Developing the Impact! Tool for Higher Education

Master Thesis

Institution: University of Twente

Faculty: Behavioral Sciences

Master: Educational Sciences and Technology

Author: Shanice Harmon

Supervisors: Dr. A.J. Visscher and Hannah Bijlsma

Date: August 20, 2018

UNIVERSITY OF TWENTE.

Abstract

Student Evaluations of Teaching' (SET's) are instruments are used to collect student perceptions of teacher quality in contexts of higher education (Marsh & Dunkin, 1992). Nevertheless, SET's have been suggested to have three limitations: they are rarely used for formative evaluation (Stieger & Burger, 2010); they are mostly presented in paper format, rather than digital format (Stieger & Burger, 2010); and the items in the instrument are often critiqued in terms of validity and reliability (Coles, 2002; Greenwald & Gilmore, 1997; Rowley, 1996; Rowley, 2003; Wachtel, 1998; Worthington, 2002). In response to this, researchers from the University of Twente aimed to develop a SET called the Impact!l, which considers these mentioned limitations in its design.

To support the development of Impact!, this study sought to answer the following research question: *To design a valid and reliable instrument for measuring student perceptions of teaching quality after a lecture in higher education by means of the Impact! tool, which items should be included in the Impact! questionnaire?* This study is structured in three stages and uses Pride's (2010) model for questionnaire design. In stage 1, focus groups were facilitated for experts to evaluate the items in first draft of the questionnaire. In stage 2, cognitive interviews were conducted to determine whether the target audiences comprehended the items in the second draft of the questionnaire. In this stage 3, the third draft (pilot version) of the questionnaire was added into the Impact! tool, and tested in lectures at institutions of higher education. It was determined that the third version of the questionnaire is loaded on one single factor, which is predicted to be the underlying construct of 'teaching quality'. The third version of the questionnaire was also measured to have an acceptable reliability of .75.

Table of Contents

Introduction

Problem Statement

Institutions of higher education seek to improve the quality of teaching in lectures on an international level (Centra, 2003; Dochy, Segers, & Sluijsmans, 1999). Correspondingly, there has been an increasing demand to evaluate the quality of teaching in specifically higher education (Centra, 2003; Dochy, Segers, & Sluijsmans, 1999). Feldman (2007) states that, "many colleges and universities are currently putting emphasis on good teaching, and on designating, honoring, and rewarding good teachers" (p.12). Evaluating the quality of teaching is important, because it has the potential to improve teacher performance for better student learning and achievement (Rowley, 2003). One possible method to evaluate the quality of teaching is by means of Student Evaluations of Teaching (SET's; Marsh & Dunkin, 1992). SET's are student feedback instruments that have the potential to collect student perceptions of teaching quality in higher education (Marsh & Dunkin, 1992).

Although SET's have already been in use for several decades, this research-project has acknowledged them to have three main limitations. The first limitation is that SET's have been mostly used for summative evaluations at the end of a course, rather than for formative evaluations after one single lecture (Stieger & Burger, 2010). Narasimhan (2001) discovered that collecting student perceptions of the quality of teaching directly after a lecture, helps teachers adapt to the students' needs in the following-up lecture. Direct feedback makes it possible for teachers to improve their lesson plans (Narasimhan, 2001). The second limitation is that the majority of existing SET's have been in paper format, rather than digital format (Stieger & Burger, 2010). Korfhage (2008) recognized technological application to be practically beneficial for data collection; he states that digital systems make it possible to obtain, store, and retrieve data from respondents in a systematic manner. Means and Olson (1997) also argue that digital systems are advantageous, because they give students and teachers the ability to communicate district wide, nationally, and internationally. Richardson (2005) suggested to that its best to obtain these student perceptions of teaching quality through questionnaires, because they are efficient instruments for data collection. The third limitation is that content in the SET's has been continuously critiqued in terms of validity and reliability. Scholars have identified confounding variables that might influence student ratings when responding to items (Coles, 2002; Greenwald & Gilmore, 1997; Rowley, 1996; Rowley, 2003; Wachtel, 1998; Worthington, 2002); this limitation is further discussed in the 'Theoretical Framework' section of this report.

Research Aim

This research project aimed to develop a new set of items for an SET that is called Impact!. The development of the items for the Impact! tool had the purpose to support the increasing demand to evaluate the quality of teaching in higher education, as well as help overcome the limitations of existing SET's. These items sought to measure student perceptions of the quality of teaching, directly after a lecture in higher education by means of a digital application. And for the items in the SET to be reliable and valid, there were quantitative and qualitative evaluations of the instrument.

Research Questions

Main Research Question:

To design a valid and reliable instrument for measuring student perceptions of teaching quality after a lecture in higher education by means of the Impact! tool, which items should be included in the Impact! questionnaire?

Supporting Research Questions:

These additional research questions help answer the main research question:

1. Literature review

    A. Which constructs measure teaching quality in higher education?

    B. What are appropriate items to be in the questionnaire for the Impact! tool in higher education?

2. Focus groups and cognitive interviews

    C. Which items are relevant in the questionnaire for the Impact! tool in higher education?

    D. Which items should be reformulated in the questionnaire for the Impact! tool in higher education?

3. Pilot testing

    E. What is the reliability of the items in the questionnaire for the Impact! tool in higher education?

    F. What is the construct validity of the items in the questionnaire for the Impact! tool in higher education?

Theoretical Framework

The following section consists of concepts, definitions and references to relevant scholarly literature and existing theory that is relevant for the research-project. The theoretical framework helps

explain the meaning, nature, and challenges related to the main topic of interest (ie. SET's) (Gabriel, 2008).

Student Evaluations of Teaching (SET's)

As stated in the introduction, the research-project sought to develop items for a SET. SET's are defined as instruments that collect student perceptions of their teacher's teaching quality in contexts of education (Marsh & Dunkin, 1992). They can help teachers understand in what areas of instruction they can improve, so that they can then adapt their lessons to student's needs. A SET has been explained to, "take the form of students' ratings of their level of satisfaction or their self-reports of other attitudes towards their teachers or their course units" (Richard, 2005, p. 388). The degree of satisfaction and attitudes indicate how a student assessed a teacher's performance with his or her personal expectations on a positive to negative scale. And although SET's often have one or more open-ended items that allow students to communicate their perceptions toward their instructors' teaching style and efficacy, these instruments typically contain rating scales that feature Likert-type items (Onwuegbuzie & Leech, 2006). Data from SET's can be collected mainly through questionnaires, in either digital or non-digital form (Marsh & Dunkin, 1992). Supportively, it is most commonly the task of the teacher to ask students to complete the SET's (Marsh & Dunkin, 1992).

Marsh and Dunkin (1992) conducted research to identify the main purposes for student feedback in educational contexts. Although four purposes are present in educational contexts, the first purpose was identified to be the most common:

1. Diagnostic insight to teachers about the effectiveness of their teaching
2. A measure of teaching effectiveness to be used in administrative decision making.
3. Information for students to use in the selection of course units and teachers
4. An outcome or process description for use in research on teaching.

Supportively, there is substantial evidence that SETs are useful to improve teaching quality (Centra, 1993; Marsh & Dunkin, 1992; Marsh & Roche, 1994). In the majority of these studies, teachers have been randomly assigned to control groups, in which student perceptions of teaching quality is collected with SETs. Afterwards, the control groups are compared on subsequent SET results. Moreover, Cohen's (1980) meta-analysis found that teachers who used formative SETs were rated .30 standard deviations higher at the end of their course than those teachers who did not use formative SETs. Marsh (1979) also discovered that feedback with consultation led to better examination performance, affective outcomes, and higher SET scores. Nevertheless, there is not enough evidence about the type of consultative feedback.

Taking the previous research into account, it can be assumed that if SETs support teaching quality, there will also be an improvement in student learning outcomes. Theal and Franklin (2001) have found positive correlations between SET scores and amount learned. However, Berk (2005, p 56.) explains that "key characteristics of students, such as ability, attitude, motivation, age, gender, and maturation, and of the institution, such as class size, classroom facilities, available technology and learning resources, and school climate, can affect student performance irrespective of what an instructor does in the classroom." Thus, it is suggested that the only way to determine whether a SET is improving student learning outcomes is to test students in the particular context (Berk, 2005).

Summative and Formative SET's

Most educational contexts depend on student feedback from summative evaluations, which are conducted at the end of a course (Stieger & Burger, 2010). These types of evaluation require students to recall their teachers behaviour and experiences from the course in its entirety. It is however questioned whether students' summative evaluations reliably predict teaching quality in a particular course. Studies have identified evidence that the evaluations are impacted by 'order effects' (Feldman, 1977). Some studies concluded that the last lectures of the course have a greater impact on the summative evaluation; this is related to the 'recency effect' (Steiner, 1989; Dickey & Pearson, 2005). Other studies reported the opposite by suggesting that the summative evaluation is influenced by people's first impression at the first lectures of the course; this is explained to be the 'primacy effect' (Feldman, 1977). An alternative form of evaluation that might prevent 'order effects' from biasing the student feedback results is formative evaluation (Stieger & Burger, 2010). It continuously takes place throughout the entirety of the course and allows students to judge their lecture based on their current memory.

The reason why the formative SET are rarely applied in higher education is most likely due to time constraints and other procedures, such as numerous evaluations, overwhelming data, and high administrative investment (Stieger & Burger, 2010). In order to overcome these disadvantages, it is recommended that the instrument is simple to use for both students and teachers. Stieger & Burger (2010) suggested that the following factors are incorporated into a formative evaluation tool: "(a) the procedure should be quick and easy for teachers and students, (b) the number of questions asked should be kept to a minimum, and (c) aggregation of data should be easy to administer in order to get the results quickly. It is evident that the effectiveness of a formative evaluation tool depends on the methodology of collecting data" (p. 164).

A Further Need for SET Research

Overall, many remain critical towards the quality of SET's. Scholars have found confounding variables that have been identified to influence student ratings of teaching quality (Coles, 2002; Greenwald & Gilmore, 1997; Rowley, 1996; Rowley, 2003; Wachtel, 1998; Worthington, 2002). Characteristics related to course, instructor, student, time of feedback, and the construction of instruments can have an impact on students perceptions of teaching quality. For course characteristics, students perceptions are affected by selectivity, level of course, subject area, and workload (Coles, 2002). Coles (2002) identified that students perceptions of teaching quality are scored lower when class sizes are larger, in earlier phases of their studies and when students courses are obligatory rather than optional. It has been proven that the instructors characteristics of rank, experience, reputation, research skills, class gender minority, class status and the instructors physical appearance can all be influential in student feedback (Rowley, 2003). For student characteristics, the expected grade, ethnic background, gender and age has also been found to influence results of student perceptions of teaching quality (Worthington, 2002). Additionally, the time-period of when the student perceptions are obtained has been identified as an influencing factor of validity (Narasimhan, 2001; Greenwald & Gilmore, 1997). Narasimhan, (2001) suggested that collecting students perceptions of teaching quality at the end of a course unit is not beneficial to their learning experience; instead, earlier feedback would be of more immediate value, in order for the instructor to improve their lecture. Greenwald and Gilmore (1997) found that students' perceptions in the middle of a course unit impacted their academic achievements.

Impact!: A SET that Evaluates Formatively and Digitally

Considering that SET's can lead to better evaluations when they are formative and in a digitally format (Stieger & Burger, 2010), the University of Twente has developed the Impact! tool for this purpose in secondary education. The Impact! tool is an SET that simple and quick to use, handles data electronically, and keeps administrative efforts small. Moreover, it has the capability to insert items that allow students to give teachers feedback for formative evaluation.

The developers of the Impact! tool had already developed items to formatively evaluate teaching quality in secondary education (Bijlsma, Visscher, Veldkamp, & Dobbelaer; under review). The instrument incorporated the following constructs into its items: a supportive and positive classroom, organized and structured classroom management, providing clear instruction, quality of the teacher-student interaction, cognitive activation for deep learning, and assessment of students' learning.

However, the developers of the Impact! tool had not yet developed items to formatively evaluate teaching quality in higher education; these are expected to differ from those in secondary education. The reason for the expected difference is the variation in structure of the two contexts. In secondary education, class sizes are generally smaller, student participation is less limited, the speed of lectures are often

adapted to the students, students have a lower academic workload, students are told what to do in most situations with instruction, teachers engage in instructional supervisory practice, students receive incentive to participated from others, and teachers have more classroom management responsibilities (e.g. preparing classroom assignments) (Boyer, 1983; University of Wisconsin-Platteville, 2018). In higher education, class sizes are generally larger, student participation is limited, the speed of lectures are fast-paces and generally more rigorous, students have a higher workload, teachers do not play a supervisory role, students require self-discipline, students have to take control of their own organization and time management, students supply their own motivation, and teacher have less classroom management responsibilities (McKeachie, 1987; University of Wisconsin-Platteville, 2018).

Because of the different teaching practices in educational contexts, the items for the Impact! tool in secondary and higher education are expected to not be interchangeable. Supportively, this research-project aimed to develop items that were appropriate to formatively evaluate teaching quality in secondary education, which were then implemented into the Impact! tool.

Method

Model for Questionnaire Design

This research project was structured in three stages, which were guided by six questions to answer the main research question. The research project had a mixed methods approach, because it included quantitative and qualitative research (Collins, 2003). The development of the SET was supported by a literature review, focus groups, cognitive interviews, and pilot testing. Samples for this research included experts and target group respondents.

The methodology of the research followed the structure of a five-step model for questionnaire design from Pride (2010). This model was chosen, because it supports the aim of this research-project: to design items for an evaluation instrument that are valid and reliable (which can be determined by construct validity, face validity, content validity, and reliability) (Pride, 2010). The first step in the model is to determine the background of the research, which consists of the target audience, purpose, goal, and objective; this was formally discussed with the developers of the Impact! tool. The second step in the model is to conceptualize the preliminary items with relevant constructs suggested by previous researchers specialized in SET's. This was done in the form of a literature review that compared and contrasted different constructs that related to effective teaching practices. The third step in the model is to select an appropriate scale of measurement for the questionnaire format, which was identified after developing a draft of item for the Impact! tool. The fourth step in the model is to establish the face and content validity of the questionnaire. Face validity is a measure of how representative an instrument is 'at face value,' and content validity is the estimate of how much a measure represents every single element of

a construct (Shuttleworth, 2009). These forms of validity were identified when focus group respondents were interviewed about the items in the instrument and the target audience took apart in readability tests. The fifth step in the model is to measure the construct validity and reliability by analyzing data after piloting the items with the Impact! tool. Construct validity defines how well a test or experiment measures up to its claims and reliability is the internal consistency of a measure (Shuttleworth, 2009).

Stage 1: Literature Review

In stage 1, the items for the Impact! draft questionnaire were developed. To come up with these appropriate items, the following information was taken into account: constructs that measure teaching quality and criteria requested from the developers of Impact!. The following research question guided this part of the research: *A. Which constructs measure teaching quality in higher education?*

To find literature linked to the research question, social science discourse was examined. The literature commonly addresses the following key-terms: student evaluations of teaching, student perceptions of teaching, formative evaluations/feedback, summative evaluations/feedback, student evaluations/feedback, and teaching quality in higher education. The search engines that were used to find literature are JSTOR and Google Scholar, which were both accessible via the University of Twente online library. While searching for relevant literature, the language was limited to English.

To determine appropriate constructs that measure teaching quality in higher education, a literature review was conducted to find SET's that include constructs that measure teaching quality in higher education. The type of constructs that were identified from the literature review reflect the degree to which a teacher performs specific teaching characteristics or practices (Feldman, 2007). To identify the constructs, the items of the five existing SET's were individually reviewed. These five existing SET's were selected to be a part of the literature review, because their motives for evaluation closely align with those of the Impact! tool in higher education. SET's that did not address teacher behavior were not included. Supportively, the literature review listed the constructs in independent paragraphs for the reader to compare (to see the full list of these constructs, refer to the chart in Appendix A).

Aside from the constructs identified in the literature review, the criteria requested from the developers of Impact! tool were also considered to formulate appropriate items for the Impact! draft questionnaire. These criteria were all communicated by Adrie Visscher and Hannah Bijlsma at the beginning of the research project and included the following (to see a full description of each criteria, refer to the chart in Appendix B) :

1. Topic: Student Evaluation of Teaching (SET) in higher education

2. Objective: To help teachers identify the quality of their teaching practices in higher education

3. Target Audience: Teachers in higher education

4. Target Context: Lectures at institutions in higher education

5. Assessment Type: Formative

6. Data Collection Instrument: Questionnaire

7. Form of Medium: Questionnaire is accessible using a digital application via a digital system

8. Psychometric Format: Scale and Survey

9. Number of Items: 15-20

10. Point of View: First person narrative; 'teacher- oriented'

11. Tense: Past- tense

12. Incorporation of constructs: constructs that reflect teaching characteristics or practices

Based on the results from the literature review and the criteria listed in Appendix B, a preliminary list of items for the Impact! draft questionnaire was created. The following research question guided this part of the research: *B. What are appropriate items to be in the questionnaire for the Impact! tool in higher education?* The preliminary list of items incorporated all of the teaching dimensions identified in the literature review. Afterwards, changes were made to the preliminary list of items for the Impact! questionnaire in higher education. It was agreed upon that the number of items should be reduced; the inclusion and exclusion of the items depended on the importance of the teaching dimension in higher education. Furthermore, improvements were made towards the formulation of the items; these aligned with the following suggested guidelines for effective questionnaire development (Lee, 2006): write simple, clear, and short; be specific and precise; use appropriate language; make them answerable; one topic per item; avoid negative items; avoid biased or loaded items; use appropriate emphasis for keywords.

Stage 2: Focus Groups & Cognitive Interviews

Stage 2 was divided into two parts: focus groups with experts and cognitive interviews with the target audience. The purpose of these two parts was to review the items in the questionnaire for the Impact! tool in higher education. Focus groups allowed a diverse group of individuals to give their insights about the subject of interest (Kitzinger, 1994); it helped determine the face validity of the items. Cognitive interviews helped to explore whether respondents understand a questionnaire in the way intended by the researcher (Collins, 2003); this helped determine the content validity of the items. Based

on the respondents given feedback, the items in the questionnaire for the Impact! tool in higher education were changed.

All respondents from the focus groups and cognitive interviews were volunteers who were asked to review the questionnaire for the Impact! tool in higher education. The respondents from the focus groups had to be teachers in higher education, and also be specialized in the field of Statistics, Psychology, Learning Sciences, or Educational Sciences. The respondents from the cognitive interviews had to be students who are currently a part of an undergraduate or graduate program in higher education. They were selected through goal-oriented sampling, and were asked for consent to participate via one-on-one approach, phone call, email, or private messaging. Respondents were all required to speak the English language, but ranged in education, gender, ethnicity, nationality, and age. For the focus group, 18 potential respondents were contacted, and total of 14 were willing to participate. For the cognitive interviews 10 potential respondents were contacted, and a total of 7 were willing to participate.

Stage 2 was guided by the two research questions:

C. Which items are relevant in the questionnaire for the Impact! tool in higher education?

D. In what way do the items have to be reformulated in the questionnaire for the Impact! tool in higher education?

The goal of the focus groups was to receive feedback from experts was used to determine the relevance and formulation of the item. To achieve this, two focus groups were organized. Each focus group included a minimum of six respondents. The focus groups were facilitated by Hannah Bijlsma and Shanice Harmon at the University of Twente. They were scheduled to take 1 hour and 30 minutes on April 7th and April 14th, 2018. Commentary from the respondents was written down by the facilitators. However, respondents also had the option to write down further comments on a sheet of paper, which were collected and reviewed by the facilitators afterwards.

Questions posed in the first round of the first focus group asked experts which teaching aspects they thought should be a part of a questionnaire that evaluates teaching quality in higher education. This round was done to determine the relevance of constructs in the Impact! questionnaire. Respondents were asked to write down a minimum of five constructs on a piece of paper. Each one of the suggestions had the potential to be classified into Feldman's (1976; 2007) categories.

Questions posed in the second round asked experts which constructs they preferred to be excluded in a questionnaire that evaluates teaching quality in higher education. This round was done to determine which constructs were not considered relevant. Respondents had to choose between the constructs in Feldman's (1976; 2007) categories. These constructs were individually listed on pieces of A4 paper that were hanging around the room. Respondents were asked to select and mark two constructs

they preferred to exclude using a red sticker. To evaluate these results from this round, the top three most frequently marked constructs were identified.

Questions posed in the third round related to the quality of the items in the questionnaire for the Impact! tool in higher education. This round was done to determine the relevance and alternative formulations of the presented items. Facilitators guided an open discussion about the items in chronological order. Respondents were given a hand-out that listed the items; the list also gave respondents the option to write comments in a chart that they did not verbally express in the open-discussion. To evaluate the results from this round, two sets of data were considered. First, whether the majority of the respondents rated the item as relevant. Second, participants remarks towards the formulation of the items. The respondent's ratings and remarks were identified from verbal responses during the open-discussion, as well as from non-verbal responses that were written on the hand-out.

The decision to include, exclude, or change an item in the questionnaire was based on the review of the focus groups results. Results were collectively reviewed on two points: first, which items did experts recommend being included and excluded; second, what are better alternative formulations of the items. Afterwards, the developers of the Impact! tool had a meeting to agree upon the adjustments in the questionnaire before the cognitive interviews.

The goal of the cognitive interviews was to evaluate the target groups comprehension of the items in the questionnaire for the Impact! tool in higher education. Depending on whether the respondents' interpretations of the items were correct or incorrect, the formulation of the items was adjusted. To achieve this goal, seven cognitive interviews were organized. The cognitive interviews were facilitated by Shanice Harmon at the students' location of convenience. They were scheduled to take 30- 45 minutes between April 23rd and April 28th, 2018. The commentary from the respondents was written down and notes included direct quotation from the participants involved. The facilitator of the cognitive interviews aimed to ask questions that assessed the students' comprehension of items in the Impact! questionnaire. During the cognitive interview, the target audience was asked how they understood each item, or to explain each item in their own words.

To evaluate the results from the cognitive interviews, one set of data was analyzed by the use of two methods. First, the respondents' comprehension of the items; this was analyzed using the Cognitive Classification Scheme (CCS) from Willis (2015), which is an instrument that supports the measurement of comprehension error. The CCS helped classify respondent's remarks about the items into one of the following categories: inaccurate instruction, vague topic/unclear question, complex topic, undefined/vague term, transition needed, unclear instruction to respondent, question too long, complex or awkward syntax, erroneous assumption, several questions. And second, participants' additional remarks towards the items, which were optionally shared by respondents.

The decision to include, exclude, or change an item in the questionnaire was based on the review of the cognitive interview results. Results were collectively reviewed on two points: first, which items were not correctly comprehended by the target audience; second, which additional remarks from the target audience were helpful to improve the formulation of the items. Based on the review, the developers of the Impact! tool had a meeting to agree upon changes that needed to be made to the items in the questionnaire.

Stage 3: Pilot Test

In Stage 3, the questionnaire was implemented in the Impact! tool and piloted in higher education. The goal was to measure the construct validity and reliability of the items, after they had been piloted at lectures in higher education using the Impact! tool, as well as gain informal feedback about the usage of the tool. This stage was guided by the following research questions: *E. What is the reliability of the items in the questionnaire for the Impact! tool in higher education? F. What is the construct validity of the items in the questionnaire for the Impact! tool in higher education?* The respondents for this stage were teachers who were willing to test the items in the questionnaire for the Impact! tool at the University of Twente. The respondents were selected through goal-oriented sampling. Only teachers who instructed lectures in higher education between May 4th and May 31st, 2018 were asked to participate. They were contacted via one-on-one approach, phone call, email, or private messaging. 55 potential respondents were contacted, and total of 20 respondents were willing to participate. The respondents ranged in education, gender, ethnicity, nationality, and age. They were teachers of the subjects of the faculties in Behavioral, Management and Social Sciences, Mathematics and Computer Sciences, and Science and Technology.

Participating respondents were sent a PowerPoint slide that gave their students instructions to evaluate the lecture using the Impact! tool. The PowerPoint instructed students to access an online-link that leads directly to the items in the questionnaire for the Impact! tool in higher education. Accessing the online-link was suggested to be possible with any digital devices (i.e. phone, laptop, computer, tablet). Students took a maximum of 5 minutes to respond to all the items in the Impact! tool. After students responded to the items, their feedback was anonymously communicated to the teachers.

Afterwards, the results were analyzed to measure the construct validity and reliability of the items in the questionnaire for the Impact! tool in higher education using the IBM SPSS Statistics software. To measure the reliability of the questionnaire, the average Cronbach alpha of the whole questionnaire was evaluated. And to measure the construct validity of the questionnaire, exploratory factor analysis (EFA) was applied, which lead to the recognition of factors that loaded individual items (Kim, 1978). Items that included Likert-scale responses were analyzed, but items that had an open-response option were not

analyzed. The reason for this if that only the quantitative data could be analyzed using a psychometric method for construct validity and reliability (Shuttleworth, 2009). In order for pilot groups to be included for reasonable analysis, there had to be a minimum of 6 students who responded to the items in the Impact! tool (Shuttleworth, 2009). As a result, only 13 of the 24 executed pilots were considered for analysis.

Proceeding the pilot, teachers and students were also informally asked questions about their experience with the Impact! tool. And in case they had further suggestions or concerns, students were asked to approach the researcher after their lecture, and teachers were asked to contact the researcher via email.

Results

This section includes the analyzed results from the literature review, focus groups, cognitive interviews, and pilot tests. Each section is followed with the updated version of the Impact! questionnaire for higher education.

Stage 1

Literature Review

The literature review aligns with the 12th criterion request by the Impact! developers: the incorporation of constructs that measure student perceptions of teaching quality from existing SET's. The items in these SET's are suggested to indicate the degree of a teachers' pedagogical performance in the classroom (Liston, Borto, & Whitcomb, 2008). The particular SET's that were chosen to be incorporated are the Student Evaluation of Educational Quality (SEEQ), Uvalon, Impact! tool, Nationale Studenten Enquête (NSE), and the Behavioral, Management, and Social Sciences (BMS) summative course evaluation instrument. These SET's were chosen, because they feature the characteristics of effective teaching, and also align with goal of this research project: to measure students perceptions of teaching quality in contexts of education. And although the SET's differ, their constructs can be grouped into similar categories, using framework from Feldman (1997, 2007); refer to appendix A to see a chart that lists these classifications in detail. Some of the constructs identified in the SET's were chosen to be adapted into the items in the questionnaire for the Impact! tool in higher education.

The most thoroughly tested SET is the Student Evaluation of Educational Quality (SEEQ) from Dr. Herbert Marsh (1984, 1987) of the University of Sydney. It has been used in over 50,000 courses, which includes educational contexts that are international. Moreover, the SEEQ has been determined to have a high quality: "Its reliability is good when based upon 10 to 15 or more student responses. And the ratings have successfully been validated against the retrospective ratings of former students, student

learning as measured by objective examination, affective course consequences, and staff self-evaluations of their own teaching effectiveness" (Marsh 1982; p.2). The SEEQ includes the following nine constructs that are suggested to measure teaching quality in higher education:

1. Organization and clarity
2. Breadth of coverage
3. Instructor enthusiasm
4. Individual rapport
5. Group interaction
6. Learning and value
7. Examinations/grading
8. Assignments and Readings
9. Difficulty and workload

The Uvalon is another SET, which has been designed by Vorst & Can Engelenburg (1992) from the University of Amsterdam. The Uvalons' constructs are based on the SEEQ from Marsh, and the Evaluation of Lecturing (Evalec) from De Neve and Janssen (1982). Vorst & Can Engelenburg (1992) acknowledge that their instrument includes psychometric features that are related to teaching behavior, teaching characteristics, and student behavior. The constructs in the Uvalon are similar to the ones in the SEEQ:

1. Structure
2. Explication
3. Stimulation
4. Validation
5. Instruction
6. Conversation
7. Interaction
8. Learning and value
9. Examination
10. Literature
11. Workload and difficulty.

A further SET is the questionnaire in the Impact! tool. As already mentioned in the introduction section, this instrument has been designed by researchers from the University of Twente in 2017 (Bijlsma, Visscher, Veldkamp, & Dobbelaer; under review). It aims to collect student perceptions of teaching quality at the end of lessons in secondary education. The constructs featured in this SET are based on

results of a meta-analyses about teaching practices which are known to be effective for student learning. The researchers from this study categorized the teaching practices into seven general categories:

1. Creating a supportive and positive classroom climate

2. Well-organized and structured classroom management

3. Clear instruction

4. Adaptive instruction

5. Focus on quality of the teacher-student interaction

6. Cognitive activation to promote deep learning

7. Assess students' learning during the lesson

Another SET is the Dutch National Student Survey or Nationale Studenten Enquête (NSE, 2017). The NSE (2017) is a nationwide survey that is conducted annually among students in Dutch higher education in the form of an online questionnaire. An assessment of the responses, theme scores and weighting factors of the NSE indicated that it is significantly reliable (Brenders & Roggenweg, 2013). And unlike the SEEQ, Uvalon, and Impact!, the NSE includes constructs that measure teaching quality, as well as annual students' perceptions of their course program. There are eighteen constructs in the NSE, but only the following six measure teaching quality:

1. Program content
2. Skills acquired
3. Teachers and lecturers
4. Testing and assessment
5. Study load
6. Quality care

The final SET is the Behavior, Management, and Social Sciences (BMS) course evaluation instrument. This SET is used by the University of Twente to evaluate teaching quality at the end of a course from the BMS department. The items included a total of sixteen different constructs:

1. Quality of subject matter
2. Quality of study material
3. Learning objectives
4. Course organization
5. Course environment

6. Workload and Difficulty
7. Knowledge gained/ learning
8. Clarity
9. Quality of Lecture
10. Assignments
11. Student-teacher interaction
12. Examination
13. Language-use of teacher (ie. English)
14. Enthusiasm
15. Expertise of knowledge
16. Presentation skills

Impact! Questionnaire- 1st Draft

Below are the items in the first draft of the questionnaire for the Impact! tool in higher education:
1. The teacher helped me learn valuable content during the lecture.
2. The teacher challenged me intellectually about the lecture content.
3. In the lecture, the teacher connected to what I already knew. (option 1)
   The teacher connected the lecture with content to what I already knew (prior knowledge). (option 2)
4. The teacher's lecture added value to the material to be studied for this lecture.
5. The teacher presented the lecture content clearly. (option 1)
   The teacher explained the lecture content clearly. (option 2)
6. The teacher clearly stated the learning objectives of the lecture.
7. The teacher presented content that aligned with the learning objectives of the lecture.
8. The teacher summarized what we had learned at the end of the lesson.
9. If I did not understand the lecture content, the teacher made sure I understood it. (option 1)
   In case I did not understand it, the teacher clearly clarified the lecture content (option 2)

10. The teacher made me more interest in the lecture content. (option 1)
    The teacher raised my interest about the lecture content. (option 2)
11. The teacher's lecture was well structured.
12. The teacher's lecture was at a good pace.
13. The teacher encouraged me to interact with other students about the lecture content during the lecture.
14. The teacher encouraged me to interact with him/her about the lecture content during the lecture.
15. The teacher created a good environment for learning. (option 1)
    The teacher created a comfortable environment for learning. (option 2)
16. The teacher made me feel comfortable asking him/her questions.
17. The teacher gave meaningful answers when questions were asked about the lecture content.
18. The teacher spoke English clearly.
19. During the lecture, the teacher checked if we understood the lecture content.
20. The teacher's lecture could be improved if he/she…
21. This is what I liked about the lecture…

Stage 2

Focus Groups Data
This section lists a combination of the results from the two focus groups.

First Round:
Below are the results of the 56 teaching aspects that were suggested by the respondents:
1. What did the student learn?
2. What was most important/relevant?
3. Which questions are still unanswered?
4. What information did students miss?
5. How would the student qualify the lecture in 3 sentences?
6. What can be improved?
7. Main points clear?
8. Do I know how to pass the exam?
9. Can I apply the content taught?
10. Topics (taking into account the reading material)
11. Flow (media/storyline)
12. Understanding (language use, logical storyline, easy to follow)
13. What was missed?
14. Structure of the lecture
15. Speech (clear, tempo, quality of English)
16. Match of content and difficulty with learning goals and student expectations
17. Interaction; did the teacher stimulate active thinking (teaching methods)?
18. Classroom activities; do they help master content?
19. Explanations; were they clear, and relevant for the program?
20. Pace; speech, between topics?

21. Answering questions
22. Relationship with students; is the teacher accessible, interested in the student's well-being?
23. Connection to prior knowledge
24. Motivating
25. Contributing to learning goals and expectations
26. Feedback, answering questions of students
27. Presentation and explanation
28. Presentation of the content; interesting and relevance
29. Providing insight in why the content is important
30. Use of examples
31. Usefulness of assignments
32. Relation to reading materials
33. Was the meeting useful in terms of delivery and new information
34. Did the teacher responded well to questions?
35. Were you encouraged to spend time on studying at home?
36. Did the teacher manage the lesson well?
37. Did the teacher create a good learning atmospheres?
38. The teacher made me excited about the content?
39. If I had read the materials I would know the same as after this lecture
40. The teacher was enthusiastic about what he taught
41. I learned a lot
42. Structure of the lecture
43. Understanding of theories
44. Room for questions
45. Interaction
46. Depth of subject matter
47. Interactivity
48. Openness to questions
49. Interesting content
50. Opportunities to apply content/ knowledge
51. Openness to look into different side-topic on request
52. Learning goals and successful criteria
53. Teacher/students and student-student interaction
54. Quality instruction
55. Quality feedback
56. Formative assessment

Second Round:
The three most common constructs that participants chose to be excluded in the Impact! questionnaire are: organization and classroom management, assessment and grading, and general satisfaction.

Third Round:
Refer to Appendix C to read the feedback that focus group respondents gave for each item.

Table I

Decisions made to change items in the first draft of the Impact! questionnaire in higher education; based on the focus group data from the first, second, and third round:

Item	Reason to Change Item
Pre-Item: The teacher helped me learn valuable content during the lecture.	This item was eliminated from the questionnaire, because it was rated low in relevance. Moreover, the many focus group participants found the term 'valuable' to be too subjective and vague.
2. The teacher challenged me intellectually about the lecture content.	In this item, the phrase 'challenged me intellectually' was replaced with 'made me think deeper', because focus group participants thought an alternative wording would make it easier to understand the item.
3. In the lecture, the teacher connected to what I already knew (option 1). The teacher connected the lecture with content to what I already knew (prior knowledge) (option 2).	This item was eliminated from the questionnaire, because focus group participants questioned whether it is relevant for every lecture.
4. The teacher's lecture added value to the material to be studied for this lecture	This item alternatively changed the phrase 'added value to the material to be studied for this lecture' to 'helped me understand the study material better', because focus group participants had difficulties comprehending the meaning of 'added value'.
5. The teacher presented the lecture content clearly (option 1). The teacher explained the lecture content clearly (option 2).	The second version of this item was chosen to be used in the questionnaire, because focus group participants thought it has more of a positive influence towards student learning.
6. The teacher clearly stated the learning objectives of the lecture.	No changes were made to this item.
7. The teacher presented content that aligned with the learning objectives of the lecture.	This item was eliminated from the questionnaire, because focus group participants thought it was too similar to item 6.
8. The teacher summarized what we had learned at	This item was eliminated from the questionnaire, because focus group participants thought that if

the end of the lesson.

item 7 is included, then this item can be excluded; items 7 implies that the teacher behavior in item 8 is fulfilled.

9. If I did not understand the lecture content, the teacher made sure I understood it (option 1). In case I did not understand it, the teacher clearly clarified the lecture content (option 2).

This item was eliminated from the questionnaire, because focus group participants did not think that the behavior addressed in this item is realistic for the teacher to fulfill. The participants explained that the teacher cannot address all of the students' individual needs during the lecture.

10. The teacher made me more interest in the lecture content (option 1). The teacher raised my interest about the lecture content (option 2).

Focus group participants discussed that enhancing student achievement is the result of keeping a student's attention; thus, they agreed that option 1 is more relevant, because that is closer to the construct that it measures.

11. The teacher's lecture was well structured.

No changes were made to this item.

12. The teacher's lecture was at a good pace.

No changes were made to this item.

13. The teacher encouraged me to interact with other students about the lecture content during the lecture.

This item was eliminated from the questionnaire because focus group participants questioned whether this item is relevant for all lectures; this is because, some class sizes and lecture content are not suitable for interaction.

14. The teacher encouraged me to interact with him/her about the lecture content during the lecture.

This item was eliminated from the questionnaire because of the item limit in the questionnaire; the item was not considered a priority by the Impact! developers.

15. The teacher created a good environment for learning (option 1). The teacher created a comfortable environment for learning (option 2).

This item was eliminated from the questionnaire because of the item limit in the questionnaire; the item was not considered a priority by the Impact! developers.

16. The teacher made me feel comfortable asking him/her questions.

This item was eliminated from the questionnaire because of the item limit in the questionnaire; the item was not considered a priority by the Impact! developers.

17.The teacher gave meaningful answers when questions were asked about the lecture content.

The phrase 'about the lecture content' was added to the item, because it helps clarify which questions it is referring to.

18. The teacher spoke English clearly.

This item was eliminated from the questionnaire, because the focus group participants agreed that this item is not relevant for formative feedback; instead, it should be a part of a course evaluation.

22

19. During the lecture, the teacher checked if we understood the lecture content.	No changes were made to this item.
20. The teacher's lecture could be improved if he/she…	No changes were made to this item.
21. This is what I liked about the lecture…	No changes were made to this item.

Impact! Questionnaire- 2nd Draft

Below are the items in the second draft of the questionnaire for the Impact! tool in higher education.

Pre-item: I studied all the reading material in preparation for the lecture.

*based on the the answers of the student, the scores are presented to the teacher

1. The teacher asked questions that made me think about the lecture content.
2. The lecture helped me understand the study material better.
3. The teacher explained the lecture content clearly.
4. The teacher clearly stated the learning objectives of the lecture.
5. The teacher made me (more) interested in the lecture content.
6. The teacher's lecture was well structured.
7. The teacher's lecture was at a good pace.
8. The teacher made us actively process the lecture content (e.g. questions, discussions).
9. During the lecture the interaction between teacher and students was pleasant.
10. The teacher gave meaningful answers if questions were asked.
11. The teacher checked if we understood the lecture content.
12. I now understand the lecture content.
13. This is what I liked about the lecture…
14. This is what I still do not understand about the lecture content…
15. The teacher's lecture could be improved if he/she…

Cognitive Interviews Data

Refer to Appendix D for the feedback that cognitive interview respondents gave for each item.

Table II

Decisions made to change items in the second draft of the Impact! questionnaire in higher education; based on cognitive interview data:

Item	Reason to Change Item
Pre-Item: I studied all the reading material in preparation for the lecture.	The terms 'studied' and 'reading material' should be excluded from the item, because they do not relate to one another. Instead, the item should address the construct of 'preparation' in a different manner.
2. The teacher asked questions that made me think about the lecture content.	The phrase 'asked questions' should be excluded from the item, because students can think about the lecture content, without directly having to be asked questions. And in order to measure whether student are engaged in higher-order thinking, the item should be reformulate in a way that measures students 'deep' learning.
3. The lecture helped me understand the study material better.	The term 'study material' should be explained different in the item, because the majority of the cognitive interview respondents found it to be vague.
4. The teacher explained the lecture content clearly.	No changes were made to this item.
5. The teacher clearly stated the learning objectives of the lecture.	The phrase 'at the beginning' should be excluded from the item, because it is already assumed by the majority of the cognitive interview respondents that the learning objectives are presented early on in the lecture.
6. The teacher made me (more) interested in the lecture content.	The term '(more)' should be excluded from the items, because it makes the term have a double connotation; this can cause the item to have low validity, because it is understood in two different ways.
7. The teacher's lecture was well structured.	The term 'teacher's' should be excluded in the item, because it is. Furthermore, the phrase 'was well structured' should be replaced with 'had a

	good structure', because it sounds more natural.
8. The teacher's lecture was at a good pace.	The cognitive interview respondents understood the term 'pace' differently. Thus, the item should also clarify which type of 'pace' it is addressing.
9. The teacher made us actively process the lecture content (e.g. questions, discussions).	This item should be deleted from the questionnaire, because it measures the construct of deep-thinking/ learning, which is already addressed in the 2nd item.
10. During the lecture the interaction between teacher and students was pleasant.	This item should also begin with the term 'the teacher', because then it is more consistent with the other items. Moreover, the term 'pleasant' should be replaced with 'positive', because it can potentially measure the construct of 'student-interaction' more objectively.
11. The teacher gave meaningful answers if questions were asked.	The term 'meaningful' should be replaced with 'helpful', because it is less vague.
12. The teacher checked if we understood the lecture content.	No changes were made to this item.
13. I now understand the lecture content.	No changes were made to this item.
14. This is what I liked about the lecture…	No changes were made to this item.
15. This is what I still do not understand about the lecture content…	This item should include the term 'still' before 'do not understand about the lecture content'; it makes it more clear that the item aims to measure whether the teacher clarified the students understanding of the study material.
16. The teacher's lecture could be improved if he/she…	This item should not begin with the term 'the teacher', because its aim was changed to measure the quality of the lecture, and not teaching behavior.

*Applicable to all items: Items that include the term 'teacher' was changed to 'teacher(s)', because the Impact! tool has the potential to evaluate more than one teacher.

Impact! Questionnaire - Third Draft (Pilot Version)
Below are the items in the third draft of the questionnaire for the Impact! tool in higher education:
    Pre-Item: I prepared well for the lecture.
    1.   The teacher(s) made me think deeper about the lecture content.

2. The teacher(s) increased my understanding of what I prepared for the lecture.
3. The teacher(s) explained the lecture content clearly.
4. The teacher(s) clearly stated the learning objectives at the beginning of the lecture.
5. The teacher(s) made me interested in the lecture content.
6. The lecture had a good structure.
7. The teacher(s) presented the lecture content at a good pace.
8. The quality of the PowerPoint was good.
9. The teacher interacted in a positive way with the students during the lecture.
10. The teacher(s) gave helpful answers if questions were asked.
11. The teacher(s) checked if we understood the lecture content.
12. I now understand the lecture content.
13. This is what I liked about the lecture:
14. This is what I do not understand about the lecture content:
15. The lecture could be improved if…

Stage 3

Summary of Pilot Test Data

A total of 16 items were piloted amongst the test groups; these are the items that the Impact! developers agreed upon for the third draft of the questionnaire for the Impact! tool in higher education. Because of methodological limitations, only items 2-12 (from the third draft of the questionnaire) could be analyzed descriptively for reliability and construct validity. Moreover, the results from 13 out of the 20 executed pilots were analyzed; the other 7 pilot groups were excluded, because there were not enough respondents for effective analysis of reliability and construct validity.

The research-project aimed to measure the reliability of the questionnaire for the Impact! tool in higher education. According to Tavakol and Dennick (2011) Cronbach Alpha can range between: >.9 (excellent), >.8 (good), >.7 (acceptable), > .6, >.5 (poor), and <.5(unacceptable). Correspondingly, the questionnaire had an average Cronbach's Alpha score of .75, which is considered to be acceptable. Pilot 7 had the highest Cronbach Alpha with an excellent score of .94. Pilot 1 had the lowest Cronbach Alpha with a poor score of .53. Hence, none of the pilots were identified to have a Cronbach's Alpha score with an unacceptable reliability (<.5).

The analysis also helped determine which items are most frequently identified as having the lowest and highest reliability (i.e. mode of the results). The data indicated that item 10 ('The teacher interacted in a positive way with the students during the lecture') was most frequently identified as having the lowest reliability (4 out of the 13 pilots). And item 3 ('The teacher(s) explained the lecture content clearly') was most frequently identified as having the highest reliability (4 out of the 13 pilots). These findings suggest that the developers of Impact! should include item 3 in the next version of the questionnaire, but might want to reconsider the formulation of item 10.

To determine the underlying common factor(s) of the interrelated items in the questionnaire, all of the data from the 13 pilots was combined for analysis. The EFA results indicated that there is one underlying common factor that is responsible for the covariation in the data, because it is the only one with an eigenvalue greater than one. This factor has a component number of 4,708, and variance of 39,234. This means that this single component is significant in terms of underlying the items in the questionnaire.

Correspondingly, the 12 quantitative items from the questionnaire are highly loaded on the one underlying common factor (refer to *Scree Plot I*). The average component value is .617. Item 3 has the highest component value of .82. Item 10 has the lowest component value of .47. These findings suggest that having only one factor means that all items fit onto a single theoretical construct, which the Impact! developers assume to be 'teaching quality'.

Scree Plot I
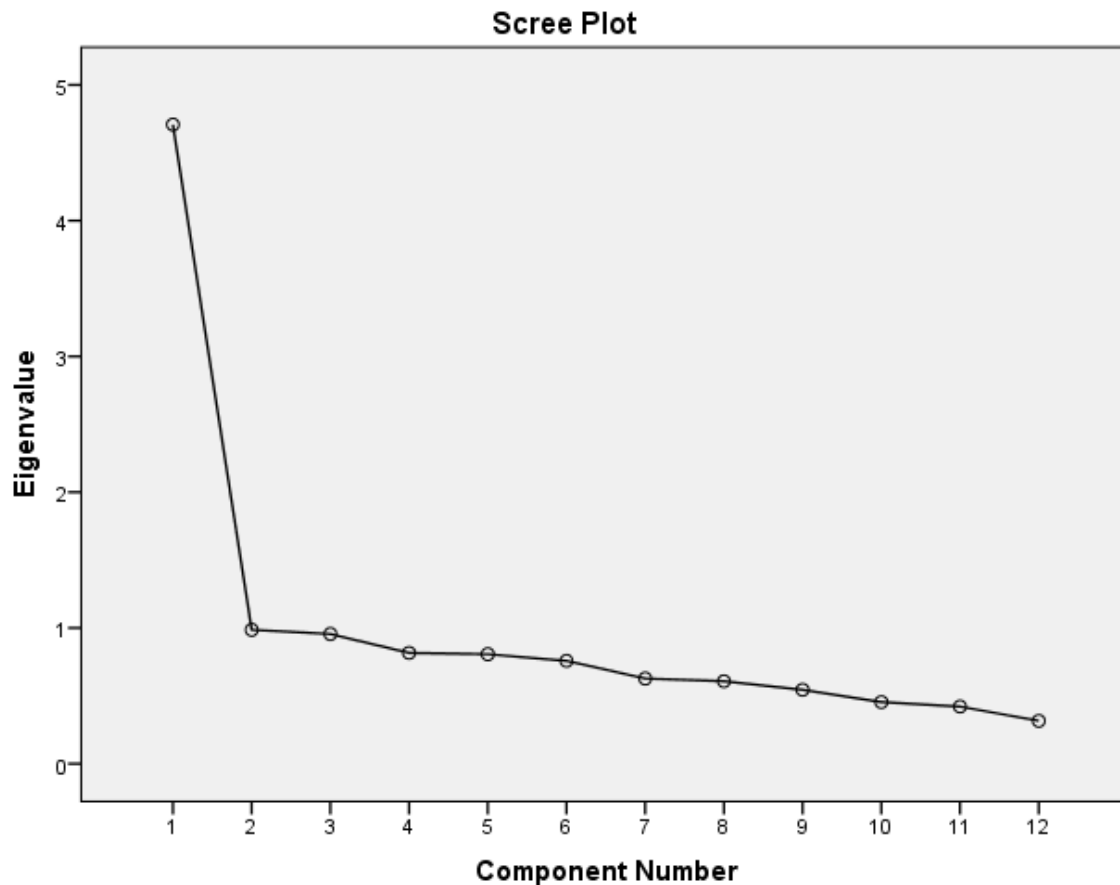
Informal Feedback about the Impact! tool from the Pilot Groups

The Impact! developers also received informal feedback about the pilot study from students and teachers involved in the pilot groups. Students shared the following points of concern and suggestions:

- The introduction to the Impact! tool is in Dutch and not English
- The first item in the Impact! tool asks about the average score in the course, but this is not always known by students in every course
- Add the 'NA' option to the item that asks about lecture preparation
- Alternate the term 'PowerPoint' to 'presentation slides'
- The questionnaire is to extensive
- The questionnaire is too repetitive to be done every lecture
- The list of items is out of order in the Impact! tool
- Add additional text boxes for people to explain their answers
- Difficulties accessing the items in the Impact! tool whenever using the web-browser Firefox or Internet Explorer

Additionally, teachers shared the following points of concerns and/or suggestion:

- Students shared non-relevant information in the open-response items; there was a particular case that was about a students' mental health
- More pilot tests would be beneficial to make a judgment on whether there was improvement in teaching quality

Discussion and Conclusion

Below are the final items that have been adjusted to be in the questionnaire of the Impact! tool in higher education:

Pre-item: I prepared well for the lecture.
1. The teacher(s) made me think deeper about the lecture content.
2. The teacher(s) increased my understanding of what I prepared for the lecture.
3. The teacher(s) explained the lecture content clearly
4. The teacher(s) clearly stated the learning objectives at the beginning of the lecture.
5. The teacher(s) made me interested in the lecture content.
6. The lecture had a good structure.
7. The teacher(s) presented the lecture content at a good pace.
8. The quality of the presentation slides was good.
9. The teacher interacted in a positive way with the students during the lecture.
10. The teacher(s) gave helpful answers if questions were asked.
11. The teacher(s) checked if we understood the lecture content.
12. I now understand the lecture content.
13. This is what I liked about the lecture:
14. This is what I do not understand about the lecture content:
15. The lecture could be improved if…

The research-project sought to answer the following main research question: *To design a valid and reliable instrument for measuring student perceptions of teaching quality after a lecture in higher education by means of the Impact! tool, which items should be included in the Impact! questionnaire?* The items (1-12) were determined to have an acceptable reliability of .75. Moreover, the items all loaded on one single factor, and thus related to one construct, which the Impact! developers assume to be 'teaching quality'. Thus, the pilot version of the questionnaire can be considered to include items that constitute a valid and reliable instrument for measuring student perceptions of teaching quality after a lecture in higher education. And because the development of the questionnaire followed an in-depth design procedure from Pride (2005), by means of cognitive interview, focus groups and pilot group testing, the final items are assumed to cover its goal of measurement: student perceptions of teaching quality in higher education. Nonetheless, the Impact! developers decided to make only one additional change to the items in the pilot version of the questionnaire. This change was based on the informal comment provided by the participants in the pilot group; in item 9 the term 'PowerPoint' was changed to the phrase 'presentation slides'.

The items in the questionnaire for the Impact! tool in higher education aims to be implemented in lectures at the institution of the University of Twente. These are meant to be relevant for teachers in university contexts to improve their teaching quality. However, in order to enhance the effectiveness of the questionnaire, there are several theoretical and methodological limitations that are suggested to be addressed through additional testing. These limitations are related to the research process and findings from the focus groups, cognitive interviews, and pilot study.

To begin, the focus groups were limited in terms of the specialization of respondents, discussion points, and methodology. The focus groups included respondents from the University of Twente that are specialized in the fields of Statistics, Psychology, Learning Sciences, and Educational Sciences. These experts had a profound understanding of constructs related to collecting student perceptions of teaching quality; thus, they could give significant feedback about the content in the questionnaires. The focus groups could have also been enhanced by developing discussion related to the scaling of the items. This discussion did not happen, because the time of the focus groups were limited to 90 minutes. This limitation prevent discussion that went further than the relevance and formulation of the items in the questionnaire. Inviting respondents who are specialized in the fields of Linguistics and Communication Sciences could have positively expanded the diversity of responses in the focus groups. They might have been able to given more qualified suggestions towards the phonological and grammatical features of the item in questionnaire. But since did not have contact with people specialized in these two fields, these other types of respondents were excluded. Finally, due to practical considerations, the methodology of

documenting the respondents' comments were limited to note-taking. However digital recordings might have been more useful for Impact! developers to extract more detailed information from the discussion.

Furthermore, the cognitive interviews helped determine whether students understood the items in the questionnaire correctly, but it would have been helpful to conduct further cognitive interviews with the aim of comparing the responses of Dutch students to those of other International students. And although both were included, there was not a sufficient number of Dutch students to conduct an effective reliability analysis. This comparison is suggested, because it can give insight into whether these two groups of students comprehend the items differently or not. The results from this comparative analysis could be relevant, because the Impact! tool is aimed to be implemented at the University of Twente, which is an institution with a dominantly Dutch student population.

The evaluation of validity and reliability from the pilot group also had limitations. To begin, the pilot group results only assessed the quantitative responses using factor (items 1-12 in the pilot questionnaire). There was not any analysis of the qualitative responses that students provided in the open-response questions. Using the data that was collected in the present study, it would be insightful to further evaluate the quality of these responses; for future research, this can be done by looking at repeating patterns and trends amongst the comments. Correspondingly, the evaluation of the questionnaire had a methodological limitation, because Cronbach's Alpha was used. Sijtsma, K. (2009) argued that this form of reliability measurement has been identified to have two problems: "First, alpha always has a value, which cannot be equal to a test group results reliability given the inter-item covariance matrix and the usual assumptions about measurement error. Second, in practice, alpha is used more often as a measure of the test's internal consistency than as an estimate of reliability.

In the research-project students and teachers gave informal recommendation of what should be taken into account by the developers of the Impact! tool to improve the quality of the questionnaire. But for even better feedback from target group participants after the pilot of the questionnaire, it is suggested that students and teachers are asked questions by means of formally structured interviews or surveys; this is a more systematic method of determining the instruments strengths and weaknesses. Such formal questions would allow an objective assessment of the target groups attitude, satisfaction and perceptions towards the Impact! tool: *How do you respond to being forced to use the Impact! tool? How many items are you willing to answer in the Impact! tool before you get bored, frustrated, or consider quitting? How many times are you willing to use Impact! tool throughout the duration of the course? What elements could be included in the Impact! tool to make it more engaging for you to use? What do you think about the design of the Impact! tool? Do you feel like the product was design for you? Is there anything distracting about the Impact! tool?*

Moreover, the overall research could have also been strengthened if it asked the research question: *To what extent is the questionnaire biased or not biased in regard to other confounding variables?* This question is crucial, because as was mentioned in the theoretical framework section, a combination of researchers have found these confounding variables to influences SET results: course characteristics, instructor characteristics, student characteristics, and time of feedback; Coles, 2002; Greenwald & Gilmore, 1997; Rowley, 1996; Rowley, 2003; Wachtel, 1998; Worthington, 2002). But because this research question was beyond the objective of this research-project, it was not taken into account and it is recommended to investigated in further research.

Next, when the questionnaire was piloted in the target group, there were technical problems that prevented students from accessing the items in the Impact! tool. Students experienced difficulties accessing the items in the Impact! tool whenever they used the web-browser Firefox or Internet Explorer. Thus, it is suggested that these technical problems are fixed in the Impact! tool, and piloted once more for ecological validity.

The last limitation is that the pilot of the Impact! tool lasted a total of three weeks with a total of 13 pilot groups. Although the size and duration of target group respondents allowed for effective data analysis, a larger sample group would have given a better evaluation of the validity and reliability, as well as the applicability of the questionnaire. To determine whether the Impact! tool mediates teacher quality for improved student achievement, it is best to benchmark the questionnaire amongst target groups, or use a longitudinal study to compare the results over an even longer period of time; both methods could help determine to what extent the teacher quality has increased, and whether students' achievements has improved. Also, extending the size and duration of the pilot would also give researchers the opportunity to determine whether the tool matches across all faculties of the University of Twente: Behavioral, Management and Social sciences, Engineering Technology, Electrical Engineering, Mathematics and Computer Sciences, Science and Technology, and Geo-Information Science and Earth Observation. Since lectures vary between subjects in the different programs, it would be useful to conduct more pilot tests amongst all of these five faculties and consequently compare the relevance of the questionnaire in each one.

Overall, this research-project had the objective to develop items for a new SET instrument that is to be used in higher education. The SET has three characteristics that were identified to not exist in other student feedback questionnaires in higher education. First, the questionnaire obtains student perceptions of teacher quality at the end of a lesson; this allows teachers to improve their next session accordingly. Second, the questionnaire obtains student perceptions of teacher quality through the means of a digital feedback system; this enhances the efficiency of data collection and analysis. And third, since there have

been several methodological doubts in terms of measuring student perceptions of teacher quality, this research has included extensive validity and reliability testing.

References

Abend, Gabriel. "The Meaning of Theory." *Sociological Theory* 26 (June 2008): 173–199;
Swanson, Richard A. *Theory Building in Applied Disciplines*. San Francisco, CA:
Berrett-Koehler Publishers 2013.

Berk, R.A. (2005). Survey of 12 Strategies to Measure Teaching Effectiveness. International
Journal of Teaching and Learning in Higher Education **17**(1): 48-62.

Bijlsma, H. J. E., Visscher, A. J., Veldkamp, B. P., & Dobbelaer, M. J. (under review). Does
smartphone-assisted student feedback impact teaching quality? Technology, Pedagogy
& Education.

Boyer, E. L. (1983). *High school: A report on secondary education in America*. Harper & Row, Inc., 10
East 53rd Street, New York, NY 10022.

Brenders, P., & van Roggenweg, G. (2013, August). Response, reliability, and validity of the Dutch
national student survey. In *EAIR 35th Annual Forum in Rotterdam, the Netherlands*.

Burgess, T. F. (2001). Guide to the Design of Questionnaires. *A general introduction to the
design of questionnaires for survey research*, 1-27.

Centra, J. A. (1993). *Reflective Faculty Evaluation: Enhancing Teaching and Determining
Faculty Effectiveness*. *The Jossey-Bass Higher and Adult Education Series*.
Jossey-Bass Inc., 350 Sansome St., San Francisco, CA 94104.

Centra, J.A. (2003). Will teachers receive higher student evaluations by giving higher grades
and less course work? *Research in Higher Education* 44(5): 495–518.

Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-
analysis of findings. *Research in higher education*, *13*(4), 321-341

Coles, C. (2002). Variability of student ratings of accounting teaching: evidence from a
Scottish business school. *International Journal of Management Education*,
2(2), 30-40.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. Quality
of Life Research, 12, 229-238.

De Neve HMF, Janssen PJ. Validity of student evaluation of instruction. Higher Education. 1982;11(5):543–52.

Dickey, D., & Pearson, C. (2005). Recency effect in college student course evaluations. *Practical Assessment, Research and Evaluation*, *10*(6), 1-10.

Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, *24*(3), 331-350.

Feldman, D. C. (1976). A contingency theory of socialization. *Administrative science quarterly*, 433-452.

Feldman, K. A. (1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Research in Higher Education*, *6*(3), 223-274.

Feldman KA. Identifying Exemplary Teachers and Teaching: Evidence from Student Ratings. In: Perry RP, Smart JC, editors. The scholarship of teaching and learning in higher education: An evidence-based perspective. New York: Springer; 2007. p. 93–143.

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: interviews and focus groups. *British dental journal*, *204*(6), 291.

Greenwald, A. G., & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course   workload in student ratings of instruction. *Journal of Educational Psychology*, *89*(4), 743.

Joyce, B. R., & Showers, B. (2002). Student achievement through staff development.

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues* (No. 14). Sage.

Kitzinger J. (1994) 'The methodology of focus groups: the importance of interaction between research participants', *Sociology of Health* 16 (1): 103-21.

Korfhage, R. R. (2008). Information storage and retrieval.

Lee, S. (2006). Constructing effective questionnaires. *Handbook of human performance technology, Hoboken, NJ: Pfeiffer Wiley*, 760-779.

Leech, N. L., & Onwuegbuzie, A. J. (2007). A typology of mixed methods research designs. *Quality & Quantity*, *43*(2), 265–275. doi:10.1007/s11135---007---9105---3

Liston, D., Borko, H., & Whitcomb, J. (2008). The teacher educator's role in enhancing
teacher   quality.

Marsh, H. W., Overall, J. U., & Kesler, S. P. (1979). Validity of student evaluations of instructional
effectiveness: A comparison of faculty self-evaluations and evaluations by their students. *Journal of Educational Psychology*, *71*(2), 149.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity,
potential baises, and utility. *Journal of educational psychology*, *76*(5), 707.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings,
methodological issues, and directions for future research. *International journal of educational research*, *11*(3), 253-388.

Marsh, H. W. & Dunkin, M. J. (1992) Students' evaluations of university teaching: a
multidimensional perspective, in: J. C. Smart (Ed.) Higher education: handbook of theory and research, volume 8 (New York, Agathon Press).

Marsh, H. W., & Roche, L. A. (1994). *The Use of Students' Evaluations of University
Teaching to Improve Teaching Effectiveness: Final Project Report*. Australian Government Pub. Service.

McKeachie, W. J. (1987). Teaching and Learning in the College Classroom. A Review of the Research
Literature (1986) and November 1987 Supplement.

Means, B. (1993). *Using technology to support education reform*. US Government Printing
Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328.

Means, B., & Olson, K. (1997). *Technology and education reform: Studies of education
reform*. Diane   Publishing.

Mohsen Tavakol and Reg Dennick. Making Sense of Cronbach's Alpha. International Journal
of Medical Education. 2011; 2:53-55 Editorial

Narasimhan, K. (2001). Improving the climate of teaching sessions: the use of evaluations by
students and instructors. *Quality in Higher Education*, *7*(3), 179-190.

Neuman, W. L. (2013). *Social research methods: Qualitative and quantitative approaches*.
Pearson education.

Onwuegbuzie, A. J., & Leech, N. L. (2006). Linking research questions to mixed methods data analysis procedures 1. *The qualitative report*, *11*(3), 474-498.

Pride, Ferrell (2010). Marketing. South-Western Cengage Learning

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in higher education*, *16*(2), 129-150.

Rowley, J. (1996). Measuring quality in higher education. *Quality in Higher Education*, *2*(3), 237-255.

Rowley, J. (2003). Designing student feedback questionnaires. *Quality assurance in education*, *11*(3), 142-149.

Richardson, J. T. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & evaluation in higher education*, *30*(4), 387-415.

Shuttleworth, M. (2009). External validity. *Explorable Psychology*.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107.

Steiner, D. D., & Rain, J. S. (1989). Immediate and delayed primacy and recency effects in performance evaluation. *Journal of Applied Psychology*, *74*(1), 136.

Stieger, S., & Burger, C. (2010). Let's go formative: continuous student ratings with Web 2.0 application Twitter. *Cyberpsychology, Behavior, and Social Networking*, *13*(2), 163-167.

Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education*, *20*(3-4), 165-184.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, *2*, 53.

Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: a search for truth or a witch hunt in student ratings of instruction?. *New directions for institutional research*, *2001*(109), 45-56.

University of Wisconsin-Platteville, 2018. Major Differences between High School and College. Retrieved from https://www.uwplatt.edu/counseling-services/major-differences-between-high-school-and-college

Vorst HCM, Van Engelenburg B. UVALON UvA-pakket voor onderwijsevaluatie. Amsterdam, The Netherlands: Psychological Methods Department, University of Amsterdam, 1992.

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, *23*(2), 191-212.

Watkins, D. (1994). Student evaluations of university teaching: A cross-cultural perspective. *Research in Higher Education*, *35*(2), 251-266.

Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University Press.

Worthington, A. C. (2002). The impact of student perceptions and characteristics on teaching evaluations: a case study in finance education. *Assessment & Evaluation in Higher Education*, *27*(1), 49-64.

Appendices

Appendix A

Table III
Comparison Chart of SET constructs

SET Construct Category#	Description of SET Construct Category (Feldman, 1976; 2007)	SET constructs from Marsh (1984, 1987)	SET Constructs from Vort and Ban Engelburg (1992)	SET Constructs from Impact!	SET Constructs from NSE (2017)
1	Nature and Value of the Course Material (Including Usefulness and Relevance); Perceived Outcome or Impact of Instruction	Learning/ Value	Learning/Value/Validation	Cognitive Activation for Deep Learning	General skills acquired / Scientific skills acquired
2	Clarity of Course Objectives and Requirements	Organization/ Clarity	Instruction	Clear Instruction/ Adaptive Instruction	N/A
3	Teachers Stimulation of Interest in the Course and Its Subject Matter; Teachers Enthusiasm for the Subject or Teaching	Instructor Enthusiasm	Stimulation	N/A	N/A
4	Teacher Preparation; Organization of Course	Organization/ Clarity	Structure	Organization/ Classroom Management	N/A
5	Intellectual Challenge and Encouragement of Independent Thought (by the Teacher and the Course); Teachers Encouragement of Questions and Discussion, and Openness to Opinions of Others	Group Interaction	Interaction	Student-Teacher Interaction	N/A
6	Teachers Sensitivity to, and Concern with Class Level and Progress; Teachers Availability and Helpfulness, Teachers Concern and Respect for Students	Individual Rapport of Instructor	Conversation	Classroom Climate	Quality care/ Academic guidance and counselling
7 *excluded	Clarity and Knowledge of Subject Matter, and Intellectual Expansiveness	Breadth of Coverage	Explication	N/A	N/A

8	Teachers Fairness; Impartiality of Evaluation of Students; Quality of Examinations	Examinations/ Grading	Examinations	Assessment of Students	Testing and assessment
9 *excluded	Nature and Usefulness of Supplementary Materials and Teaching Aids	Assignments/ Readings	Literature	N/A	N/A
10 *excluded	Difficulty of the Course and Content (and Workload)	Workload/ Difficulty	Workload/ Difficulty	N/A	Study load
11	Other	General Satisfaction	General Satisfaction	General Satisfaction	General Satisfaction/ Experience gained through internships/ Group size/ Applied research/ Study facilities/ Pursuit of excellence/ Program schedules/ Preparation for a professional career/ Information provided on course program/ Role of course management in internships/ Internationalization

Appendix B

Table IIII
Criteria for the Impact! Tool in Higher Education

SET Construct Category#	Description of SET Construct Category (Feldman, 1976; 2007)	Constructs from Marsh (1984, 1987)	(1992)	Constructs from Vort and Ban Engelburg

Criteria #	Criteria Category	Description of Criteria	Guidelines for the Application of Criteria
1	**Topic**	Student Evaluation of Teaching (SET) in higher education	Items measure student perceptions of teacher quality in higher education
2	**Objective**	To help teachers identify the quality of their teaching practices in higher education	Items help teachers develop a plan to improve their weaker points of teaching, and maintain their stronger points of teaching
3	**Target audience**	Students and teachers in higher education	Items need to be answerable for students to evaluate teacher quality in higher education

4	**Target context**	Lectures at University, colleges, vocational schools, and other institutions in higher education	Items refer to teacher quality at lectures in higher education
5	**Assessment Type**	Formative assessment; used after learning to provide feedback on teacher quality and to support planning for the proceeding lecture	Items address the evaluation of one lecture
6	**Data Collection Instrument**	Questionnaire	Items can be implemented in a questionnaire
7	**Form of Medium**	Questionnaire is accessible using a digital application via a digital system	Items can be added into the Impact! tool for access using a digital application
8	**Psychometric Format**	1. Scale 2. Survey	Items have the following format: 1. 4-point Likert scale *includes 'not applicable' option 2. Open-ended questions
9	**Number of Items**	15-20	The prototype questionnaire has maximal 20 items, and the archetype questionnaire has maximal 15 items
10	**Point of View**	First person narrative using the 'teacher-oriented' approach in past tense	Students perspective of the teacher's behavior written in a passive voice; items begin with "The teacher..."
12	**Incorporation of Relevant Constructs**	Constructs that support teaching effectiveness	Items measure student's perceptions of teaching quality from the following instruments: 1. Student Evaluation of Educational Quality (SEEQ) from Marsh (1987) 2. Uvalon from Vorst & Can Engelenburg (1992) 3. Questionnaire for the Impact! tool in Secondary Education 4. Nationale Studenten Enquêt (NSE, 2017)
13	**Other**	Guidelines to consider when constructing the questionnaire	Guidelines to construct an effective questionnaire (Lee, 2006): 1. Write simple, clear, and short 2. Be specific and precise 3. Use appropriate language 4. Make them answerable 5. One topic per item 6. Avoid negative items 7. Avoid biased or loaded items 8. Use appropriate emphasis for keywords

Appendix C

Data Input from Focus Groups

Third Round:
Item 1: The teacher helped me learn valuable content during the lecture.
Item 1 was rated low in relevance. Participants thought that this item is too vague; the exact definition of the term 'valuable' was not clear to them. The participants questioned whether all students would have the same interpretation of the item.

Item 2: The teacher challenged me intellectually about the lecture content.
Item 2 was rated high in relevance. Participants thought that the phrase 'challenged me intellectually' should be reformulated. They provided several suggestions. Most of the participants agreed that the phrase 'think about' is a suitable alternative to measure the teaching theme of learning.

Item 3 (option 1): In the lecture, the teacher connected to what I already knew.
Item 3 (option 2): The teacher connected the lecture with content to what I already knew (prior knowledge).
Item 3 was rated high in relevance. Nonetheless, participants questioned whether it is necessary for every lecture. Participants thought that option 2 is formulated better; it aligns with the criteria listed in appendix B, which emphasizes that items should begin with 'the teacher'. One participant also suggested to address the teaching theme differently by replacing 'what i already knew' with 'prior knowledge'; but, this suggestion faced opposition, because participants doubted whether students are familiar with such a pedagogical concept.

Item 4: The teacher's lecture added value to the material to be studied for this lecture.
Item 4 was rated high in relevance. However, participants did not think that this item is always necessary, because students are not required to read prior to every lecture. It was agreed that this item should have a 'not applicable' option. Although most of the participants were content with the formulation of the item, some of the participants had difficulties comprehending the meaning of 'added value'.

Item 5 (option 1): The teacher presented the lecture content clearly.
Item 5 (option 2): The teacher explained the lecture content clearly.
Item 5 was rated high in relevance. The participants thought that option 1 and 2 of the item measure different constructs; thus, they cannot be replacements of one another. The participants came to a consensus that option 2 is more important, because it is more influence towards enhancing students learning and achievement.

Item 6: The teacher clearly stated the learning objectives of the lecture.
It was not clear whether item 6 was rated high in relevance. Several participants preferred to replace the term 'communicated' with 'stated'.

Item 7: The teacher presented content that aligned with the learning objectives of the lecture.

Item 7 was rated high in relevance. The participants agreed that it is crucial for a teacher to link all of the lecture content to the learning objectives. Participants again figured that if this item is included, then the 6. item can be excluded; this is because, the items implies that the teacher behavior in the 6. item is fulfilled.

Item 8: The teacher summarized what we had learned at the end of the lesson.
Item 8 was rated low in relevance. Participants again figured that if the 7. item is included, then this item can be excluded; because, the 7. item implies that the teacher behavior in this item is fulfilled. Also, in order for this item to align with the criteria in appendix B, the term 'lesson' should be replaced with 'lecture'.

Item 9 (option 1): If I did not understand the lecture content, the teacher made sure I understood it.
Item 9 (option 2): In case I did not understand it, the teacher clearly clarified the lecture content.
Item 9 was rated low in relevance. The participants did not think that the behavior addressed in this item is realistic for the teacher to fulfill. The participants explained that the teacher cannot address all of the student's individual needs during the lecture. Because of this, the subject in this item should also be changed from 'I' to 'we'. Participants again figured that if the 19. item is included, then this item can be excluded; because, the 19. item implies that the teacher behavior in this item is fulfilled.

Item 10 (option 1): The teacher made me more interest in the lecture content.
Item 10 (option 2): The teacher raised my interest about the lecture content.
Item 11 was rated high in relevance. Participants discussed that enhancing student achievement is the result of keeping a student's attention; thus, they agreed that option 1 is more relevant, because that is the construct that it measures.

Item 11: The teacher's lecture was well structured.
Item 11 was rated high in relevance. Participants thought that the students learning and achievement is dependent on how well a teacher structures their lecture. There were no remarks towards the formulation of this item.

Item 12: The teacher's lecture was at a good pace.
Item 12 was rated high in relevance.

Item 13: The teacher encouraged me to interact with other students about the lecture content during the lecture.
Item 13 was rated high in relevance. However, the participants questioned whether this item is relevant for all lectures; this is because, some class sizes and lecture content are not suitable for interaction. Instead of measuring the degree of interaction, the participants suggested two alternative constructs of measurement: engagement or active-processing.

Item 14: The teacher encouraged me to interact with him/her about the lecture content during the lecture.
Item 14 was rated high in relevance. However, some participants from focus group 1 agreed that if the 13. item is included in the questionnaire, then this item should be excluded; they are too similar, because they

both measure the degree of interaction. One other participant mentioned that it would be beneficial to include an alternative item that addresses whether a teacher took personal interest in their students. A suggested change towards the formulation of the item was the replace the phrase 'me to interact with him/her' with 'group interaction'.

Item 15 (option 1): The teacher created a good environment for learning.
Item 15 (option 2): The teacher created a comfortable environment for learning.
It was not clear whether this item 15 was rated high in relevance. Participants in focus group 1 were in favor of option 1. But, focus group 2 participants discussed that this item can be better formulated by describing 'environment' with an alternatively; one suggestion was 'open learning environment'.

Item 16: The teacher made me feel comfortable asking him/her questions.
Item 16 was rated high in relevance.

Item 17: The teacher gave meaningful answers when questions were asked about the lecture content.
Item 17 was rated high in relevance. Participants from focus group 2 thought that this item is a better alternative to the 9. Item. It was suggested that the item is formulated in such a way that it is shortened for easier comprehension: The teacher gave meaningful answers when questions were asked.

Item 18: The teacher spoke English clearly.
Item 18 was rated low in relevance. Participants from both focus groups agreed that this item is not relevant for formative feedback; instead, it should be a part of a course (summative) evaluation. There was also discussion amongst participants whether the term 'clearly' is too vague; an alternative suggestion was 'proficient'.

Item 19: During the lecture, the teacher checked if we understood the lecture content.
Item 19 was rated high in relevance.

Item 20: The teacher's lecture could be improved if he/she…
Item 20 was rated high in relevance. All participants thought that this item was necessary; because it allows for students to provide additional feedback towards a teaching theme that was not previously addressed. Additionally, it was suggested that the item is phrased more general; i.e. 'Points of improvement for teachers…'

Item 21: This is what I liked about the lecture…
Item 21 was rated high in relevance. There were no remarks towards improving the formulation of the item.

Appendix D

Data Input from Cognitive Interviews

Pre-Item: I studied all the reading material in preparation for the lecture.

Item 1 was clear to the majority of participants. However, several participants noticed that the terms 'studied' and 'reading material' do not relate in definition. Additional remarks related to the meaning of the terms in the item. Participants commented that 'reading material' does not have to be 'studied', but 'read'. Participant 7 suggested to reformulate 'reading material' to be 'material' or 'study material'. Participant 6 commented that there are cases where the 'reading material' is not necessary for the preparation of the lecture.

Item 2: The teacher asked questions that made me think about the lecture content.
Item 2 was clear to the majority of the participants. 5/7 participants comprehended that the item asked whether the teacher posed questions that resulted in deeper think. Participant 7 did not give a response towards the understanding of this item, because it was considered to be too vague to comprehend. Additional remarks related to the vagueness of the item. Participants 4 and 7 thought the item could address the specific type of 'lecture content'. Moreover, participants 3 suggested the following alternative item to prevent erroneous assumption or awkward syntax: 'The teacher asked me questions about the lecture content that helped me get a better understanding.'

Item 3: The lecture helped me understand the study material better.
Item 3 was not fully clear to the majority of the participants. 6/7 of the participants interpreted the item to address 'the teacher', and not the 'the lecture. This comprehension error indicates that the topic was carried over from a previous item. Additional remarks related to the syntax of the item. Participant 3 questioned why the item did not begin with 'the teacher', and participants 6 and suggested to change 'the lecture' to 'the teacher'.

Item 4: The teacher explained the lecture content clearly.
Item 4 was clear to the majority of the participants. Most participants did not rephrase the item when explaining what it meant to them. Participant 6 assumed that when a teacher explains the lecture content clearly, they have used examples, discussions and explanations. However, participant 7 assumed that when a teacher explains the lecture content clearly, they have not used any other mediums (i.e. PowerPoint presentations). Additional remarks related to the vagueness of the item. Participant 1 thought that the item is too broad, because it does not address which parts of the lecture the teacher explained clearly. Participant 3 thought that the item does not distinguish between 'the content used in the lecture' vs 'the content used outside of the lecture'. Furthermore, participant 6 commented that the items continuously addresses different types of subject matter (e.g. reading material, study material and lecture content); thus, it was suggested that the Impact! questionnaire only mentions one of these.

Item 5: The teacher clearly stated the learning objectives of the lecture.
Item 5 was clear to the majority of the participants. All participants understood that the teacher presented the learning objectives in the lecture. 4/7 of the participants assumed that the item refers to the learning objectives being at the beginning of the lecture. Additional remarks related to the undefined reference period of the item. Three participants commented that they preferred to have the item specify that the learning objectives are states 'at the beginning', while one participant commented otherwise.

Item 6: The teacher made me (more) interested in the lecture content.

Item 6 was clear to the majority of the participants. In order for their interest to increase in the lecture content, participants associated the teacher's behavior to be motivating, passionate, enthusiastic, and not boring. Additional remarks related to the syntax of the item. Participants 1 and 3 thought that the term 'more' caused the item to ask two different questions. Participants 5 and 7 preferred to exclude the term 'more'. Participant 4 suggested to take out the brackets around the term 'more'. Participants 5 was confused about the meaning of the term 'more'.

Item 7: The teacher's lecture was well structured.
Item 7 was clear to the majority of the participants. Additional remarks related to the vagueness and awkward syntax of the item. Participants 5 and 6 commented that the item it too broad; it should specify what part(s) of the lecture it is referring it. Furthermore, participant 7 suggested an alternative formulation of the item: 'The teachers lecture had a good structure'.

Item 8: The teacher's lecture was at a good pace.
Item 8 was not clear to the majority of the participants. Participants had three different interpretations of the term 'pace'. Some participants thought the term 'pace' related to the speed of the teachers speech. Participants thought the term 'pace' related to the speed of the teachers presentation. Participants thought the term 'pace' related to both the speed of the teachers speech and presentation. Additional remarks related to the clarity and syntax of the item. Participants were not certain of the definition of the term. Participants 3, 4 and 6 suggested to specify which type of 'speed' the item is focusing on. Furthermore, participant 7 suggested an alternative formulation of the item: 'The teachers lecture had a good pace'.

Item 9: The teacher made us actively process the lecture content (e.g. questions, discussions).
Item 9 was clear to the majority of the participants. Additional remarks related to the clarity and repetition of topic in the item. Participant 6 found the examples 'discussion, questions' necessary, because it helped explain what was being referred to in the question. Participant 4 thought it would be helpful to add another example, such as 'assignments' into the parentheses. However, participant 5 commented that the terms 'discussion' and 'questions' were a distraction from a previous phrase of the item; they noticed that 'actively process' became irrelevant when comprehending the question. Moreover, participant 4 thought that this item repeated the constructs that were included in item 2.

Item 10: During the lecture, the interaction between teacher and students was pleasant.
Item 10 was clear to the majority of the participants. All of the participants interpreted the item to ask whether the interaction between student and teacher was pleasant to social interaction. Additional remarks related to the vocabulary in the item. Participant 1 suggested to replace the term 'pleasant' with 'positive'.

Item 11: The teacher gave meaningful answers if questions were asked.
Item 11 was clear to the majority of the participants. All participants interpreted the item to ask about the quality of the response a teacher gave to questions. Additional remarks related to the vagueness of the term 'meaningful' in the item. Participant 5 and 7 commented that the term 'meaningful' should be explained. Participant 7 suggested to replace the term 'meaningful' with 'helpful'.

Item 12: The teacher checked if we understood the lecture content.

Item 12 was clear to the majority of the participants. All participants interpreted the item to ask about whether the teacher checked their understanding of the lecture content. Additional remarks related to the vagueness and complexity of the item. Participant 3 thought that 'lecture content' should be specifically explained or defined. Participant 5 questioned the validity of this item; they think it is too hard for the teacher to administer the students learning.

Item 13: I now understand the lecture content.
Item 13 was clear to the majority of the participants. All of the participants interpreted the item to ask whether they understand, and consequently recall the lecture content. However, in comparison to the other items, the participants seemed to take a longer time to interpret this one. Additional remarks related to the syntax and vagueness of the item. Participants 1, 2, and 5 questioned why 'the teacher' is not addressed in the item. Participant 7 suggested the following alternative formulation of the item: 'The teacher made me understand the lecture content better'. Participant 4 thought that the item was not specific enough, because it does not explain what type of material defines 'lecture content'.

Item 14: This is what I liked about the lecture…
Item 14 was clear to the majority of the participants. All of the participants interpreted the item to ask whether they can understand, and consequently recall the lecture content. Additional remarks related to the syntax of the item. Participant 5 suggested the following alternate formulation of the item: 'What I liked about the lecture…'

Item 15: This is what I still do not understand about the lecture content…
Item 15 was clear to the majority of the participants. All of the participants interpreted the item to ask about content that was not clear to them. Additional remarks related to the syntax and topic of the item. Participants 3, 5 and 7 gave the following alternative formulations of the item: 'this is what I do not understand about the lecture content…', 'what I still don't understand…', and 'there something I did not understand about the lecture'. Participant 6 thought that the item could be combined with number 13.

Item 16: The teacher's lecture could be improved if he/she…
Item 16 was clear to the majority of the participants. All of the participants interpreted the item to ask which part(s) of a lecture need improvement. Additional remarks related to the syntax of the item. Participant 3 thought that the terms 'he/she' should not be included in the item.

Appendix E

Data Input from Pilot

Table V lists the Cronbach's Alpha of the questionnaire from the individual pilot groups, items with the highest and lowest reliability from the individual pilot groups, and the Cronbach's Alpha if the item with the highest and lowest reliability is deleted from the individual pilot groups.

Table V
 Data Output from SPSS : Reliability Analysis

	N: # of Students responding to the items in the Impact! tool	Cronbach's Alpha of Questionnaire	Item with the highest reliability	Cronbach's Alpha if item with the highest reliability is deleted	Item with the lowest reliability	Cronbach's Alpha if item with the lowest reliability is deleted
Pilot 1	17	.527	9	.441	10	.637
Pilot 2	22	.839	9	.834	10	.849
Pilot 3	12	.806	3	.760	4	.841
Pilot 4	24	.803	12	.751	4	.822
Pilot 5	22	.841	3	.803	11	.860
Pilot 6	17	.615	11	.513	1	.663
Pilot 7	9	.940	2	.926	12	.951
Pilot 8	37	.604	3	.536	4	.624
Pilot 9	32	.577	4	.483	6	.662
Pilot 10	7	.849	3	.797	2	.898
Pilot 11	20	.908	6	.894	10	.913
Pilot 12	20	.800	1	.752	10	.840
Pilot 13	15	.693	12	.621	8	.738