of no differences in the final exam results as given by the infographic and as based on students’ actual performance. The Null hypothesis was rejected for 4 out of 20 items (5 performance indicators consisting of 4 achievement groups each) investigated, indicating an increase in the final performance for the students in 2013 for all 4 items (achievement groups).

When looking at these differences between the cohort of 2011/12 and the cohort of 2013 it is important to keep in mind that the examination results in 2012 were significantly worse as compared with 2011 most likely due to a harder MCQ part (see Table 7.11, section 7.6.2). A comparison of students’ indicated performance based on the 2011 cohort only with students’ actual performance in 2013 for all 20 items resulted in one significant difference for indicator 4 (Average lateness by week 8) \(t(114)=-2.052, p=0.042\) for the group of
students being slightly late with their lab work in 2011 (M=61.00, SD=18.87, N=56) and
2013 (M=68.22, SD=19.01, N=60) indicating a better performance in the final exam with a
small to medium effect size of \( d = -0.38 \).

Given that 16 out of 20 items for the 2011/12 cohort (or 19 out of 20 for the 2011 cohort
only) provided an accurate indication (the Null hypotheses could not be rejected) for the
final examination results of the 2013 cohort and that the inaccurate indicators were indicat-
ing a conservative lower result, we conclude that an adaption of the performance indica-
tors is not necessary. However it can be assumed that the combination of the results of all
three years would strengthen the performance indicators and weaken the influence of the
low final exam results form 2012, as 2011 data only would have provided more accurate
indications for 2013.

7.6.5 Assessing risk factors

The definition of risk factors was discussed in Chapter 4, Section 4.4. It was shown that in
2011 70% of the students not sitting the final exam or scoring under 50% could have been
identified by week 5 based on their pre-course (COMP150) result, their attendance in the
first week and their number of submitted lab tasks by week 5. After mid-semester examina-
tion these predictions became more accurate with 84% correctly predicted. In this section
predictions based on those risk factors are assessed for the 2013 cohort of students. Again,
the aim of this investigation is twofold: firstly a high accuracy of the predictions would re-
ineforce previous observations that students in 2013 performed similarly in the final exam
compared to 2011 and 2012 and secondly the investigation might lead to conclusions to
adapt the risk factors for higher precision based on the results in 2013.

As a measure of accuracy we state the percentage of correctly indicated students (positives)
from the group of students actually scoring under 50% in the final exam or not sitting the
exam as well as the percentage of students correctly not indicated to fail (negatives) from
the group of students actually scoring over 50% in the final exam. The percentage of the
overall correct predictions for the entire population is also stated. Table 7.20 summarises
the predictions for all 3 years and provides a colour key for the charts at the end of the ta-
ble.
Table 7.20: Assessment of risk factors for 2011 to 2013

<table>
<thead>
<tr>
<th></th>
<th>By Week 5</th>
<th>After Mid-Semester Exam</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Week 5</td>
</tr>
<tr>
<td></td>
<td>N=184</td>
<td></td>
<td>Positives: 69.0</td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td>Negatives: 77.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Overall: 75.0</td>
</tr>
<tr>
<td></td>
<td>N=176</td>
<td></td>
<td>Mid-Semester</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Positives: 84.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Negatives: 73.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Overall: 76.7</td>
</tr>
<tr>
<td></td>
<td>N=201</td>
<td></td>
<td>Week 5</td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td>Positives: 56.8</td>
</tr>
<tr>
<td></td>
<td>N=195</td>
<td></td>
<td>Negatives: 84.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Overall: 73.1</td>
</tr>
<tr>
<td></td>
<td>N=205</td>
<td></td>
<td>Mid-Semester</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Positives: 76.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Negatives: 79.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Overall: 77.9</td>
</tr>
<tr>
<td></td>
<td>N=189</td>
<td></td>
<td>Week 5</td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td>Positives: 67.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Negatives: 76.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Overall: 73.2</td>
</tr>
<tr>
<td></td>
<td>N=205</td>
<td></td>
<td>Mid-Semester</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Positives: 90.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Negatives: 72.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Overall: 78.8</td>
</tr>
</tbody>
</table>

Key:
- Actual Performance Final Exam (FE)
  - % of Students
    - Not sitting FE
    - Under 50% in FE
    - Passing FE
- Predicted Performance Final Exam (FE)
  - % of Students
    - Positives
    - False Negatives
    - False Positives
    - Negatives
For a summary of the data refer to the contingency tables for 2011 and 2012 in Section 4.4, Tables 4.1 and 4.2. The results from 2013 are summarised in Table 7.21 below.

<table>
<thead>
<tr>
<th></th>
<th>2013 Week 5</th>
<th></th>
<th>2013 Mid-Semester</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>At Risk Sign</td>
<td>No At Risk Sign</td>
<td>Total</td>
<td>At Risk Sign</td>
</tr>
<tr>
<td>Failed Final Exam</td>
<td>29</td>
<td>16</td>
<td>45</td>
<td>40</td>
</tr>
<tr>
<td>Didn’t Sit Final Exam</td>
<td>24</td>
<td>9</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>Passed Final Exam</td>
<td>30</td>
<td>97</td>
<td>127</td>
<td>35</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>83</strong></td>
<td><strong>122</strong></td>
<td><strong>205</strong></td>
<td><strong>99</strong></td>
</tr>
</tbody>
</table>

\( \chi^2 = 39.95, \ p < 0.001 \)

\( \chi^2 = 71.42, \ p < 0.001 \)

The prediction of students who will not sit or score under 50% in the final exam becomes more accurate when the mid-semester results are incorporated. Students scoring under 60% in the mid-semester are identified to be at risk of failing the final exam. However false positive predictions also increase with the effect that the proportion of correctly indicated negatives (students who score 50% and over in the final exam) decreases. This trade-off can be observed for all years. However each year the incorporation of the mid-semester result leads to an improved overall prediction. The percentage of the overall correctness of the predictions are similar for all three years being between 73.1% and 75.0% for the predictions by week 5 and 76.7% and 78.8% after the mid-semester examination. The most accurate prediction of students at risk to fail the final exam was seen in 2013 after the mid-semester exam with 90.1% correctly identified from the group of students actually failing the final exam and the highest overall correctness of 78.8%.

The assessment of the risk factors showed that students’ performance in 2013 resulted in similar proportions of correct predictions for successful and failing students than in 2011, and 2012. In fact, the predictions after mid-semester examination were the most accurate in 2013 with correct predictions for 78.8% of the entire population of 189 students. Therefore an adaption of the risk factors based on the examination results of the 2013 cohort is not necessary. As 2013 had the highest proportion of students not sitting the final exam it is interesting to see that in fact 24 of the 26 students in this group were indicated to fail the final exam taking their mid-semester examination results into account.
7.6.6 Discussion

In order to evaluate changes in students learning outcome, as assessed by the mid-semester and final examination, in this section five different aspects were explored:

1. Differences in student cohorts based on students’ COMP150 pre-course grades (available for 60% to 65% of the students) could be observed when comparing the mastery test results. Students in 2013 scored significantly lower than in 2011/2012 with a small effect size.

2. Differences in mid-semester and final examination results when comparing 2011 results with 2013 results showed no significant differences for the entire cohorts. However when comparing the halves of higher and lower performing students, significant differences were observed indicating lower mid-semester results for higher achieving students with a medium effect size.

3. Differences between students’ mid-semester and final exam result when compared per student showed that the consistent, significant decrease in scores had the lowest effect size in 2013. An intervention, providing students with detailed feedback guiding revisions for their mid-semester exam was not improving their final exam results, as there was no statistically significant difference for the final exam results between the group of students doing the revisions and the students not doing the revisions.

4. Differences between the indicated performance as provided by the infographic and students’ actual final examination results in 2013 showed an overall better performance for the actual final exam results. This observation was statistically significant for 4 out of 20 achievement groups. This increase was mainly caused by low final exam results in 2012, which informed the performance indicators. By excluding the 2012 cohort, one significant difference remained, indicating an improved performance in the final exam for students being slightly late with the submission of their lab work with a small to medium effect size.

5. Accuracy of the prediction of students at risk to fail or not to sit the final exam demonstrated an increased accuracy of 90.1% correct predictions incorporating students’ mid-semester result and the highest overall proportion of correct prediction of 78.8% when compared with the predictions in 2011 and 2012.
These observations suggest that students’ learning outcome in 2013 was on the same general level as in 2011 and 2012. When accounting for the variations in the examination questions by only comparing the results from 2011, where the same final examination was written as in 2013, differences in the results are diminished further. The question to what extent the significantly lower mid-semester results of the higher achieving half of students in 2013 was caused by (1) a slightly different set of exam questions or (2) the observed variations in student cohorts or (3) by a presumed influence of the infographic, remains unanswered. It is possible that higher achieving students were more confident about their progress in the course based on the performance indicators provided by the infographic and therefore were not as well prepared for the mid-semester examination. Such a conclusion would coincide with 47% of students reporting via the questionnaire on feeling more confident after attending to the performance indicators. However, the final examination results appeared as being extremely similar. Therefore reducing the impact of the infographic on the mid-semester results only remains questionable.

By disregarding differences for just the mid-semester examination results, as discussed above, and accounting for variations caused by different sets of examination questions, students’ learning outcome as assessed by the examinations, appears largely unaffected by the feedback provided. In fact it is surprising how robust students’ examination results appear, given that there are many possible influences on how students perform in a course. We might take this as a positive finding suggesting a solid course structure and staff members operating on a reliable level.
7.7 Summary

<table>
<thead>
<tr>
<th>Infographic (IG)</th>
<th>Collection &amp; Immediate Attention</th>
<th>Cognitive Response</th>
<th>Prolonged Attention</th>
<th>Immediate or Latent Reaction</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions to evaluate degree of engagement: Did student(s)…</td>
<td>… collect and pay attention to the IG?</td>
<td>… understand the information provided?</td>
<td>… refer back to the IG? &quot;map&quot; their own performance?</td>
<td>… perceptions and expectations change? … behaviour change? … find the information useful &amp; influential? … value diagnostic course data?</td>
<td>… cohorts differ? … results improve? (mid-sem &amp; final)</td>
</tr>
<tr>
<td>Results</td>
<td>99.5% collected and paid attention to the Infographic.</td>
<td>Approximately 80% indicated no difficulties of understanding the information.</td>
<td>94% indicated that they referred back at least once. 80% looked up at least one performance indicator.</td>
<td>Very subtle changes in students' perceptions and behaviour. 70% to 80% reported on usefulness. 84% agreed that such inf. should be shared.</td>
<td>Small effect of students in 2013 performing not as well in COMP150. No changes for examination results.</td>
</tr>
</tbody>
</table>

Figure 7.12: Results for the different stages in the engagement process and outcome

Students’ engagement with the infographic was evaluated in the previous five sections following a model proposed by Price et al. (2011) and adapted for our scenario (Figure 7.12). Each stage in the engagement process was investigated to assess the proportion of students actively engaged. For most stages this investigation was carried out regarding different perspectives employing various units of analyses (e.g. self-reported vs. naturally occurring).

Besides observing averages and frequencies in the single data sets, significance tests were used to compare the control group formed by students taking the course in 2011 and 2012 with the experimental group of 2013 for different aspects regarding students engagement with the infographic at different stages (see Section 7.1 to Section 7.5). As a last step, in Section 7.6, the question to which degree students’ learning outcome, as assessed by the two examinations in 2013, was potentially affected by the information provided, was addressed. Students in 2011 and 2013, provided with the same set of final examination questions, performed almost identically. That is surprising as many factors can influence students’ performance and we actually found significant small differences between the cohorts of 2011 and 2013 based on the mastery test component in the pre-course COMP150.

Given that there was no difference in the way students performed in the final exam the question of the degree of engagement at different stages is of much interest. Which have been the hurdles in the engagement process students could not take? How to address the potential problems leading to disengagement? And were the units of analysis rich enough to answer these questions?
The infographic was collected and attended to (Stage 1 & 2) by 99.5% of students initially active in the course based on at least two completed lab tasks, which involved the attention of the infographic. There is a slight chance that students copied task solutions from other students, but such behaviour is uncommon at the beginning of the course.

Students’ cognitive response (Stage 3) concerned with the degree of understanding was quantified by analysing a sample of 36 infographic related task solutions and 5 questionnaire items. Despite the different nature of the data sets (performance data vs. self-reported data) the results are very similar: on average 82% of the collected task solution were correct and 78% of the students agreed with the statements assessing the communication of important aspects of the infographic. Given that similarity we can conclude confidently that approximately 80% of students had no trouble understanding the information provided at the beginning of the course.

The evaluation of prolonged attention (Stage 4) regarding the question whether students referred back to the infographic voluntarily, was based on students’ self-report only. In addition to the possibility of measurement errors caused by common method biases ( Podsakoff et al., 2003) there was a self-selection effect for the group of students responding to the questionnaire detected by statistically significant better results in the final exam for this group. Both effects need to be kept in mind when interpreting the results of 94% of students reporting that they at least once referred back to the infographic and 80% of students reporting that they used on average 3 out of 5 performance indicators. Again over 80% of students are still engaged as the course continues and indicate prolonged attention.

Immediate or latent reaction (Stage 5) was investigated based on self-reported confidence data, course performance data and questionnaire responses to assess students’ perceptions, expectations and attitudes as well as “visible” reactions. Interview responses were used to illustrate certain facets of the investigation. The way students responded to the confidence survey over the last three years was consistent and the experimental group in 2013 was mainly unaffected. We observed that 7 of the 10 affected items (of 125 in total) were related to small, consistent effects that students felt tasks were more difficult and they felt less confident. In contrast, some students left comments with the survey indicating that they felt positive about having managed particular hard tasks.

Potential effects on students’ actual study behaviour were difficult to track down. The data sets investigated (Lab attendance, Time spent in the labs, Number and time of submissions and Drop-out rates) appeared to be similar for all three years indicating little or no change. Consistent, statistically significant changes between 2011 and 2013 as well 2012 and 2013 were seen for the higher achieving students submitting fewer tasks in 2013, which lead to the conclusion that students felt more confident based on the information provided. This
conclusion was in-line with the 47% of students responding via the questionnaire that they felt more confident after attending to the performance indicators. However a higher drop-out rate of students scoring under 50% in the mid-semester indicated that these students felt less confident about their prospects in the course and decided not to sit the final examination, which in turn led to a higher rate of students from this group passing the final exam.

Despite very subtle behavioural changes students indeed valued the information provided. 76% of the students responding to the questionnaire agreed that the infographic section “What students say” was a good preparation for the course and 70% felt that the performance indicators provided a good indication about their own performance. In a more general sense, 84% of the students responding to the questionnaire valued the information by agreeing with a statement that diagnostic course data should be shared with students and 95% ticked at least one scenario of how such information would help them.

Considering these rather enthusiastic responses in the light of subtle actual changes in perceptions and reactions and the unchanged examination results, as mentioned at the beginning, is a bit of a puzzle. Is it possible that students would have performed worse without the information provided? The differences in student cohorts as seen by significantly lower results for the pre-course mastery tests in 2013 might suggest this possibility, even though the effect size was small. Alternatively the findings suggest that high engagement, as seen for stages 1 to 4 and partly for stage 5 in terms of valuing the information and reporting on a motivational effect, not necessarily triggered action. Price et al. (2011) pointed out that for action taking place, motivation is only one aspect and that opportunity, means and guidance are required as well. In our case students had the opportunity to catch up with their lab work, review their mid-semester exam and prepare for the final exam during practically unlimited lab time (students are allowed to attend additional lab streams and catch-up sessions are offered) and with demonstrator assistance. However on an individual level, time and resources are in fact limited and other courses may demand extra work on a similar scale. So motivation and knowing what is required is an important prerequisite but not the key for action. Inaction can be caused by various reasons and one might be that students are already operating at a high level with little room for improvement.

It can be concluded, based on the high degree of engagement, that students have been well aware about course characteristics, increasing task difficulty and the fact that working consistently is important. We can also conclude that warning signs based on COMP150 final scores and mid-semester examination results were most likely received. A high proportion of students reported that they paid attention to these performance indicators. But we also need to conclude that such information is not sufficient to trigger substantial action, which
in our case was expected to be visible as improved attendance, submission rates or final exam performance.

If such behavioural changes are not considered as the ultimate outcome of a feedback intervention, but positive students’ attitudes towards and high engagement with the feedback are likewise valued as beneficial results, which contribute to students’ development into self-regulated learners as well as contribute to a relationship of acceptance and trust between teaching staff and students, we can conclude that the infographic had a very positive impact. Opening the black box of course characteristics by offering insights which are considered as valuable for students’ self-regulated learning, is not a matter of good will, but our responsibility.
Chapter 8

Contributions and Future Work

“Confusion over expectations of what feedback can achieve, and the difficulties associated with establishing a dialogue between staff and students, should not be underestimated. Inevitably, these will impact on the extent of engagement with feedback that can be achieved. It is unlikely that, without a clear and shared understanding of the nature of feedback by all parties involved, feedback can be improved.”

(Price, Handley and Millar, 2011, p. 881)
8.1 Introduction

In the previous chapter the different stages of the engagement process with feedback were assessed and discussed. Despite a high engagement with the infographic, students’ actual study behaviour was not changed measurably for the aspects we investigated. Given these findings it is important to reflect critically on the process leading to the scoping study and the contributions resulting from this process. This critical reflection will inform the future work discussion. The questions guiding this chapter can be summarised as:

- What are the limitations of the research findings?
- How to address those in future work?

Section 8.2 presents a summary of the process, highlighting various contributions, which are discussed with the results of the scoping study in mind. Potential limitations regarding the approach and methodology are revealed and result in suggestions for future work. Selected future work ideas are discussed in more detail in Section 8.3.

8.2 Contributions – A critical reflection

High failure rates in CS1 courses testify that learning to program is hard and frustrating for a lot of students. In searching how to support first year programming students we investigated the notion of feedback from a higher education perspective. This theoretical basis was used to assess feedback practice in the CS1 context and various opportunities to integrate research from CSEd to support feedback processes were revealed. A gap in the research became apparent when searching for ways to support students’ self-regulated learning, which requires self-directed goal setting, self-monitoring and the adjustment of strategies to reach these goals. Here the provision of information about successful and risky study behaviour by communicating course characteristics and performance indicators is essential to support students’ self-regulation, but related research was not found to be established in CSEd. In the next step we analysed the predictive value of naturally occurring course data for students’ final performance at different stages in our course. All data sets correlated significantly with students’ final examination results and we were able to define risk factors as well as performance indicators. In a scoping study students were supplied with this information in form of an infographic to explore the impact of feedback on self-regulation level as well as to learn about students’ attitudes towards diagnostic course data in general. Within this process various contributions were achieved and will be discussed in more detail in the next three sections:
1. The development of a model to guide effective feedback and the application in the CS1 context.

2. The definition of predictive factors for students’ final achievement in COMP160.

3. The conducting of a scoping study providing feedback to support students’ self-regulated learning.

This project developed from a very wide, theoretical perspective on feedback on different levels into a specific scoping study to investigate students engagement with feedback on self-regulation level for one particular first year programming course. The question of what hindered students in the scoping study from acting on the information, needs to be considered and will inform the future work discussion. Section 8.2 is set out to summarise the various contributions made in the course of this project and to reflect critically on those by considering the results of the scoping study as a backdrop for the discussion.

8.2.1 A model to guide effective feedback

Given that feedback is regarded as one of the main influences on student learning, the question on how to support first year programming students was approached by a literature review of higher education research to answer the question *What constitutes effective feedback?* A number of key topics emerged from the literature review:

1. The notion of feedback as “closing the gap between desired and actual performance” was used by numerous authors (e.g. Sadler, 1989) and consequently calls for three main aspects to be addressed by feedback: (1) the clarification what the desired performance is (goals), (2) regular and detailed information about the actual performance (progress) and (3) corrective advice and strategies in relation to the actual performance to attain the desired performance (improvements).

2. A definition of different levels of feedback was proposed by Hattie and Timperley (2007). Feedback on the learner as a person (self-level) was debated as mainly ineffective, but the authors state, in increasing order of effectiveness, three additional levels (1) the non-generalizable feedback concerned with a specific task (task level) (2) the feedback directed at the processes to understand and perform tasks (process level) and (3) feedback targeting students’ ability to self-direct, -monitor and -regulate their learning (self-regulation). The importance of feedback at self-regulation level was acknowledged by other authors as well (e.g. Nicol and Macfarlane-Dick, 2006) and seen as the most “sustainable” form of feedback (e.g. Carless et al., 2011).
3. The emphasis on students’ active role in the feedback process, an underlying theme for several authors (e.g. Liu and Carless, 2006; Boud and Molloy, 2012), was manifested conclusively in a model of the temporal dimension of student engagement with assessment feedback by Price et al. (2011) where learning outcome can only be influenced when feedback was collected, attended to, understood and acted on. The authors claim that continuous dissatisfaction with feedback leads to earlier disengagement in the process to an extreme that students do not even collect assessment feedback.

4. The communication of effective feedback is characterised in the literature as: (1) being regular and timely, (2) with opportunities to apply the feedback and (3) encouraging teacher and peer dialog as an ideal form of communication to ensure that (4) feedback is personal and specific to the learner’s situation.

All four key aspects were feeding into a general framework, established at the end of Chapter 2, stating questions to guide the assessment and improvement of feedback practices for specific learning situations.

In Chapter 3 this framework, guiding effective feedback, was used to critically reflect on COMP160 as a typical introductory programming course. By addressing the question of *How to translate effective feedback practice into the CS1 context?* we wanted not only to learn about the quality of our own feedback practices in this course but also illustrate the application of the framework in practice. Whenever possible related research topics from CSEd were introduced to provide entry points into the literature, which are useful to plan and conduct feedback interventions in the CS1 context.

Different course components were considered at different levels: (1) lab work, small tasks during lectures and the two examinations on *task level*, (2) managing programming processes and course requirements on *process level* and (3) multi-stage assessments, peer feedback, self-assessment and students’ understanding of course characteristics on *self-regulation level*. On each level current feedback practices to clarify goals, actual progress and activities for improvement were discussed.

We found excellent, task-specific feedback provided in a tutor-assisted lab session, but feedback regarding examination results was neglected in our own course and revealed as a missed opportunity to guide students’ learning. On process level a discrepancy was found between our expectations of how students should manage the programming processes (e.g. in terms of planning and testing a solution) and course related processes (e.g. in terms of preparing for lab work and examinations) as well as missing explicit goals and structured feedback to guide these processes. On self-regulation level the question of what kind of
feedback supports students’ development into skilful self-regulated learners was addressed. In Chapter 3 we discussed various aspects supporting students’ self-regulation and found the topics of peer feedback and self-assessment well established in CSEd research. However, the question of on what basis students, unfamiliar with the course structure and the subject matter, decide on performance goals and self-monitor their course performance remained unanswered. No research was found in CSEd on using course performance data as feedback information. We decided to address the question of what defines successful or risky study behaviour by investigating the predictive value of course performance data at different stages for students’ final examination result and to make this information available to our students.

Considering the aim of the resulting infographic to support students’ self-regulated learning by communicating course characteristics and performance indicators, two possible shortcomings become apparent: Firstly, we need to consider the challenge of self-regulation. Given that several authors point out that self-regulation is a skill, which needs to be developed over time (e.g. Zimmerman, 1998 or Nicol and Macfarlane-Dick, 2006) and given that students’ conceptions of feedback are influenced by prior experiences, e.g. by the Dialogic Feedback Cycle in school as described by Beaumont et al. (2008) we need to ask to what degree we can expect first year university students to take responsibility and regulate their own learning. It is possible that targeting students’ self-regulated learning in a first year course is problematic because students do not yet accept the importance of their active role in the learning process. Secondly, Hattie and Timperley (2007) clearly state self-regulation as the highest level and view the different levels as interrelated: “Feedback aimed to move students from task to processing and then from processing to regulation is most effective.” (p. 91).

This perspective suggests that feedback at lower levels needs to be well established before students can “move” to a higher level. In our case the mere knowing that staying on top of the material is a key success factor is ineffective if students have no means to progress through the necessary lab work because of missing feedback on task or process level. While we think that active students get sufficient feedback on task level in the tutor-assisted lab sessions in COMP160, we raised concern about the missing implementation of regular feedback on process level. For example students with no strategies to plan a program solution will most likely struggle with the increasing complexity of the tasks. Furthermore the dependence on tutor assistance in terms of error detection and correction prevents students from working in a self-directed way, even when the tasks are understood and all necessary concepts can be applied.
The scoping study showed that submission rates and examination results remained mainly unaffected in 2013. Consequently we need to ask to what degree (1) students' readiness to self-regulate their learning and (2) flaws of the current feedback practice on task and process level impacted the effectiveness of the feedback on self-regulation level. On one hand it is possible that, although students saw the danger of failing or getting a low grade, they felt not inclined to act on these information. On the other hand it is also possible that even though they wanted to act on it, a lack of process-oriented feedback and accordingly strategies to deal with the complexity of the subject matter prevented any fruitful action and therefore improvements.

The results of the scoping study indicate that our students not only valued the information provided but also that (1) 80% of the students responding to the questionnaire paid attention to the performance indicators during the course, (2) 49% of the students could see how diagnostic course data could help them to take more responsibility for their learning and (3) 61% how the information would help them to take action when necessary. These results indicate that most of our students were aware of their active role in the learning process to some degree. How much this awareness triggered some action is hard to say. Although the average time per submitted task did not significantly increase, the negative correlation of this unit of analysis with students' final exam result was strongest in 2013. So, it is possible that lower achieving students might have attempted the lab work to a higher degree, but could not complete to a higher degree, because of the limitations discussed above.

In conclusion there are two strands to consider: (1) the developmental aspect of self-regulation and (2) the availability of effective feedback on all three levels. Both aspects will be addressed as possible future work topics in Section 8.3.

8.2.2 A case study of predictive factors for students’ final achievement

If students are expected to self-regulate their learning, course characteristics and performance indicators need to be communicated. On the basis of this information students are able to derive performance goals and self-monitoring their own performance in relation to these goals by linking their actual performance to a projected final achievement as drawn from past years’ data. The two questions of What performance indicators are available? and What indicates risky or successful study behaviour? were guiding the investigation of naturally occurring COMP160 course data from 2011 and 2012 regarding their correlation with students’ final exam result as presented in Chapter 4. All six units of analysis showed significant correlations of varying strengths (here ordered from strongest to weakest):
1. Mid-semester exam result
2. Pre-course grades
3. Number of submissions (number of completed tasks)
4. Time of submission (as average lateness per submitted task)
5. Time spent in labs (including average time per completed task)
6. Attendance of lab sessions

Furthermore, we could show that students at risk of failing the final exam can be predicted by using naturally occurring course performance data. A combination of risk factors based on (1) pre-course grades, (2) attendance of the first lab session, (3) number of submissions by week 5 and after mid-semester the (4) mid-semester exam score, resulted in a correct identification of 62% (86 of 139) of students dropping the course or scoring under 50% in the final exam by week 5 and almost 80% (99 of 125) after mid-semester for 2011 and 2012 (see contingency tables at the end of Section 4.4). Approximately 20% of students were incorrectly identified. These students did significantly worse in the final exam than their unidentified classmates. Considering the common structure of the course, we expect pre-course grades, early struggle with the lab work, mid-term test results and time on task to be general indicators for students’ final performance in CS1 courses. As the communication of risk factors and early warning signs is vital information to enable students to judge course demands and their actual progress realistically, we hope to initiate a discussion amongst CS1 teachers of integrating such information into the feedback practice.

Because our primary aim was to add meaning to students’ course performance data and not to provide a precise prediction of their final exam result, the development of a predictive model, combining several units of analysis in a linear regression model, was not pursued in this project. By keeping the single performance indicators separated, as illustrated in the infographic, we intended to show the influence of different components and thus to highlight areas for improvement or revision on a personal level. We could derive 5 performance indicators to (1) illustrate upfront that success in the course is highly influenced by the persistence to stay on top of the lab work and to (2) inform students about how their individual performance at different stages in the course relates to the performance in the final exam. Students responding to the questionnaire attested that these two goals were met as 87% of the students agreed with the statement “It was clear from this section that working consistently pays off at the end” and 70% agreed with the statement that “Looking up my own results provided me with a good sense on how I’m doing in the course”. Still 23% answered on a neutral range, which might be explained by the 20% of students reporting that they never used the performance indicators. Here the question why students did not engage with the information needs to be considered. Besides addressing the limitation regarding visual aspects and the form of delivery of the information, which will be discussed
in the next section (8.2.3), it is worthwhile considering how a precise diagnoses of the final grade by using a predictive model or the integration of other behavioural aspects could enrich the information.

Not all units of analysis are appropriate for the illustration of successful or risky study behaviour. The time students spent in the labs was negatively correlated with their final exam score, which would have resulted in a performance indicator showing that spending long hours in the labs is harmful for the final performance. We assume that longer time spent in the labs indicates a lower degree of efficiency or dedication. Here further analysis would be necessary to extract such performance indicators.

Not always is a “positive message” clear-cut. For example, the increase of the number of students dropping COMP160 after mid-semester in 2013 (see Section 7.5.2), was discussed as a possible reaction to the performance indicator “Mid-Semester Exam Score” which illustrated low chances to pass the final exam with a score of under 9 out of 15 points in the mid-semester exam. At first sight that is not a desired reaction. We argued that despite the lowest completion rate in 2013 of 83% (172 of 207) of students sitting the final exam for all three years, the success rate (students scoring 50% and over in the final exam) in 2013 was higher than in 2012. Reflecting on these results, we concluded that in 2013 students at a very high risk to fail decided, on an informed basis, to drop the course. This is by no means a negative reaction as those students could concentrate on other courses and save themselves time and frustration in our course. As another example we found that submission rates for higher achieving students were significantly lower in 2013 than in 2011/12 (Section 7.5.2, Table 7.7). This indeed should raise some concern. We concluded that the performance indicators, illustrating the relation of high submission numbers with good and excellent grades, leads to a confidence, which we might not have intended. However, considering the most popular answer choice on the statement “Diagnostic information about my final performance would help me:”, which was answered by 68% of the students responding to the questionnaire with “to balance my workload better”, we found that students are in need of pragmatic decision support to juggle the demands of all their courses. Here again, in a wider perspective, the decision to cut back when feeling safe, might contribute to a lower stress level and better performance in other courses, when still operating on a high level in the course in question. In our case the lower submission rates in 2013 had no impact on students’ final mark when compared with 2011 (Section 7.6.2, Table 7.12).

In conclusion we could show that there are performance indicators to illustrate successful and risky study behaviour, which can be used by students to derive performance goals and monitor the actual performance regarding these goals. By keeping the performance indicators separated, areas of improvement become apparent. For example a mid-semester score indicating a final performance under the desired level indicates the need for revision. In the
Discussion we identified the potential to (1) identify additional predictive factors and (2) develop a predictive model based on the combination of predictive factors.

8.2.3 A scoping study supporting students’ self-regulated learning

After performance indicators have been identified, the question of How to effectively communicate course characteristics and performance indicators to support students’ self-regulated learning?, was addressed. For the scoping study, conducted in the second semester in 2013, an infographic was developed to inform students in an engaging way about the (1) course structure and key events, (2) former students’ perceptions about the course and the lab task difficulty, (3) important course characteristics and (4) performance indicators for different stages of the course. As described in Chapter 5 these design objectives resulted in four sections, which were loosely connected by a timeline layout. An interview study, involving six students, was conducted to evaluate to which degree the design objectives were met and students felt attracted and inclined to attend to the information. The results of the interview study informed the redesign of the infographic. Students’ very positive reactions and interest during the interviews was taken as encouragement towards the idea of the printed format and the kind of information to display. The redesign resulted in a compact graphical design, fitting all four aspects mentioned above on a single A3 page. We hope that this concrete example of packaging course related information inspires other educators to pursue similar ideas.

The redesign of the section with the performance indicators resulted in donut charts showing the actual distribution of students’ final examination result for one achievement group in 2011/12. On reflection it would be interesting to investigate how students actually interpreted these distributions. During the design phase we decided not to remove any outliers from these distributions, showing, for example for one achievement group, that a very small proportion of students achieved at B-grade level when over 75% of students failed the final exam. Even in this extreme example we need to ask, if students are likely to overestimate their chances to do reasonably well and how the information was actually interpreted. It is possible that an alternative graphical representation would be more effective in conveying the message we intended, but we did not compare different graphical styles regarding their effectiveness. To which degree students felt that the information was relevant for them personally is another interesting question. By clearly stating in the design that this information is past years’ data (consider section name “What numbers in 2011/12 said”), students might have felt a certain doubt to which degree these data are applicable to their situation. In general the questionnaire responses indicated a high acceptance of the infographic as 94% of the students referred back to the information voluntarily and 80% used the performance indicators later in the course. However to investigate the questions of stu-
8 Contributions and Future Work

students’ (1) actual interpretation of the information and (2) genuine acceptance in more detail, further studies would be necessary.

The printed format of the infographic ensured short development cycles, low production costs and easy delivery, but also bears shortcomings given the static nature of the format. If performance indicators were interactive, presented by a software application for a website, mobile phone or available on the lab computers, students’ actual course performance data could influence the presentation indicating achievement groups and projected final performance. Combined with a predictive model, more precise diagnoses would be possible, which reduces the room for interpretation and which might also improve the perceived personal relevance. Applying simulation and animation techniques could show in a compelling way how changes of study behaviour are likely to influence the final performance.

An interactive version of the infographic could provide (1) instant availability of up-to-date, personal course performance data, (2) a precise projection of future performance and (3) simulation. Whether this would increase students’ engagement with the information is speculative. We would need to know why 20% of the students responding to the questionnaire did not engage with the performance indicators. Given that the group of higher achieving students reported a significant lower number of predictors used, we can conclude that higher achieving students were not as much concerned about their performance than lower achieving students. In fact 11 of the 17 students, reporting that they never used the performance indicators, scored over 75% in the final exam. Targeting the engagement with feedback of high achieving students is pointless, as those students most likely do well regardless of any feedback intervention. However, there were also lower achieving students reporting that they never used the performance indicators, but we don’t know why. One participant (P1) in the interviews commented on the use of the infographic: “… after that first lab I just sort of forgot about this”. Here an increased presence of the information, for example when logging into the system, would remind students of the resources available and might encourage an higher prolonged engagement.

Students’ stages of engagement with feedback as proposed by Price et al. (2011) and matters of effective feedback influenced the way the infographic was rolled-out and evaluated (Chapter 6). By including the infographic in the lab book it was ensured that all students received the information. The initial engagement with the infographic was very high because students were required to solve a number of infographic related tasks to complete labs 1 to 3. As the submission of lab tasks in our course is always accompanied by personal dialog with a tutor, misunderstandings could be instantly addressed and corrective advice, on how to read this information, could be provided. This kind of personal, detailed and timely feedback was important to avoid later disengagement with the infographic because it was not understood. In fact a high prolonged engagement was achieved as almost all stu-
students (94%) responding to the questionnaire reported that they referred back to the infographic at least once. As the questionnaire responses were the only data set to assess prolonged attention, the conclusions need to be considered in the light of self-selection and common method biases, both suggesting limited generalizability and validity of self-reported data sets. For a fuller understanding of students’ actual interactions with the infographic, a counterbalance with potentially unbiased data sets is desirable. Here again, the consideration of some interactive software seems promising as those interactions can be recorded when they actually take place (ethics permitting).

During the discussion of the question of To what degree did students engage with the feedback provided? in Chapter 7 various other limitations of the way the data were collected and analysed were addressed, which might impact on the validity of the conclusions regarding students’ engagement and learning outcome. Firstly, a year-to-year comparison is influenced by many factors besides the intervention itself. That was illustrated by significant differences between the two years 2011 and 2012 forming the control group (e.g. for no explainable reason students in 2012 spend less time per submitted task in the labs than in 2011), which made the detection of subtle behavioural changes difficult. Variability within the control group was also problematic when discussing the final examination results as students performed significantly different in 2011 and 2012. This highlights, secondly, the need to reflect critically on the format of a final examination to assess students’ learning outcome as well as on the quality of examination questions. If students’ results in the MCQ part differ by more that 10% with a medium to large effect size, but there is no difference in the short answer part and their mid-semester results, the question should be raised to what degree paper-based written MCQs are a valid assessment of students’ programming ability and therefore a solid basis for assessing the effectiveness of an intervention. We addressed this problem by using nearly identical, embargoed sets of examination questions in 2011 and 2013, which was taking care of the variability caused by different questions but not of the problem that examinations might have a limited validity to assess students’ actual learning outcome. Here alternative measures, for example based on the quality of the lab-based programming tasks, would add to the discussion as to what degree the actual learning outcome improved.

In contrast to the tricky assessment of students’ subtle behavioural changes and learning outcome as discussed above, the conclusions regarding students’ attitudes towards diagnostic course data are rather clear-cut. Both interview studies resulted in our confidence that students want to gain insight into course characteristics and factors influencing their final performance, rather to depend on hope to do well. This was reiterated by the responses to the questionnaire, where 95% of the students ticked at least one scenario how they thought diagnostic course data would help their learning. These empirical results form an
important starting point into further research on how to foster students’ active role in the learning process by providing feedback at self-regulation level.

In the critical reflection on the scoping study, which provided (1) a concrete example of how naturally occurring course performance data can be turned into relevant feedback information and (2) empirical results of how such information affected students’ study behaviour as well as students’ general attitudes towards diagnostic course data, we isolated several areas for revision and improvement. Firstly, further investigation of students’ actual interpretation and genuine acceptance of the information could lead to a more effective form of graphical representation. Secondly, the printed format was questioned and the implementation of an interactive version was considered to eliminate issues of misinterpretation and disengagement as well as to provide rich detail about the actual use of the information as opposed to self-reported data. Thirdly, we found the expressiveness of a study based on comparing data on a year-to-year basis limiting given the generally high variability. Concern was also raised to what degree examination results are valid reflections of students’ learning outcome. However, one of the main contributions and a guide for future work was the finding from the scoping study that our students valued the information provided. These are encouraging first results, which should lead to further advances towards the provision of diagnostic course data.

8.3 Future Work

As a next step, a follow-up project could build on the experiences gathered in the scoping study and address the limitations discussed in the previous section (8.2). The following aspects might assist the planning and conducting of such a research project:

- Considering feedback, namely (1) the developmental aspect of self-regulation and (2) the availability of effective feedback on all three levels to be influential on students’ ability to act on the feedback provided.

- Considering the information provided, by the (1) identification of additional predictive factors and (2) the development of a predictive model based on the combination of predictive factors.

- Considering the form of delivery and evaluation, regarding (1) further investigation of students’ actual interpretation and genuine acceptance of the information leading to a more effective form of graphical representation, (2) the implementation of an interactive version to eliminate issues of misinterpretation and disengagement as well as to provide rich detail about actual use of the information as opposed to self-
reported data and (3) the improvement of the evaluation methodology and the assessment of learning outcomes.

- Considering the provision of diagnostic course data as a promising feedback practice towards more transparency for students.

In the following sections, these opportunities for future work are taken into account when discussing a follow-up research project to target the role of feedback in the context of students’ self-regulated learning.

### 8.3.1 Assisting the development into self-regulated learners

As a first step in preparation of such a project the theoretical basis needs to be reassessed. In particular the developmental characteristic of students’ self-regulation should be taken into account. Zimmerman and Schunk (2001) provide a variety of theoretical perspectives that are guiding the research on how students become self-regulated learners. One of the five questions leading the discourse is “How does a learner acquire the capacity to self-regulate when learning?” (p. 7). Consensus is not always achieved and controversies between the theoretical standpoints remain, for example “whether external reinforcers are essential to intervention programs” (p. 287). That students hold different conceptions of feedback was already discussed in section 2.2.6. Based on these findings McLean, Bond and Nicholson (2014) conclude that feedback practices need to be aligned with students’ conceptions of feedback:

*Students understand the process of feedback in different ways; therefore educators need to take account of these differences, if they want better engagement with feedback. For example, students who experience Feedback as telling expect to be told. If they are provided with feedback in some other form, e.g. questioning or discussion, they may perceive that no feedback has been given.*

(p. 9, original emphasis)

If we learned more about where our students are in their development as self-regulated learners and their conceptions of feedback, we could better tailor the feedback information and assist these developmental aspects.

Most likely an exploration of students’ expectations and needs of feedback, as discussed above, would question current feedback practices on all levels. If students want to take action, what are the limiting aspects? How should we support students who struggle in their quest to master the first programming language? And specifically, in regards to the findings of the scoping study, how can students’ realisation of the risk of failing be turned into fruitful action? Here the wealth of literature into the “novice programmer” (e.g. Soloway
and Spohrer, 1989; Fincher and Petre, 2004; Simon, Lister and Fincher, 2006; Robins, Haden and Garner, 2006) would inform in a broader context about problems students are facing and might extend the discourse on what kind of feedback on task and process level is particularly important to first year programmers. We already initiated this discussion in Chapter 3 by identifying opportunities to improve feedback in our course with entry points into the literature.

8.3.2 Improving diversity and personal relevance of feedback

As a second step in preparing a follow-up research project, the search for additional predictive factors could be undertaken with the aim to illustrate successful or unsuccessful learning strategies. As one example, students’ compile behaviour was studied by Tabanao, Rodrigo and Jadud (2011) and showed that patterns of frequent compilation and certain error types indicated lower performing students. For example we know that “trial and error” when working on a programming task is often not a successful strategy (e.g. Edwards, 2004). Perkins (1989) observed: “Extreme movers, however, move too fast, trying to repair code in ways that, with a moment of reflection, clearly will not work” (p. 266). Illustrating a possible negative correlation of compile frequency with final exam results or submission numbers could put emphasis on a methodical and reflective approach to programming. Patterns might be also revealed by looking into students’ actual hours on task as opposed to the use of not task-related programs and internet sites or the frequency of swapping task and not task-related context during the lab session. If those patterns can be connected with success or failure they might be turned into valuable feedback information.

The combination of predictive factors into a diagnostic model is a tempting endeavour. Predicting students at risk of failing the course is one aspect. In the light of the Learning Edge Momentum (Robins, 2010) such identification is desirable as soon as possible to intervene and stop the development of a negative learning momentum. Another aspect is to support students in self-monitoring their performance as attempted in this research. If precise predictions, based on students’ performance data, are possible, students might engage with such information on a higher level. Examples for attempting diagnostic models using various data sources to predict retention (Sarker, Tiropanis and Davis, 2013), final grades (Kotsiantis, 2011) and students at risk of failure (Macfadyen and Dawson, 2010) by employing methods of learning analytics or data mining can be found, but are rarely discussed in the context of feedback. Arnold and Pistilli (2012) report on the use of a system called “Course Signals” to inform educators as well as students about students’ “risk statuses” in form of a traffic light, which lead to significant higher retention rates for the experimental group.
If personal performance predictions are shared, which leave limited room for interpretation, research is necessary to enquire into students’ reactions: What are the effects on students’ self-efficacy and motivation? What kind of advice and guidance is needed if students want to improve their prospects? What degree of scaffolding is necessary to avoid discouragement on one hand and overconfidence on the other hand?

### 8.3.3 Improving student engagement with feedback through interactivity

In the development phase of a follow-up project, the implementation of an interactive version to eliminate issues of misinterpretation and disengagement should be considered. Beside the advantage of the instant availability of students’ personal course performance data and a projection of future performance derived from those data, such a software-based system could provide rich detail about the actual use of the information as opposed to self-reported data. Here an opportunity to learn more about students’ interactions with the feedback provided and to conclude on improvements of current feedback practices emerges.

However, an assessment of the requirements regarding content and functionality of an interactive system is a prerequisite before the development. For example Cooper et al. (2007) advocate a goal-directed interaction design, which put the user’s goals in the centre of the design process. Design research (e.g. interviews or observation), involving potential users, is the first step. To model different user groups and their specific requirements, as found during the design research, the authors recommend the use of “personas”, archetypes with clearly defined behaviour and goals. Later in the design process interface variants are discussed with the personas’ needs in mind. Frequent usability testing of the software prototypes is a vital part of the interaction design process and should involve a group of typical users. According to Krug (2000) and Nielsen (n.d.) a small number of participants (3 to 5) can be sufficient to encounter most major usability issues. They also point out the need to keep testing after fixing the problems in an iterative design process.

Applying such user-centred interaction design methods would reveal students’ aims and needs regarding an interactive feedback system and clarify what type of information to include and what form of delivery is most desirable.

### 8.3.4 Improving evaluation methodology by the use of mixed methods

The assessment of the effectiveness of such an interactive feedback system remains problematic. As matters of engagement and usage can be evaluated based on students’ interactions with the system and extend the existing quantitative data sets, we still learn little about students’ actual interpretations and beliefs about their progress in the course. For
example frequent dialog with students involved in the study may result in insights, which cannot be gained from the quantitative data only.

The use of mixed research methods, a methodical combination of qualitative and quantitative research, is suggested to provide a better understanding of the research problem than by using quantitative or qualitative data solely (Creswell, 2012). Even though we used interviews at the end of the course, which would be considered as a form of embedded design (Creswell, 2012, p. 541), as we collected the qualitative data as part of a quantitative design, other types of mixed methods designs would allocate more weight to the qualitative data. For example in convergent parallel design qualitative and quantitative data are collected simultaneously to investigate the same aspect (Creswell, 2012, p. 540). Data sets are analysed separately and the results are compared for an interpretation.

Awareness that CSEd research needs to integrate methods, already developed in social and behavioural sciences, to inquire into teaching and learning and that knowledge of the subject alone is insufficient to tackle questions in the educational context, was raised by Tenenberg and McCartney (2010). In Fincher, Tenenberg and Robins (2011) this discussion is extended and the authors suggest, “that there are cases were we need to craft our own methods, or reshape tools to better suit the phenomena that we wish to investigate” (p. 27). To study novice programming students regarding their cognitive strategies and their identities as programmers, in two examples the researchers describe the tools they applied for an in-depth investigation of students’ activities, locations for those activities as well as students’ stories and emotions around those activities. Different artefacts produced by students such as photographs, diary grids (records of hourly activities) or emotional seismographs (curves to describe positive or negative feelings over time) are used to aid recollection of events and feelings in semi-structured interviews with the students. The depth of information collected from these interviews supports the authors’ argument to explore novel qualitative approaches.

For example, aspects of students’ engagement could be highlighted if students are interviewed or observed at the time of expected usage of the feedback (e.g. after getting the mid-semester results). This way the central role of examination results as one major unit of analysis to assess the effectiveness of the feedback, as used in this research, would be diminished in favour of closer investigation of other aspects apart from learning outcomes such as students’ conceptions of and satisfaction with the feedback.
The question to which degree common forms of assessment in introductory programming courses, and MCQs in particular, can be used to assess programming ability is discussed in the CSEd literature (e.g. Lister, 2001; Shuhidan, Hamilton and D’Souza, 2010) and concerns about the validity and reliability of assessment instruments in introductory programming courses are raised (e.g. Lister, 2010; Petersen, Craig and Zingaro, 2011; Simon et al., 2012). Consequently, using examination results to address research questions might lead to false conclusions. How to evaluate the quality of MCQs is part of the investigation of student answers of 12 MCQs to assess students’ code reading and writing skills in Lister et al. (2004). Those discussions and recent advances into the provision of a public repository for exam questions for CS1 and CS2 students (Sanders et al., 2013) might lead to more comprehensive examinations in the future, which would also increase their validity and reliability as research instruments.

8.3.5 Opening the black box – our responsibility for more transparency

Despite the difficulty of assessing the effectiveness of feedback, the question what we consider as effective feedback interventions remains. Is a feedback intervention, which does not result in a significantly improved learning outcome per se “ineffective”? What value is attached to an outcome that students feel more in control of their learning because course characteristics and influential factors are made transparent up-front? How to weigh the support of developmental aspects, which cannot be measured by examinations or submission rates?

The Otago University Teaching and Learning Plan (University of Otago, 2010) states as “Objective 3: To develop the lifelong learning skills of students.” which is accompanied by teaching guides such as “Encouraging students to take increasing responsibility for their learning as they progress through the University.” and “Encouraging students to take responsibility for their decisions.” (p. 7). Here the importance of students’ development into self-regulated learners is manifested and should lead to further advances to provide feedback which enables students to “take responsibility for their decisions” by opening up the black box of course characteristics and performance indicators. From this perspective, effective feedback cannot be solely determined by students’ learning outcome, but moreover to what degree it supports students’ informed decisions regarding their learning. This shift in perspective needs to be incorporated when planning, conducting and evaluating future feedback interventions.
8.4 Closing remarks

A PhD project can be an isolating experience as the author is often the sole investigator in the project. Apart from this emotional drawback there is also no compensation for the bias this investigator brings into the study. Working in a team would have had several advantages besides a more objective approach to the research project. In my role as a tutor in COMP150 and COMP160 I knew most students, who were involved in the study, personally. Students might have answered differently during the interviews and when filling-in the questionnaire if asked by an independent investigator. It cannot be denied that the research presented in this thesis is highly influenced by the particularity of the group of students involved, their attitudes towards me as an investigator as well as by my own beliefs, perspectives and interpretations. Given these limiting aspects towards the generalizability of the study, future work plays an important role to situate similar feedback interventions into a wider context.

As a very practical contribution, the infographic will be updated and produced as a poster for the computer lab in 2014. This is necessary because, due to the department’s efforts towards paperless course, the lab book, which included the infographic, will be only printed on demand.

Personally I enjoyed the research journey as well as the teaching part. At the moment I am looking forward to another semester of teaching in the labs and I hope to be part of the CSEd community in the long term.
References, Appendices
References


References


Appendix A – Infographic Example I (Charles Joseph Minard)

Charles Joseph Minard (27 March 1781 – 24 October 1870 in Bordeaux)
Florence Nightingale (12 May 1820 – 13 August 1910)
Appendix C – Infographic (initial version)
Appendix D – Infographic (final version)
Section 1: Key Events & Advice

Section Key Events provides some information about what is happening in COMP160, parts that count towards your final mark and the concepts introduced each week. In section What we say you’ll find bits and pieces of advice.

1) How is your final mark composed? Please fill in the numbers:
   Lab Tasks ___%  Mid-Semester Exam ___%  Final Exam ___%  = 100% Final Mark

2) How do you get each 1% internal mark towards your final grade? You need to...
   A) attend the scheduled lab session  C) submit a finished lab task on time
   B) submit a finished lab task at any time  D) submit a partially finished task

3) Why do students fail the course even though their final mark is 50 (C-) or more? Because they...
   A) did not attend all lab sessions  C) scored under 50% in their final examination
   B) submitted less than 21 lab tasks  D) were late with half of the lab tasks

Section 2: Task Difficulty and Students’ Comments

In section What students say you’ll find a selection of students’ comments from 2011 & 2012 and a graph to visualise students’ ratings of the lab task difficulty.

1) If you consider the quotes, do you think students experienced the lab tasks in a similar way?
   A) yes  B) no  C) not sure

2) What did students rate to be the second hardest lab task? Lab number

3) What lab number is set for the first half of week 7? Lab number

4) Why do you think it is important to do this lab on time? Because it...
   A) is more difficult than the previous lab  C) needs to be submitted for terms requirements
   B) is a catch-up session  D) it is a good preparation for the mid-semester examination

Section 3: Performance Indicators from 2011 and 2012

By closely looking at data from 2011 & 2012 we realised that there are some indicators for how students would perform in their final examination in COMP160. For example the COMP150 grade was moderately (in 2011) to strongly (in 2012) correlated with students exam results in COMP160. Another strong correlation was seen by the total number of lab tasks handed in. The strongest indicator, however, was the mid-semester examination result.

The charts in section What the numbers in 2011/12 said look quite complicated and there is a lot of information here. For five separate indicators we see:

- How many students were in each group for an indicator.
- The average score of the group’s COMP160 final exam in the centre of the circle.
- The group’s actual distribution for the final exam results in terms of A-level (dark green), B-level (green), C-level (orange) and falling students (red).

Please have a look at the Mid-Semester exam result indicator.

1) For the group of students achieving 9 or 10 points out of 15 as their mid-sem-semester exam result, what was their average final exam score? ____________________________ %

2) How many students were in that group? ____________________________

3) Can you estimate how many from this group could actually pass the final exam? Approximately ____________________________ %

Let’s say your goal is to complete COMP160 with a B-grade. Preferably you need to score above 65% in your final exam. (For an A-grade it would be over 80% and for a C-grade over 50%).

4) Based on the data from 2011 & 2012, what do you need to achieve to be confident to actually reach that goal?
   Number of labs by week 5 ____________________________
   Number of labs by week 11 ____________________________
   Mid-semester result out of 15 ____________________________
   Average lateness per submitted lab task by week 8 ____________________________
Dear Student

The infographic at the back of the COMP160 lab book was something new this year. Therefore we would like to know what you think about this kind of information.

Section “Key Events”

This section (“Key events”) made it easy to understand the structure of the course.

Strongly agree □ □ □ □ □ Strongly disagree

I looked up this section …
□ Often □ Occasionally □ Seldom □ Never (after Lab 3).

Section “What students say”

This section (“What students say”) clearly showed that the tasks’ difficulty increases steadily over the first 5 weeks.

Strongly agree □ □ □ □ □ Strongly disagree

Seeing other students’ difficulty ratings and comments was a good preparation on what to expect in this course.

Strongly agree □ □ □ □ □ Strongly disagree

I wish other courses would provide similar information.

Strongly agree □ □ □ □ □ Strongly disagree

I looked up this section …
□ Often □ Occasionally □ Seldom □ Never (after Lab 3).

Section “What we say”

The bits and pieces of advice in this section (“What we say”) were easy to understand.

Strongly agree □ □ □ □ □ Strongly disagree

Most of the information in this section was relevant to me.

Strongly agree □ □ □ □ □ Strongly disagree

I looked up this section …
□ Often □ Occasionally □ Seldom □ Never (after Lab 3).

Section “What numbers in 2011/12 said”

Section (“What numbers …”) clearly showed how performance before and during the course correlates with the final exam result.

Strongly agree □ □ □ □ □ Strongly disagree

It was clear from this section that working consistently pays off at the end.

Strongly agree □ □ □ □ □ Strongly disagree

This kind of information was useful.

Strongly agree □ □ □ □ □ Strongly disagree

I looked at the “donut charts” for my own results:
(please tick all that apply)
□ COMP150 final mark □ Lateness by week 8
□ Number of Labs by week 5 □ Num. of Labs by week 11.
□ Mid-semester exam score □ I never used the charts.

Looking up my own results provided me with a good sense on how I’m doing in the course.

Strongly agree □ □ □ □ □ Strongly disagree

Having this kind of information made me feel
□ More confident □ Less confident □ No change □ Can’t say.

□ Having this kind of information made me work
□ Harder □ As normal □ Less. □ Can’t say.

General

I was surprised that there is information indicating students’ later success or failure in a course.

Strongly agree □ □ □ □ □ Strongly disagree

Such diagnostic information should be shared with students.

Strongly agree □ □ □ □ □ Strongly disagree

Diagnostic information about my final performance would help me:
(please tick all that apply)
□ to balance my workload better □ to take more responsibility
□ to take action when necessary □ to feel more in control of my learning.
□ to reach my goals for a course □ other
□ to stay focussed □ I don’t think it would help.

Thank you!

Additional Comments? Any Suggestions? Please turn over.
### Appendix F – Interview Protocol

<table>
<thead>
<tr>
<th>1 Briefing (5 min)</th>
<th>2 Warm up (7 min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Thanks for coming</td>
<td>- What is the course like? Is it fun? Is it lots of work? How is it compared to other courses you did (COMP150 if applicable)</td>
</tr>
<tr>
<td>- Introduce me and my topic (Investigation of different forms of feedback in CS1)</td>
<td>- Are the lectures interesting? Is it important to go to the lectures?</td>
</tr>
<tr>
<td>- Aim of the interview and agenda: Four parts …</td>
<td>- What is the lab work like? Is it easy to understand what the tasks are about? Do you need to spend a lot of time? Did you enjoy the programming part? Or is it hard?</td>
</tr>
<tr>
<td>- Reading information sheet and signing the consent form</td>
<td>- Have the demonstrators been helpful/approachable/nice? Did they have enough time to help? When they explained something – did you understand what they were talking about? Was the marking fair?</td>
</tr>
<tr>
<td></td>
<td>- What do you think helps best to get through the tasks: friend, online resources, lecture notes, textbook or demonstrator's help?</td>
</tr>
<tr>
<td></td>
<td>- Do you think you learned a lot about programming?</td>
</tr>
<tr>
<td></td>
<td>- What was the most annoying/frustrating/hardest part? (How to support this better?)</td>
</tr>
<tr>
<td>3 Feedback Intro (3 min)</td>
<td>3a Feedback regarding lab tasks (10 min)</td>
</tr>
<tr>
<td></td>
<td>- Did you understand the task requirements, goals and marking criteria?</td>
</tr>
<tr>
<td></td>
<td>- As you worked on the task did you know how you went in relation to these requirements?</td>
</tr>
<tr>
<td></td>
<td>- Was the feedback demonstrators provided helpful that you got an impression how you went regarding task requirements and task completion?</td>
</tr>
<tr>
<td></td>
<td>- When you have been stuck - was the help you got useful to know what to do next and how to work towards the task completion?</td>
</tr>
<tr>
<td></td>
<td>- Have you ever not been signed off when you actually thought you would be finished? Was the feedback you got helpful guidance/advice to understand what was wrong and to finish the task?</td>
</tr>
<tr>
<td></td>
<td>- How to improve feedback regarding the lab tasks?</td>
</tr>
<tr>
<td>3b Feedback regarding the programming process (10 min)</td>
<td>3b Feedback regarding the programming process (10 min)</td>
</tr>
<tr>
<td></td>
<td>- Which of those have been included in the task requirements and marking criteria?</td>
</tr>
<tr>
<td></td>
<td>- Which were not covered in terms of requirements for task completion?</td>
</tr>
<tr>
<td></td>
<td>- Do you think you got feedback on those aspects while working or task submission?</td>
</tr>
<tr>
<td></td>
<td>- What feedback was provided to help you to improve on those aspects?</td>
</tr>
<tr>
<td></td>
<td>- What need to get more attention in the future? What do you think would be helpful for future programming projects?</td>
</tr>
<tr>
<td></td>
<td>- How to improve feedback on this level in general?</td>
</tr>
</tbody>
</table>
3c Feedback regarding course management (10 min)

(1) Do you think it was clear form the beginning:
   - When **happens** what, internal marks and final grade
   - Tasks get **difficult** quite quickly
   - **Frustration** is part of the process
   - Working **consistently** is important/falling behind can be dangerous
   - How submission of lab work correlates with the final exam results
   - That there are **indicators** to help you to understand how you are going

(2) How good have these aspect been reflected by the infographic? What did you find **initially** most interesting/useful to get an idea about the course?

(3) In **retrospect** was you first impression about the course (based on the information given) been correct? What did surprise you?

(4) Do you think that the information given **influenced** how you approached/studied in the course? If so …. how?

(5) During the course did you use the **data section** to see how you are going?
   - **What/when** did you use it? What did this information tell you about your performance?
   - Did you think the information was **applying** to you?

(6) In terms of **improvement** … would/did you know what needs improvement?
   - Would/did you think you can still improve? If so … what makes it hard to do so (practically/theoretically)?

(7) Three possible reaction on (example from infographic):
   1) This applies not to me
   2) I will be fine. (I am here.)
   3) I am going to fail.
      3a) I need to work harder.
      3b) I better give up.

Who are you? What do you think other students are like?

(8) How to improve this kind of feedback?

3 Debriefing (5 min)  

Any remarks, suggestions, recommendations?

Thanks and goodbye.

Notes:
# Appendix G – Mid-Semester Feedback Sheet

## Summary of Mid-semester Results

**ID:**

<table>
<thead>
<tr>
<th>Short Answer page</th>
<th>Score (out of)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>8.5 (out of 9)</td>
</tr>
<tr>
<td>8</td>
<td>5.5 (out of 6)</td>
</tr>
<tr>
<td>9</td>
<td>4.5 (out of 6)</td>
</tr>
<tr>
<td>10</td>
<td>2.5 (out of 6)</td>
</tr>
<tr>
<td>11</td>
<td>4 (out of 6)</td>
</tr>
<tr>
<td>12</td>
<td>3 (out of 10)</td>
</tr>
</tbody>
</table>

**Multiple Choice Questions:**

| Total Mark (out of 63) | 43 |
| Internal Mark (out of 15) | **10.24** |

Your score indicates some problems with the concepts introduced so far. Please revise the indicated parts of the exam.

**Lab 17 Option - Corrections for the mid-semester exam:**

1) Collect a revision sheet and work through the parts as indicated above.
2) Discuss your solutions with a demonstrator.
3) Get the 1% lab mark for Lab 17.