Expert evaluation of an on-line course in Clinical Immunology

by
Dr Walter Liebrich

Thesis presented in partial fulfilment of the requirements for the degree of Masters of Philosophy in Health Sciences Education (MPhil HSE) in the Faculty of Medicine and Health Sciences at Stellenbosch University

Supervisor: Dr Jan Petrus Bosman

December 2014

Declaration

I, the undersigned, hereby declare that the work contained in this assignment is my original work and that I have not previously submitted it, in its entirety or in part, at any university for a degree.

Signature   Tygerberg   2014

ii

Abstract

This assignment describes an evaluation by experts of an on-line course in Clinical Immunology offered to medical registrars and scientists as a supplement to a practical rotation.

Because of a lack of agreement on what constitutes quality in e-learning and to avoid the customary focus on usability evaluation, an open-ended, interpretivist approach was used here which, while not entirely novel, was unusual in an e-learning environment.

For this project it was decided to evaluate both content (subject matter) as well as instructional value using two groups of peers from various academic institutions, clinical immunology experts and e-learning experts.

Feedback was obtained through participation in a focus group or in writing. Replies were much easier to obtain from the e-learning group. Five out of seven e-learning experts provided a response, versus three out of twenty subject matter experts. Eventually most of the feedback was obtained from colleagues from the home institution.

Both groups made valuable, somewhat overlapping suggestions. Subject matter experts indicated that the course materials were of good quality and adequate on a postgraduate level. E-learning experts expressed concern about the ability of the course to facilitate learning and identified also some usability issues.

Some of the findings may well apply to other settings. A number of five evaluators in each group appeared to give a good coverage within an open-ended approach. Expert peer review offered insights that neither student feedback nor self-reflection could. Rather than imposing evaluative criteria on the experts through the use of fixed checklists, the open-ended approach allowed them to cumulatively develop their own framework tailor-made for the course.

The choice of subject matter plus e-learning experts may be helpful in similar situations of evaluating on-line courses where dual expertise is not readily available. The open-ended interpretivist approach can be used for formative evaluation only and may work well for courses that are still in development or where an amount of uncertainty about teaching effectiveness exists.

Future efforts will likely focus on implementing the recommendations, identifying sustainable ways of quality review for the current course and similar open-ended evaluation of other courses.

Opsomming: Kundige evaluering van 'n aanlynkursus in Kliniese Immunologie

Die evaluering deur kundiges van 'n aanlyn-kursus in Kliniese Immunologie word in hierdie opdrag bespreek. Hierdie kursus word bykomend tot 'n praktiese rotasie vir kliniese assistente (medies) en wetenskaplikes aangebied.

Aangesien daar nie eenstemmigheid is oor wat gehalte in e-leer behels nie, en om die gebruiklike fokus op die evaluering van gebruiksmoontlikhede te vermy, is 'n interpreterende benadering in hierdie geval gebruik. Alhoewel hierdie benadering nie heeltemal nuut is nie, is die gebruik daarvan ongewoon in die e-leer-omgewing.

Daar is besluit om vakinhoud sowel as onderrigwaarde in hierdie projek te evalueer. Twee ewe-knie-groepe van verskillende akademiese inrigtings, kundiges in kliniese immunologie sowel as kundiges in e-leer is gebruik.

Terugvoer is ontvang deur die deelname aan fokusgroeponderhoude of deur skriftelike terugvoer. Terugvoer is makliker van die e-leergroep verkry. Vyf uit die sewe e-leerkundiges het gerespondeer teenoor drie uit die twintig vakkundiges. Uiteindelik is die meeste terugvoer verkry van kollegas van die tuisinstelling.

Beide groepe het waardevolle, maar dikwels oorvleuelende aanbevelings gemaak. Die vakkundiges het aangedui dat die kursusmateriaal van 'n goeie gehalte en geskik op 'n nagraadse vlak is. Die e-leerkundiges het hul kommer uitgespreek oor die vermoë van die kursus om leer te fasiliteer en het ook 'n aantal kwessies ten opsigte van bruikbaarheid uitgewys.

Sommige van die bevindinge kan moontlik ook in ander kontekste van toepassing wees. Dit het geblyk dat ongeveer vyf evalueerders in elke groep 'n goeie verslag met die oopvrae-benadering gegee het. Vakkundige ewe-kniebespreking het insigte opgelewer wat nie moontlik was met studente-terugvoer of selfrefleksie nie. In plaas daarvan dat evaluerende kriteria deur vaste vraelyste op die kundiges afgedwing is, het die oopvrae-benadering hulle die geleentheid gebied om kummulatief hul eie toepaslike raamwerk vir hierdie spesifieke kursus te ontwikkel.

Die keuse van vakkundiges en e-leerkundiges mag nuttig wees in soortgelyke situasies waar aanlyn-kursusse geëvalueer word en die tweeledige kundigheid nie geredelik beskikbaar is nie. Die oopvrae-interpreterende benadering kan slegs vir formatiewe evaluering gebruik word en mag moontlik goed werk vir kursusse wat nog ontwikkel word en waar daar heelwat onsekerheid oor die doeltreffendheid van die onderrig bestaan.

Verdere ontwikkeling sal waarskynlik fokus op die implementering van die aanbevelings, die identifisering van volhoubare maniere van gehalte-beoordeling vir die huidige kursus en soortgelyke oopvrae-evaluering van ander kursusse.

Acknowledgements

I would like to acknowledge the following people:

To start with, I would like to thank my family who had to bear with me during the three years of Masters studies.

Being part of the MPhil in Health Science education gave me a sense of belonging. I appreciated having supportive fellow Masters students around but also a group of knowledgeable tutors. Especially Dr Brenda Leibowitz deserves mention for letting me learn the ropes necessary for this study in her module 'Educational Research for Change.' Dr JP Bosman kindly agreed to be my supervisor.

Several colleagues provided valuable input for this project. Norma Kok volunteered as a moderator for the focus group meeting. Madelé Du Plessis cross-checked my coding. Martie van Heusden translated the abstract. Prof Martin Kidd provided some much-needed advice on the use of statistical methods.

Dr Monika Esser has supported me over the years building a portfolio in Clinical Immunology teaching which led to the development of a short-course in Clinical Immunology.

Although the project was unfunded, the Fund for Innovation and Research into Teaching and Learning (FIRLT) has supported previous research leading to current work, which is acknowledged.

Finally I would like to express my appreciation towards all the peer evaluators who have taken part in this study. You will know who you are.

Thank you to you all!

Index

Chapter 1: Orientation of the study

This chapter describes the setting for this study and provides a problem statement, motivation, assumptions, research question, aims, objectives, limitations and the envisaged contribution. For a more in-depth elaboration on most of the concepts touched on here, the reader is referred to specific sections of the thesis as indicated in the text.

1.1 Background

In 2005 a one-month practical immunology bench rotation was implemented in the Immunology Unit of the National Health Laboratory Service (NHLS) Tygerberg / Stellenbosch University Division of Medical Microbiology, aimed initially at Pathology Registrars. Despite a recent review of the curriculum, medical students still had little exposure to immunology during their undergraduate studies. On becoming Clinical Registrars, some expressed an anxiety to tackle complex immunology principles on their own. They approached the course co-ordinator and asked to bridge this perceived knowledge gap. This was addressed by introducing a voluntary, on-line Clinical Immunology self-study course to supplement the practical laboratory rotation. In order to implement this, considerable technical hurdles had to be overcome. The self-study course was then evaluated as part of a research project (Liebrich and Esser, 2014). Students' needs and perceptions were captured and feedback was obtained through a structured interview conducted by an independent interviewer before and after the course.

In the pre-interviews the students confirmed the impression of shortcomings of immunology teaching in undergraduate training and indicated willingness for self-directed learning on-line. In the post-interviews it emerged that, although students perceived the course as helpful, they did not feel that their applied clinical immunology knowledge had improved significantly, which commented on the need for more clinical applicability. It was noticed on tracking that almost half the students did not make use of the course, interpreted as lack of motivation.

Based on these findings, the course was redesigned. Clinical cases and pointers to clinical applications were included in the chapters. All externally copyrighted content was removed and course materials, 16 hyperlinked pdf files altogether, were now freely downloadable and usable off-line as well to work around connectivity issues. Online tests were introduced on the learning management system, including feedback by the course co-ordinator. The course was also given a more formal standing and credit by converting it into a certified short-course, to provide more incentive to partake and complete (Liebrich and Esser, 2014; http://shortcourses.sun.ac.za/ courses/3615.html). The student throughput is still small, with only about five students in total each year, which means there is rarely more than one learner active in the course and learners can therefore interact directly with each other.

While student feedback and self-reflection by the course co-ordinators proved helpful in re-designing the course, one aspect of course assessment had not been explored yet: an evaluation by independent experts. This will be the focus of the current study.

1.2. Problem statement

Expert evaluation of courses is a well-established theme in the literature within a traditional classroom setting (see 2.1). However, the immunology course to be investigated here is not a traditional face-to-face course but an assisted on-line self-study course, i.e. an e-learning course. Does this have implications for evaluation? The particularities of the learning process using instructional technologies [see 2.2] may indeed demand additional approaches and expertise necessary for evaluation that need to be considered. However, broader quality standards which would allow evaluators to look at both processes and course content are not widely adopted and there is astonishingly little agreement on what elements constitute a 'good' course (see 2.4).

Evaluation and quality assessment in e-learning often emphasizes user-centricity and usability which then becomes the centre of expert and user review. Additionally, usability inspection often follows pre-determined and well-defined standards (see 2.3). While this approach may work well for a systematic appraisal of a range of courses, it misses the opportunity to gather expert opinions in a more open-ended manner or to encourage novel ideas and suggestions from the expert panel.

Also from a philosophical point of view, a positivistic perception of quality or usability, essentially a compliance with a pre-determined set of parameters, may be challenged. In the alternative interpretivist approach, quality or usability standards may be defined by a group of experts (or users) through an exchange of thoughts and agreement on statements (see 2.5 and 1.4 below). Quality in a constructivist view may be determined by its usefulness in an experimental setting. In these more open-ended approaches, quality is not something pre-defined, it is something which is created each time in potentially novel ways.

Finally, a focus on usability alone neglects the subject matter. In order to evaluate the complete educational offering of courses, it is important to assess content as well, similar to traditional teaching.

1.3. Motivation

The current project aims to address the shortcomings of focussing expert evaluation narrowly on usability. Both the subject matter as well as the way content is presented in an online course will be evaluated. The aforementioned lack of agreement of what constitutes a quality e-learning course and the drawbacks of a purely positivist approach and limited usability inspection opens up opportunities to explore more non-traditional ways of course evaluation. Rather than pre-determining and attempting to define what quality is, one can now ask the experts to approach the topic in a more open-ended fashion. By giving them as little guidance as possible, they may come up with novel suggestions rather than merely

detecting technical oversights by the course designer. Interpretive research has recently become more accepted in e-learning ('Interpretive information system research', see 2.5).This kind of enquiry requires embracing research models originating from the social sciences.

1.4. Assumptions

The approach selected here is an interpretivist / constructivist one (see 2.5). It is assumed that there are few agreed-on principles that dictate what makes an e-learning course 'good' and that excellence in teaching is very much open to interpretation and is highly situated within particular contexts. Some of these principles may emerge from the expert feedback. A positivist line of reasoning is not taken. Qualitative research methods are suggested to approach the question of expert feedback. Furthermore, an inductive rather than an a priori approach is selected for data analysis.

1.5. Research question

Will an open-ended, interpretivist approach to expert evaluation provide suggestions for improvement of both content and instructional practises of a selected e-learning course?

1.6. Aim

Implement an open-ended approach of expert evaluation in the context of an e-leaning course in clinical immunology with the aim of firstly improving the specific educational offering, but secondly also making more general suggestions for course evaluation in related contexts

1.7. Objectives

- Identify suitable experts within both e-learning and subject (clinical immunology) contexts (see 2.6 - 2.9).
- Identify and implements suitable means for obtaining formative expert feedback (see 2.10).
- Obtain expert opinions (see 4.3 – 4.5).
- Analyse opinions and make suggestions for course improvement (see 5.2.2).
- Based on the findings make suggestions how expert feedback may contribute in other settings (see 5.3).

1.8. Limitations

The choice of suggested feedback methods was not solely guided by scientific principles but also by personal preferences. Gathering expert's opinions by independent interviewers for example may have yielded equally valid results. Other methods were not considered here partly because of cost considerations (no funding available to pay for independent interviewers and transcriptions), time constraints and because of concerns regarding sustainability beyond this research project. The study was

also limited by using only one data gathering method per expert group due to time and financial constraints.

1.9. Envisaged contribution

The study will be partly descriptive relating the situation within a specific context. Nonetheless it is envisaged that more general proposals and theories can be derived from the observations (see 5.3.4). Although not entirely novel, suggestions on the use of open-ended formative course feedback and expert assessment of e-learning courses may contribute to the pool of knowledge even outside the contexts of e-learning or a specific subject (immunology in this case).

Any experiences, recommendations or quality assurance guidelines made here may be particularly important from the point of view that this is a much-needed contribution to local African solutions on the use of information technology in medical education (Greysen, Dovlo, Olapade-Olaopa, Jacobs, Sewankambo, and Mullan, 2011).

1.10. Summary of chapter 1 and delineation of the thesis

The current study envisages an open-ended, interpretivist approach to an expert evaluation of an online course in Clinical Immunology. This chapter briefly presented the setting and provided a justification and a brief outline for the research approach chosen.

Chapter 2, the literature review, offers a more in-depth view on the concepts touched on here. Chapter 3, methods, describes how the research was conducted. Chapter 4, results, describes expert characteristics and behaviour and the feedback that was provided. Chapter 5, the discussion, interprets the results, investigates their validity and explores implications for the current course but also the broader applicability in other settings based on the findings here as well as on published literature. Chapter 6, conclusions, briefly summarises major points and explores possible future steps.

Chapter 2: Literature review

Chapter 2 provides a literature review on topics pertinent to this study, which were briefly touched on in chapter 1 above. More concepts will be explored in the discussion, chapter 5.

2.1. What do we know about evaluation in general and expert / peer evaluation in particular?

In a non-educational context, Newcomer, Hatry and Wholey (2004) define programme evaluation as a "systematic assessment of program results and … systematic assessment of the extent to which the program caused those results." Patton (2002) compares the terms programme evaluation and quality assurance. While quality assurance looks at individual processes and uses professional-based judgement intended for staff involved, programme evaluation typically focusses more on programme processes and uses goals-based judgement intended for decision makers. He points out though that this distinction of terms has lost much of its importance as both functions have expanded and overlap (Patton, 2002). The difference between evaluation and evaluative research will be discussed in 5.3.1.

Let us start with the terms assessment and evaluation. Newble and Cannon (1994) make a clear distinction between 'assessment' and 'evaluation' (not all authors do) and they claim that assessment is primarily concerned with the measurement of student performance whereas evaluation is generally understood to refer to the process of obtaining information for subsequent judgement and decision-making. Mehrens (1991, cited in Goldie, 2006) identifies two main purposes of course assessment, the evaluation of teaching methods and the evaluation of teaching effectiveness. Worthen, Sanders and Fitzpatrick (1997, cited in Goldie, 2006) distinguish six possible project evaluation approaches which include objectives-oriented approaches, participant-oriented approaches and expertise-oriented approaches, but also management-, consumer-, and adversary-orientated approaches. The first three, in particular, are addressed by other authors as well to varying degrees.

Some, but by no means all authors allocate the terms evaluation and assessment to formative versus summative approaches. For example York University's Senate Committee on Teaching and Learning (2002) distinguishes between formative assessment and summative evaluation. They suggest strategies for both which include student ratings and peer observation. They also stress the importance of keeping formative and summative approaches strictly apart. Tuckman (1999) on the other hand distinguishes between formative and summative evaluation. In his view summative evaluation is external, highly structured and accomplished by comparing performance outcomes of students who have experienced a programme to those of students who have experienced an alternative (or no) programme. Formative evaluation (not assessment) is internal and accomplished by comparing student performance outcomes to the stated objectives of the programme. According to Patton (2002), formative evaluation has the purpose of improving an intervention, policy, or programme; it focusses on strengths and weaknesses; its desired result is to make recommendations for improvement. More on formative vs.

summative approaches below (see 5.3.3.). Overall it appears that a distinction between the terms assessment and evaluation is not clearly made by all.

There are also suggestions from the field of curriculum analysis. This arena was dominated for a long time by the ideas of Ralph Tyler (in Posner, 2004 and Grant, 2006) who suggested in 1949 a 'framework' for curriculum analysis which analyses which purposes a programme has and determining whether these purposes have been attained. More recently David Kern (Kern, Thomas, Howard and Bass, 1998) assumed that educational programmes have aims or goals, whether stated or not. He feels strongly that medical educators have an ethical obligation to meet the needs of their learners, patients, and society and he builds a logical, systematic 6-step guide to curriculum development to achieve these goals. Most concepts between course/ curriculum development and evaluation overlap. For example Muraskin (1997, cited in Goldie, 2006) states the reasons for evaluation as determining effectiveness of programmes for participants, documenting that programme objectives have been met, providing useful information about service delivery and enabling staff to make changes.

What sources can be used to appraise teaching? Berk (2005) suggests that evidence for the conceptualization of teaching effectiveness should be collected from a variety of sources, which include student ratings, peer ratings, and self-rating amongst others. Harden and Crosby (2000) state that the quality of teaching needs to be assessed through student feedback, peer evaluation and by assessing the actual 'product.' Felder and Brent (2002) and Brent and Felder (2004) propose a model for the evaluation of 'traditional' teaching, which is based on three components: learners rate, peers rate, and the instructor self-rates. Peers in this context are fellow instructors. Also the University of Michigan's Center for Research on Learning and Teaching (2014) suggests using students, colleagues and self-reflection as sources for evaluation. Van Ort, Noyes and Longman (1986, cited in Brown and Ward-Griffin, 1994) suggest three different evaluation components which involve independent observers inspecting course related materials, observing classroom teaching and evaluating student performance.

Peer-review provides one source of evidence to measure teaching performance. Van Ort et al. (1986) see peer review primarily as an institutionalised and structured process which serves to provide either summative feedback for purposes of tenure, promotion etc. or formative feedback to improve the quality of teaching of the instructor being evaluated. A less stringent method of peer evaluation is peer observation of teaching (POT) (Swinglehurst, Russell and Greenhalgh, 2008). It usually involves a fellow educator observing the teaching of another in order to provide constructive feedback. POT may be informal or well-structured. The check-lists presented in Bell (2005) may serve as just one example for a structured approach. POT has been used successfully in both traditional and e-learning environments. Peer-to-peer reflection in an e-learning context may involve course design, materials, online-interactions etc. (Jara, Mohamad and Cranmer, 2008).

When compared to student rating which is a well-known and often-used approach to course assessment, peer evaluation has been much less dominant historically (Berk, 2005). Some warn that

students may not be qualified enough to supply valuable feedback. Felder and Brent (2002) for example say that it makes little sense to only use student ratings as few students are well-equipped to judge. Berk (2005) summarises his paper by saying that peer ratings of teaching performance and course materials is the most complementary source of evidence to student rating and that both should be used in conjunction.

But do experts become involved in the evaluation process? Averch (2004) states that expert evaluation is very common in evaluating higher education programs. Broadly, these experts can come from inside an agency or from outside.

If expert evaluation comes from the inside of an agency, judgement is obtained here from those close to the programme. The alternative is outside expert peer review. Expert review may even entail the recruitment of external professional evaluation agencies. Very often, however, there is no need to employ external experts. Experts are in plentiful supply in any academic environment. The expert then becomes a peer and expert review becomes peer review. For this reason the terms expert review and peer review are often used interchangeably in this paper.

A note of caution. Peer review is not a well-defined term. It is often understood as scientific peer review used by scientific journals. It may also allude to the peer review process in approving scientific projects and allocating funding. Peers may also be used to assess professional (example clinical) performance for various purposes. Many literature searches provide hits on peer review in the context of students assessing each other. Peer review here is understood as a means of assessing teaching performance by fellow educators.

2.2. How is e-learning is different from traditional learning?

The course to be evaluated here is offered on-line. The following paragraphs will have to address the question whether the concepts of quality are similar or different between traditional and e-learning teaching approaches and whether similar of different ways of evaluation or assessment are required.

According to Jung (2011) there appear to be at least some who argue that while certain principals of quality apply to both conventional and e-learning there are some features of e-learning which should be addressed in addition. It was already pointed out in the introduction that e-learning has both a content and a process dimension (Ellaway and Masters, 2008), processes being essentially learner interactions within the learning platform. Van der Westhuizen (2003) suggests that these processes can be facilitated by the e-teacher with varying degrees of instructional effectiveness. He also adds that the web has unique technological characteristics which impact on learning, in other words it's 'affordance' (for a discussion of 'affordance' in e-learning refer to Bower, 2008). Anderson (2004) claims that e-learning affords an increase in communication and interaction capability and that this is achieved by using numerous modalities. Siemens and Tittenberger (2009) specify that e-learning makes possible the use of a range of new media. They also allude to the potential of e-learning for delivery of education. Likewise Greysen et

al. (2011) highlight the promise of increased access to high-quality education which e-learning enables but also point out the danger of possible failure. In short, e-learning opens up new opportunities and ways of teaching which are not possible or not generally used in traditional teaching and learning, but the appropriate and effective use of technology need to be evaluated.

2.3. How are e-learning courses evaluated?

One may argue that the same principles of and approaches to course evaluation apply to both e-learning and traditional teaching (Jung, 2011). When taking this view all that has been said about evaluation (see 2.1) would apply to e-learning as well, at the very least in the areas where e-learning and traditional teaching overlap. There are some in information systems research who embrace such a broad approach to e-learning evaluation. For example De Villiers (2005) recommends accepting research models originating from the social sciences.

However, usability is often seen as the major quality factor in e-learning (Davids, Chikte and Halperin, 2011; Fernandez, Insfran and Abrahão, 2011). Fernandez et al. (2011) state that usability evaluation is a procedure which is composed of a set of well-defined activities for collecting usage data related to end-user interaction with a software product. Usability evaluation methods are typically divided into empirical and non-empirical (usability inspection) approaches (Davids et al., 2011, Fernandez et al., 2011, Recker, 2005). Empirical user testing involves representative end users such as students (typically non-experts) whereas usability inspection involves experts evaluating the application employing techniques such as heuristic evaluation or walkthroughs. It is interesting to note at this point the similarities to student feedback and peer evaluation introduced above (see 2.1). While e-learning evaluations could be conducted using a variety of approaches, the field has been dominated by Jakob Nielsen. Nielsen (1992) suggest an easy-to-use, non-expensive but narrow system of 'heuristic' usability evaluation in which evaluators (experts) use a set of pre-defined metrics or design principles (heuristics) to evaluate a system (Ssemugabi, 2006). Such usability testing is generally a structured approach in which evaluators are given detailed checklists which they follow to rate a course or an application (Brooke, 1996).

2.4. Can quality of teaching in e-learning be defined?

Section 2.1 elaborated on how the quality of teaching may be evaluated and it was concluded that no one method of evaluation will provide a complete picture of teaching effectiveness and that the use of multiple sources are recommended (Brown and Ward-Griffin, 1994). Numerous sources suggest compiling a teaching portfolio (for example Senate Committee on Teaching and Learning, 2002, Berk, 2005). Also Stellenbosch University's learning and teaching policy (Stellenbosch University, 2012) proposes using a variety of information sources and evidence in order to evaluate teaching performance. They are referring to 'performance indicators' and make mention of a task team that will be established

to define good teaching and provide methods to assess good teaching. Other Universities have made progress towards that goal. York University's Senate Committee on Teaching and Learning (2002) for example has defined 'quality teaching' and gives indicators for this: "effective choice of materials; organization of subject matter and course; effective communication skills; knowledge of and enthusiasm for the subject matter and teaching; availability to students; and responsiveness to student concerns and opinions."

Institutions of higher education also have to operate within quality standards set by national regulators. In South Africa the Council of Higher Education (CHE, Higher Education Quality Committee, 2004) provides a general framework of quality assurance and course review which merges elements of assessment and evaluation: user surveys of academics involved, benchmarking against national and international reference points, student throughput and retention, impact (employability of students, addressing shortages etc.), and regular evaluation for the purpose of developing improvement plans. In the medical field, the World Federation for Medical Education (WMFE, 2003) has set well-recognised standards of quality, which are similar to the CHE guidelines, but also include governance and administration, educational resources and quite broadly mission and objectives. As mechanisms they suggest institutional self-evaluation, external peer review, or a combination of the two. Discussing these processes of quality assurance would lead too far here. What is important is that both institutional and national bodies provide frameworks for quality assurance in education, when it comes to a working definition of 'quality teaching' they are however vague. The reason for this may be that there is no one good definition for quality education. A selective bullet list of possible contributing factors is given below.

Education is much more than classroom teaching

This is true for all teaching, but particularly prominent for medical teaching and the function of the medical teacher. According to Frenk, Chen and Bhutta et al. (2010) a medical faculty member should be a teacher, steward, agent of knowledge transmission, and importantly a role model for students. Also Harden and Crosby (2000) identify numerous roles for the medical teacher: lecturer (clinical or practical teacher), role model (both on the job and as teacher), facilitator (student learning, mentorship), assessor (of student and curriculum), planner (of curricula and courses – this includes use of technology), and resource developer (teaching materials – including using technology, study guides). A good teacher does not need to be competent in all these roles, but all of them need to be covered within an institution / faculty. Excellence has to be defined and understood within all these different contexts.

Since role modelling and mentorship are not stated aims in the clinical immunology course to be evaluated this point will not be elaborated further here but may be crucially important for some courses (including e-learning courses) within a medical faculty.

Education operates at various levels and involves numerous stakeholders

So far traditional teaching and e-learning were treated as unified concepts. However teaching and learning have various dimensions depending on which level they operate and both scope and stakeholders vary widely. This will now be elaborated on in the context of e-learning.

Williams and Graham (2010) distinguish between institutional, programme, course and lastly activity levels. Scope, stakeholders, subjects of evaluation and evaluation criteria differ between those levels. According to them, on an institutional level the primary stakeholders are administrators. What needs to be evaluated are e-learning initiatives, the totality of on-line course offerings, and e-learning policies. The criteria for evaluation may include cost effectiveness, number of enrolments, completion rates and user satisfaction. On an institutional or faculty level there is ideally a whole e-learning team with various sub-experts such as instructional designers, graphic artists, programmers, media specialists (audio/video), subject matter experts and usability specialists (Siemens and Tittenberger, 2009). Chua and Lam (2007) suggest a quality assurance process that relates to five main areas: content authoring, courseware development, adjunct faculty recruitment, pedagogy and delivery. Evaluation of institutional programmes needs to address all these aspects and evaluators with different types of expertise would may need to be called upon. Stellenbosch University (ICT Task Team, 2013) has a strategy for the use of ICT in learning and teaching which aims to describe and evaluate the impact of ICT-enhance learning and teaching and suggests indicators on programme and institutional level.

On a course level the primary stakeholders are instructors and learners (Williams and Graham, 2010; Jung, 2011). What needs to be evaluated are the online courses being offered. Example criteria may include student satisfaction, learning and engagement, student access as well as specific resources and technical requirements. The staff involved on this level are typically instructors (subject matter experts) working within an institutional e-learning support environment or dual subject matter / e-learning experts.

There is a raft of literature on various aspects of e-learning to be found and some is included in this thesis. Although often not explicitly stated, much of the literature is aimed at an institutional or programme level and there is often little to be found that is helpful to an instructor on how to specifically design and evaluate a good e-learning course. Many articles and guidelines suggest various frameworks for quality e-learning education. One may attempt to compile these sources and search for common themes and develop from these a framework that would work on a particular level and in a particular setting. However, is there much agreement in the literature?

Lack of agreement in the literature

There are a number of publications which warn that there may be little agreement on quality standards in e-learning. For example Anderson and McCormick (2006) contend that there are many views on what constitutes quality e-learning. Also Pawlowski (2003) states that the quality of e-learning is not a well-defined measure. Kidney, Cummings and Boehm (2007) warn that quality in e-learning is an elusive

concept and that attributes of quality differ between learners, faculty and administration. Jung (2011) acknowledges that there is general agreement on several quality dimensions but continues to say that quality is often defined from the perspective of e-learning providers. According to him "quality is a relative and value-laden concept and may be viewed differently by various stakeholders" and "e-learning quality is a complex and multi-faceted issue." It was this lack of agreement that prompted me to adopt an interpretivist, constructivist view of quality in e-learning (see also 1.3 and 2.5).

2.5. What is an interpretivist approach and has it been used in an e-learning setting?

It has been argued above that quality in teaching, including e-learning, may not be understood in absolute terms and that different settings and different stakeholders would lead to different interpretations on what constitutes quality. The philosophical underpinning for this kind of thinking is found in a school of thought called 'interpertivism.' According to Bunniss and Kelly (2010) reality in an interpretivist view (contrasted by a positivist view) is subjective and changing. There is in fact no one ultimate truth. Taylor and White (2000) also talk about the standpoint that reality cannot be accessed in a neutral way and that humans continuously re-interpret it, a view which they call social constructionism, and its proponents would be called relativists.

This school of thought has also entered information systems research. Recker (2005) discusses this in the context of how quality is perceived by positivists and interpretivists and he suggests that in a positivist view quality is determined through its compliance with a knowable reality whereas in the interpretivist perceives quality as subject and purpose oriented. This reality is agreed on within a community. De Villiers (2005) states that interpretive research has become better recognised in informatics and she uses the term 'Interpretive information system research' for this type of research. She recommends embracing research models originating from the social sciences in information systems research and she feels that Interpretivism lends itself to such qualitative types of studies. Maree and van der Westhuizen (2007) concur and say that quantitative research tends to be linked with positivism whereas qualitative research tends to be associated with interpretivism.

2.6. What characterises experts and how can they be identified?

The Merriam-Webster online dictionary (www.merriam-webster.com/dictionary/expert) defines an expert as "one with the special skill or knowledge representing mastery of a particular subject." What is the expert's contribution in the evaluation process? According to Patton (2002) an expert or 'connoisseur' brings his perceptions and expertise to the evaluation process drawing on his or her own judgments about what constitutes excellence. Also Worthen et al. (1997) state that expertise-oriented approaches depend on the direct application of professional expertise and the provision of professional judgements of quality. They discuss possible benefits such as ease of implementation as well as limitations such as vulnerability to personal bias, overuse of intuition and possible conflicts of interest. The real

11

contribution of expert evaluation is the possibility of emergent evaluation designs, an openness to evolve an evaluation plan, and the recognition of multiple realities. According to Averch (2004) procedures that force a wide range of participants to provide their reasoning and assumptions about a program turn out to be superior for decision making compared to narrow, pre-specified, tightly centrally controlled procedures.

The next question then is how to identify experts and how to compose a group of experts for the purpose of evaluation (Averch, 2004). Experts can be found based on their 'reputation' (desired expert skills, qualifications; publication record, citations etc.). Worthen et al. (1997) suggest the use of 'recognised standards' pertaining to the qualifications of 'experts.' Averch (2004) however cautions that some desired expert skills may leave no trace in any published record. He also proposes that initially identified experts suggest further experts (snowball selection). He advocates a mixed group of experts, which should include more than technical, substantive experts and might also include general-purpose policy analysts, philosophers of evaluation, or stakeholders. Again according to Averch (2004), experts should be coherent, reliable, and have resolution. A coherent expert is one who follows dictates of logic and probability, i.e. he is rational. A reliable expert is one who gives consistent and trusted feedback, i.e. he conforms with himself (longitudinally) and with other experts (i.e. he hasn't got views that nobody else agrees on). Averch (2004) explains the term 'resolution' using the example of a weather forecaster who not only predicts the weather in a logical and consistent way, but also predicts it correctly, i.e. his forecasts become true. It is of course hard to predict whether any selected expert is going to be coherent, reliable, and will have resolution, unless there is also a track-record in an evaluative setting (example: an outside consultancy agency with qualified evaluators is used).

2.7. What kind of expertise is needed to evaluate an e-learning course?

What kind of experts should be considered for course review and what kind of expertise should they have? In the context of peer evaluation of 'traditional' teaching Brown and Ward-Griffin (1994) give commonly accepted criteria of what a peer is, namely one having "knowledge and expertise in the subject matter, accessibility to the setting and shared clinical specialty." Both Brent and Felder (2004) as well as Schultz and Latif (2006) contend that fellow faculty members (not necessarily from the same speciality) could be used as raters, but this may require special training for this purpose or even the formation of a peer review committee. But peer review should surely move also beyond subject matter and content. Berk (2005) argues that course review should have two arms, the first being peer review of documents used in a course. The second is peer observation of in-class teaching performance and Schultz and Latif (2006) describe suggestions being made that raters should also have expertise in adult learning and curricular design.

Swinglehurst et al. (2008) describe peer observation of teaching in an e-learning environment and they make reference to both technical and pedagogic expertise as a requirement. Within her 'interpretive

information system' approach to evaluation, De Villiers (2007) advocates a team of evaluators which should have both subject matter and usability expertise. In a slightly different context Biswas, Basu and Chowdhury (2013) suggest content and computer interaction experts during course development and technical experts during the course delivery phase. Similarly Chua and Lam (2007) support content peer review in the area of content authoring and supervision and mentoring by senior faculty staff during course delivery. For performing 'heuristic evaluation' Jakob Nielsen suggests usability specialists and double experts (i.e. those who also have expertise in the specific interface being evaluated). He concludes that usability specialists are better than non-specialists and that 'double experts' perform the best (Nielsen, 1992, Ssemugabi, 2006). As a result of his work, 'usability' experts are now most commonly used to perform usability evaluation of e-learning programmes.

Based on the literature cited above, a good case can be made that experts evaluating an e-learning course should have experience in both the subject matter as well as technical expertise, i.e. they should be dual experts. Because it is hard to come by possible experts meeting these criteria it was decided here to use two sets of experts, one group with knowledge of the subject matter, the other with experience in e-learning.

2.8. What is an e-learning specialist?

It was suggested above to include 'e-learning' experts for the evaluation of an on-line short-course in Clinical Immunology. But what is e-learning and how does one obtain a professional qualification in e-learning?

A definition was suggested by Tavangarian, Leypold, Nölting and Röser (2004): "E-learning refers to the use of electronic media and information and communication technologies (ICT) in education. E-learning is broadly inclusive of all forms of educational technology in learning and teaching" and Ellaway and Masters (2008) suggest similarly "e-learning is not a single technology or technique. It is a loosely defined amalgam of information communication technologies (ICTs) used in education, usually but not exclusively mediated in some way through the Internet." However, the terminology is not all that clear. Ally (2004) warns that it is "difficult to develop a generic definition. Terms that are commonly used include e-learning, Internet learning, distributed learning, networked learning, tele-learning, virtual learning, computer-assisted learning, Web-based learning, and distance learning."

Of course qualifications for ICT exist, including in South Africa, and these will not be discussed here. Professionals with an ICT background can be found in IT divisions all over the country. However e-learning is more than just ICT and involves the use of technology in learning and teaching. This is where career paths become much less distinct. Tertiary qualifications in e-learning do exist overseas. For example in the UK there is an MSc in Digital Education (formerly the MSc in E-learning) (http://online.education.ed.ac.uk/). Even in South Africa the University of KwaZulu Natal offers a degree

in Medical Informatics (http://is.ukzn.ac.za/Courses/medicalinformatics.aspx). However, many educators using technology in teaching and learning do not have a formal background (qualification) in both.

For the purpose of this study an e-learning expert is defined as someone who merges teaching and learning and the use of technology in a professional educational environment. This point will be further elaborated on in the methods and discussion sections.

2.9. What is a Clinical Immunologist?

For the purpose of this study it needs to be understood what 'Clinical Immunology' as well as what a 'Clinical Immunologist' is. Armed with a working definition of the latter one might then continue to identify suitable experts in this field.

The 'clinical practice of immunology' is defined by the World Health Organization (WHO) (Lambert, Metzger and Myamoto, 1993) as encompassing "the clinical and laboratory activity dealing with the study, diagnosis and management of patients with diseases resulting from disordered immunological mechanisms and conditions in which immunological manipulations form an important part of the therapy." Much less clear is what a 'Clinical Immunologist' might be. In fact in many countries, including South Africa, there is no medical speciality or sub-speciality with that name. Shearer (2002) laments that, except for rheumatologists, all other clinical immunologists appear to lack organized training programs, defined certification pathways, and clear career opportunities. For the United States of America where such a career path exists, Bloch (1994) describes the formal requirements for certification in this discipline.

Immunologists may also have a background in science. The British Society of Immunology (www.immunology.org) describes immunologists as clinicians OR scientists who specialise in the field of Immunology. Similar to clinicians most countries don't offer science degrees in 'Clinical Immunology.' In South Africa both scientists and clinicians register with the Health Professions Council of South Africa (HPCSA). Scientists may do so as 'Medical Biological Scientists' and a sub-category 'Immunology' exists for them (Medical and dental professions board committee for medical science, 2010). However, there is no strict requirement for all medical scientists to register with the HPCSA.

In summary, 'Clinical Immunologists' may be either scientists or clinicians working in the field of Clinical Immunology. There are no clear career paths for either in South Africa. They are more defined by the type of work they do and they may find employment within various clinical or laboratory disciplines.

2.10. What are suitable means for expert feedback in the current study?

A constructivist/ interpretivist paradigm was suggested for the current study. This requires a departure from the more positivist-inspired structured and pre-determined checklists to more open-ended methods used in the social sciences in order to more broadly explore the opinions of experts. But which method should be chosen?

14

A starting point is a reflection on the research approach which has been adopted. Ringsted, Hodges and Scherpbier (2011) broadly distinguish four categories of research in medical education: experimental, explorative, observational, and translational studies. Using Ringsted et al.'s (2011) criteria the study suggested here is best described as explorative – aimed at modelling. Modelling in this suggested study is the exploration of an open-ended approach to attain the opinions of experts on an e-learning course. Methods used in explorative studies are typically qualitative research methods (Ringstead et al., 2011). Qualitative research methods include questionnaires, interviews, or observation (Nieuwenhuis, 2007a). In a social science setting, Denscombe (2010) lists questionnaires, interviews, observation and document research irrespective of the research strategy. There are further investigative methods which are used in the particular context of evaluative research which include ratings by trained observers, surveys, role playing, focus groups, fieldwork based on semi-structured interviews, and agency records (Newcomer et al., 2004). According to Broom and Willis (2007) methods used within an interpretivist / constructivist paradigm such as the one embraced in this study include interviews, participatory or non-participatory observation, focus groups and secondary discourse analysis.

Averch (2004) describes various alternatives for obtaining judgements from experts. These may be collected individually and aggregated afterwards or they may be collected collectively. He also distinguishes structured / unstructured as well as direct (face-to-face) / indirect modes of interaction. The Department of Sustainability and Environment (2005) also suggests tools within the broader context of stakeholder engagement such as brainstorming sessions, Delphi studies etc. which may also have merit in the context of exploring expert judgement.

Out of these possible approaches an open questionnaire-type written email feedback and focus groups were considered for the current study, mostly for practical reasons.

Open questionnaires

Denscombe (2010) explains that a survey is a research strategy, not a method. He lists evaluation of educational courses and new innovations as one of the potential uses. Cross-sectional surveys provide a snapshot of a sample population in time whereas longitudinal surveys collect data at different points in a study in order to observe changes over time (Fraenkel and Wallen, 2009). According to Maree and Pietersen (2007) and Denscombe (2010) surveys collect information about, amongst others, attitudes, ideas, feelings, opinions and perceptions. Information can be obtained in a variety of ways, including email. Surveys tend to be aimed at large audiences and often use questionnaires. According to Sivo and Saunders (2006) questionnaires are also popular with information systems researchers. Preece, Rogers and Sharp (2002) suggest interviews and questionnaires for user feedback. Also the checklists used in usability evaluation (above) may also be classified as questionnaires. Questionnaires typically consist of instructions and a written list of questions (open or closed) (Maree and Pietersen, 2007). Few authors (Witteck, Most, Kienast and Eilks, 2007) describe using an open questionnaire which does not give particular directions for responding.

For the subject matter experts an unstructured written feedback by mail was therefore considered, which would provide only some minimal instruction and guidance to the evaluator but would otherwise be completely blank ('open') (see also 3.7) to allow completely unguided feedback by respondents free from pre-determined questions set by the investigator.

Focus groups

Focus groups appear particularly suited in combination with expert judgement because they can elicit detailed, introspective responses on participant's feelings to tackle important how, what, and why questions (Goldenkoff, 2004). A focus group uses a small number of participants (here: experts) who informally discuss a particular topic under guidance of an independent moderator (or the researcher) (Goldenkoff, 2004, Denscombe, 2010). They are an excellent tool for exploratory studies but also for fine-tuning or expanding existing programmes. They are particularly good for identifying the reasons behind people's likes and dislikes and produce ideas that would not emerge from other qualitative methods such as surveys because they encourage a wider range of comments (Department of Sustainability and Environment, 2005). Denscombe (2010) points out the similarity between focus groups and group interviews and also mentions the fact that focus groups may be conducted on the internet.

For the e-learning experts an interactive on-line focus group was considered which should again give as little guidance as possible to the participants (see 3.7).

2.11. How many experts are needed?

Feedback from one observer is obviously not enough, because even qualified experts may have different and subjective views on what constitutes good teaching, for example. But how many experts are needed?

Most sample sizes used in surveys are relatively high. For example Denscombe (2010) suggests that samples should not involve fewer than 30 people or items. But does this apply to expert evaluation? Mathematical estimates by Ashton (1986) propose that expert opinions could be combined and that mean group validity increased rapidly as more experts were added. At a number of five mean group validity was close to saturation. Suggestions also came from the field of computer studies. Chao and Salvendy (1994) suggest expert numbers ranging from one to six for diagnosis, debugging and interpretation tasks. When Nielsen (1994, cited in Ssemugabi, 2006) plotted the number of usability evaluators against the percentage of usability problems found the result climbed from 30% (1 evaluator), 60% (3 evaluators) to 75% (5 evaluators) and did not drastically improve by adding more evaluators thereafter. He therefore suggested a minimum of five 'usability experts' to identify most usability problems. While some question these recommendations (Woolrych and Cockton, 1986), an evaluation by about five experts in a particular field is generally seen to give sufficiently good quality results (Davids et al., 2011).

One must expect of course that not all evaluators approached will also partake if participation is voluntary. Participants are also free to withdraw at any point (Patton, 2002, Horn, 2011). To take non-

participation and withdrawal into account a target of at least ten experts in both the subject and the e-learning group was suggested prior to commencement of this study to arrive at a possible five experts or greater within each group willing to provide feedback on the course in the end. Further details will be given in the methods and results sections (3.3 and 4.1).

2.12. Summary:

In an academic environment peers can be considered experts in their respective fields. An expert evaluation of university-courses can thus become a peer evaluation. Peer evaluation is well-established in the field of education and is considered a valuable supplement to other sources of teaching efficiency such as student assessment.

E-learning differs from traditional learning because it has a more dominant process dimension and affords novel ways of teaching. One can argue therefore that their assessment should be different. Evaluation of e-learning courses is often dominated by usability evaluation in which usability experts use a set of pre-defined principles (heuristics) to identify potential problems. However evaluation methods used in traditional teaching become more broadly accepted. Despite a large number of publications on e-learning, quality in e-learning is not clearly defined and it is difficult to find agreement on parameters of quality that could be useful for an instructor on a course level. Assuming an interpretivist view, teaching quality is an entity which is not absolute but agreed-on and interpreted by various stakeholders.

Experts can be identified based on recognised standards such as qualification or publication record. For the evaluation of an e-learning course a case can be made that experts should either have expertise in the subject matter or in e-learning, where dual expertise is not easily found. E-learning specialists combine knowledge in ICTs with educational experience. The subject matter experts in this study are Clinical Immunologists, unfortunately not an established discipline but identifiable through the type of work they do.

There is a range of suitable methods which could be used to obtain expert feedback. For this study an on-line focus group is suggested for the e-learning experts and an open-ended written feedback by email for the Clinical Immunology expert group.

Based on published research, five experts should be sufficient to supply feedback on the e-learning aspect and it is assumed that five experts will also suffice to gauge the applicability of the content. Because study participation is voluntary and participants have a right to withdraw a considerably higher number will have to be contacted.

Chapter 3: Methodology

This chapter outlines the methodological approaches which were used for the current research based on the suggestions in the literature review (chapter 2). Certain methodological aspects will also be addressed more in context within the results section (chapter 4).

3.1. Project proposal and approval process

The project presented here is a practical research project within the Masters of Philosophy in Health Sciences Education programme of the Faculty of Medicine and Health Sciences at Stellenbosch University. Initial ideas were explored within the module 'Educational Research for Change' and the project was suggested as part of the 'Research Methodology' module in 2012. A suitable supervisor was then identified and ideas for the project were presented to a panel of educators, the supervisor and fellow students during the contact week in January 2013. A formal project proposal was then compiled following the instructions given by the Health Research Ethics Committee (HREC) of the Faculty of Medicine and Health Sciences at Stellenbosch University (http://sun025.sun.ac.za/portal/page/portal/Health_Sciences/English/Centres%20and%20Institutions/Research_Development_Support/Ethics) and submitted in November 2013. Reviewer feedback was received only in February 2014 and concerns centred on confidentiality and anonymity in an on-line environment which were addressed. Final approval was obtained in March 2014 with an ethics reference number S13/11/232 and a project title 'Expert evaluation of an online course in clinical immunology.'

3.2. Identification of candidates for the e-learning and subject matter expert groups

It was decided to use two separate groups of potential experts (see 2.7), one 'subject matter' and one 'e-learning' expert group rather than trying to identify clinical immunology / e-learning double experts (see 2.7 and 5.3.2).

As pointed out in the introduction (see 2.8 and 2.9) the identification of experts in both groups depended more on the kind of work they are doing than based on a clearly identifiable qualification. Sourcing of potential candidates therefore required a good deal of expertise and judgement in itself and was therefore done by the investigator himself.

In both cases experts known to the investigator were used as a starting point to identify 'expert' departments at various institutions of higher education. The e-learning experts involved as instructors in the 'Cape Higher Education Consortium' (CHEC), a collaboration between the institutions of higher education of the Western Cape of South Africa, also proved a good source. From these starting points more and more possible candidates could be identified. Care was taken to include academics both well-known and not well-known or unknown to the investigator as well as representing different institutions of higher education (Stellenbosch University, University of Cape Town, other institutions). The

investigator reserved the right to exclude experts where personal bias or conflicts of interest were suspected. Professional details, qualifications, contact details were taken from the websites of the respective institutions as far as available and were not further verified. These were also the source for a limited number of personal information, mostly categorical data (gender, institution etc.). A literature search was then performed for each potential candidate to confirm whether a publication record in peer-reviewed journals existed. All the experts were confirmed to have such a publication record except for e-learning expert who had an international 'Achiever Award as best ICT teacher' who was then still retained in the list. The initial list of subject matter experts was increased to 20 in total using the same search criteria. For a list of the experts please refer to table 1 (4.1).

3.3. Sampling

Because of the rather imprecise definition of subject matter and e-learning expertise (see 2.8 and 2.9) it was clear from the start that a good deal of personal judgement would be required to identify suitable candidates. This would necessitate a sampling technique known as 'purposive sampling.' Probability sampling on the other hand would aid to avoid investigator bias in this process by introducing an element of chance in selecting a particular expert. Possible candidates may in addition be grouped into suitable subgroups prior to random selection. This is known as stratified sampling. Moreover, candidates may suggest other suitable candidates in a process called chain referral or snowball sampling (Denscombe, 2010, Fraenkel and Wallen, 2009, Maree and Pietersen, 2007). The method used here combined elements of all these approaches.

Firstly, a list of 15 possible experts (up from the initially suggested 10) was compiled for both the subject matter and the e-learning expert groups (see above). They were then stratified into three sub-groups, as coming from Stellenbosch University, the University of Cape Town, or from other institutions. Two experts were randomly drawn 'from a hat' within each of these three strata. One additional expert who had been informed about the planned research previously was added in each group to bring the total number of prospective participants to be contacted first to 7 in both groups in the first round. This kind of 'stratified purposive sampling' was to be repeated until a total number of five positive respondents in each group was reached. After a number of negative responses in the first round, the subject matter expert group was expanded to 20 in total. Each candidate was also encouraged to suggest other possible experts in the field who could potentially be added to the original lists if they were not already included.

3.4. Making contact with the experts

Potential candidates were approached by email. Tracking options (delivery or read receipts) were not used.

The following documents were also included as attachments (see appendix 1):

- The protocol synopsis

- The electronic participant information leaflet and consent form

- Feedback forms for subject matter and e-learning experts

This provided potential experts with a range of information. For example, it gave an overview of the short-course in Clinical Immunology. It also gave them a summary of the planned current research. They were told which of the expert groups they fell under (e-learning or subject matter). They were informed on their rights (participation voluntary, right to withdraw, confidentiality but not anonymity guaranteed) as well as their duties (provide feedback in writing or in an on-line focus group meeting but no expectations thereafter). It was made clear that there would be no need for travel nor were there any costs anticipated but also that no payment would be made to them in return either.

Ensuing email contact then depended on particular questions or concerns by the experts. It was planned to exclude experts if there was an expectation of payment for services. Experts who proclaimed not to have sufficient expertise or who claimed not to be proficient in English would also be excluded.

3.5. Statistical tests

In order to gauge potential differences in expert behaviour such as response rates some limited statistical analysis was done.

Because of small overall participant numbers a Fisher Exact Test was chosen for this purpose and a free on-line service was used to calculate results (http://www.socscistatistics.com/tests/fisher/Default2.aspx). Significance levels were pre-set and a p value smaller than 0.05 was interpreted as indicative of statistically significant differences in categories between groups.

3.6. Course materials evaluated by experts

Subject matter experts were asked on the consent form to identify two course chapters. These were emailed to them in a follow-up email in pdf format. A rough course overview was also possible through the information provided in the materials above. Further information was provided on request.

The e-learning experts received similar instructions but were referred to the actual course. The short-course in Clinical Immunology is offered on the institutional learning management system (LMS) which is Moodle (version 2.5.6). Due to administrative issues, e-learning evaluators were given access to an older, currently unused on-line version of the course (2.5+; otherwise identical to the current version). Usernames and passwords were created for experts outside Stellenbosch University. All evaluators were

enrolled as 'non-editing assistants' which allowed them to explore but not edit the offering at will and did not give them access to other courses on the LMS.

3.7. Feedback by the expert groups

Subject matter experts provided feedback in writing on the feedback form provided. There was no need for further transcription.

The feedback form to be used gave again a brief overview of the Clinical Immunology short-course as well as a summary of the current study. They were asked to comment on both the course overall as well as on one chapter of their choice. They were asked to give written feedback in a 1-page 'comment box' which was suggestive of the length and detail of the feedback expected from them. There were otherwise no instructions and the format was an 'open questionnaire' as described in 2.10.

E-learning experts were informed that they were expected to give feedback during a focus group meeting on-line. They were told that Microsoft Lync would be used for this purpose. Microsoft Lync is a Microsoft propriety software for web-conferencing. Attendees who did not have the software could still attend using the freely available web-app. Information regarding this was sent to the experts outside Stellenbosch University and assistance was offered.

After three attempts at organising an on-line meeting, never more than two experts could agree on a particular time and three of them eventually agreed to provide feedback in writing (same as the subject matter expert group) whereas only two eventually attended the focus group meeting, which was also attended by the moderator and the investigator.

A moderator was identified as a PhD candidate from Stellenbosch University, who had no connection to the course or the current research project whatsoever but who had a background in educational research. Evaluators were told in advance that the moderator would ask: What do you think makes a good quality e-leaning course? Based on your input: What are the positives in the clinical Immunology course? / What can be improved on?

Only audio was shared during the focus group meeting. Attendees were informed that the meeting audio was recorded. The meeting was later transcribed by the investigator.

3.8 Coding approach

Coding was done manually following the 'inductive approach' outlined by Niewenhuis (2007b). Unlike the a priori approach, codes for further analysis should emerge here from the texts rather than being imposed on them.

For each of the expert groups, printouts of the original transcripts were analysed by the investigator. Key statements were highlighted and initial codes were identified using the table format described by Niewenhuis (2007b) which allowed keeping the original transcripts and the codes and reflective notes together. This process was repeated until higher categories and key concepts emerged

21

when analysing multiple transcripts cumulatively (Denscombe, 2010). Key statements were then cut and paste into tables along with suggestive codes and categories and 2-letter personal identification tags (compare the bullet lists in 4.3. and 4.4., where suggestive categories have been removed). The summary tables were at that stage re-analysed and statements edited and simplified. Collective key statements from both expert groups were finally transferred into a preliminary overview table. At this stage an independent researcher (a Masters student with an education background but no involvement in the study) was brought in who commented on the coding strategy overall and the codes and categories that were identified. Taking her input into account, the final overview (Table 4 in 4.5.) was compiled.

The results section (4.3. and 4.4.) describes the coding approach in context. For illustrations of the coding approach see also appendix 2.

3.9. Summary of chapter 3

This method chapter first outlined the project approval process. It then described the internet search procedure used to identify possible peer immunology and e-learning experts. From this initial pool a pre-determined number of experts was selected using a mixed purposive sampling / stratified random sampling approach. Experts were contacted by email and supplied with a number or information materials as well as informed consent forms. After been given access to the online course, e-learning experts were asked to provide feedback in an on-line focus group meeting. Subject matter experts were requested to select two sample chapters and provide open, written feedback on one of chapters. Audio transcripts and written feedbacks were then analysed manually using an inductive coding approach with the involvement of an independent researcher. Statistical analysis of data on expert characteristics and feedback behaviour was also described.

Chapter 4: Results

This chapter starts off with describing the characteristics of potential experts that were to be approached followed by the response behaviour of those who were selected and contacted. It then summarises the feedback from those experts who replied and finally presents the cumulative feedback from all experts. The findings are not interpreted here. This will be done later in the discussion (chapter 5).

4.1. Expert characteristics

As pointed out in 3.2 two groups of potential evaluators, subject matter experts and e-learning experts, were identified on the websites of institutions of higher education within the Western Cape of South Africa but also from the rest of South Africa using a purposive sampling approach. Web-searches for terms such as "Clinical Immunology" proved very ineffective in locating potential experts and this kind of approach was abandoned early on. The search process started off with experts known to the investigator, thereby locating academic divisions housing possible other experts. An effort was made to select not only candidates known to the investigator but also some less known, not only from mid-tier positions but also heads of divisions. This process was continued until a list of at least 15 experts in each group (up from the suggested 10 in the project proposal) was reached (see **table 1**).

Not all the characteristics within the two expert groups were equally distributed (see **table 1**). For example the subject matter group had an overrepresentation of males when compared to the e-learning expert group which was significant (75% versus 27%; $p < 0.05$ Fisher exact test). There was also a trend for the subject matter expert group being better known to the investigator. The representation of institutions (Stellenbosch University, University of Cape Town, other) and seniority level were more similar. Two of the experts in the contact list for the subject-matter group came from non-academic environments (one private company, one National Health Laboratory Service).

4.2. Feedback behaviour

From the initial list of 15, seven experts each were drawn at random using a stratified randomised approach (see 3.3). Due to an initial lack of positive responses in the subject matter group, a further 5 experts were drawn in a non-stratified manner and contacted thereafter. An additional 5 potential evaluators were added to the subject matter group after this and further contacts were subsequently just picked in the order they appeared on the table. Hence a total of 20 in the subject matter group and a total of 15 in the e-learning group.

Table 1: Summary of expert characteristics

	Contact	Response	Feedback	Institution + Position	Relation to investigator	Gender

Subject matter experts

	Contact	Response	Feedback	Institution + Position	Relation to investigator	Gender
SA	April (1) + reminder	No reply	None	UCT Staff	Known	Female
SB	April (1)	Declined < 1 week	None	UCT Staff	Known	Male
SC	April (1)	Declined < 1 week	None	UP HOD	Unknown	Male
SD	April (1)	Declined < 1 week	None	SU HOD	Known	Male
SE	April (1)	Declined < 1 week	None	SU Staff	Known	Female
SF	April (1)	Declined < 1 week	None	UP HOD	Unknown	Male
SG	April (1) + reminder	Replied 15 weeks	None	SU Staff	Known	Male
SH	April (2)	Declined < 1 week	None	UCT HOD	Acquainted	Male
SI	April (2) + reminder	No reply	None	UCT Staff	Acquainted	Male
SJ	April (2) + reminder	No reply	None	UWC Staff	Acquainted	Male
SK	April (2) + reminder	Replied 4 weeks	Written 18 weeks	SU Staff	Known	Female
SL	April (2) + reminder	No reply	None	Private Staff	Known	Male
SM	May + reminder	Replied 5 weeks	None	SU Staff	Known	Male
SN	May	Replied 4 weeks	Written 4 weeks	SU Staff	Known	Male
SO	May	Declined < 1 week	None	UCT HOD	Acquainted	Male

	Contact	Response	Feedback	Institution + Position	Relation to investigator	Gender
SP	May	No reply	None	UCT Staff	Unknown	Male
SQ	May	No reply	None	UCT Staff	Acquainted	Female
SR	July	No reply	None	UCT Staff	Known	Male
SS	July	Replied < 1 week	Written 3 weeks	NHLS Staff	Acquainted	Male
ST	August*	Replied <1 week	[Written 9 weeks]	SU staff	Acquainted	Female

Table 1 continued - E-learning experts

EA	April (1) + reminder	Replied 4 weeks	None	UWC HOD	Acquainted	Female
EB	April (1)	Replied < 1 week	Written 10 weeks	UCT Staff	Acquainted	Female
EC	April (1)	No reply	None	UFS Staff	Unknown	Female
ED	April (1)	Replied < 1 week	Written 10 weeks	UCT Staff	Acquainted	Female
EE	April (1)	Replied < 1 week	Written 14 weeks	SU Staff	Known	Female
EF	April (1)	Replied < 1 week	F. Group 10 weeks	SU Staff	Known	Male
EG	April (1)	Replied < 1 week	F. Group 10 weeks	SU Staff	Known	Male
EH	Not contacted			UCT Staff	Unknown	Male
EI	Not contacted			CPUT Staff	Unknown	Female
EJ	Not contacted			UP Staff	Unknown	Female
EK	Not contacted			CPUT Staff	Acquainted	Female
EL	Not contacted			CPUT Staff	Unknown	Male
EM	Not contacted			SU Staff	Unknown	Female
EN	Not contacted			UWC HOD	Acquainted	Female
EO	Not contacted			UCT HOD	Unknown	Female

SU = Stellenbosch University, UCT = University of Cape Town, UWC = University of the Western Cape, UP = University of Pretoria, UFS = University of the Free State, NHLS = National Health Laboratory Service, CPUT = Cape Peninsula University of Technology
* In the subject matter group, experts SP onwards were only added after the first draw. Expert ST was recommended by another potential participant who declined; she was included at a very late stage (replacing an expert that had not been contacted yet); her feedback was received after write-up and is not included here – treated as 'committed' (not 'feedback received') in the following analysis

The following response behaviours were observed. Some contacts never replied, even after receiving a reminder email ('no reply'). Some sent an email back but gave various reasons for not being able or willing to participate ('declined'). Others replied that they were willing to participate ('committed'). Out of these a few then ignored follow-up requests. Some pledged to provide feedback after the cut-off date set for this study. The remainder submitted their feedback in time ('feedback received').

The response behaviours in the two groups were dissimilar (see **tables 2** and **3**). In the e-learning group one round of contact was sufficient. From the seven contacts made, five resulted in a feedback

(71%). A sixth evaluator who committed initially did not respond to further requests. As pointed out in the methods section, a feedback through a focus group meeting was envisaged. After two attempts no more than two experts could ever agree on a meeting time. It was therefore decided to schedule a final third meeting and offer the option to either partake or submit written feedback. Two of the evaluators chose to participate in the focus group meeting and three submitted a written report, resulting in five feedbacks overall.

The response behaviour in the subject matter group was much more negative overall. One month and two rounds of contacting later not a single of the 12 candidate thus far had committed to providing feedback. After 5 rounds of contacts only 6 out of 20 experts contacted committed to providing feedback. One of them who committed initially did not respond to further requests. From the remaining five, three feedbacks were received. The other two pledged to submit their feedback later. At this stage the feedback rate is therefore 15%. This is significantly lower than the 71% by the e-learning group (p < 0.5 Fisher exact test). Once the two outstanding experts will have submitted (see below), the subject matter expert feedback rate will rise to 25% and the difference between the two groups will no longer be significant.

Table 2: Response behaviour of study participants

	no reply	declined	committed	feedback received
M (17) [66%]	5 [1.0] *	6 [**1.4**]	6 [0.9]	4 [0.8]
F (10) [37%]	3 [1.0]	1 [**0.4**]	5 [1.2]	4 [1.4]
known (13) [48%]	3 [0.8]	3 [0.8]	7 [1.3]	5 [1.3]
acquainted (10) [37%]	3 [1.0]	2 [0.7]	4 [1.0]	3 [1.0]
unknown (4) [15%]	2 [**1.7**]	2 [**1.7**]	0 [**0.0**]	0 [**0.0**]
SU (10) [37%]	0 [**0.0**]	2 [0.8]	8 [**2.0**]	5 [**1.7**]
UCT (10) [37%]	5 [1.7]	3 [1.2]	2 [0.5]	2 [0.7]
Other (7) [26%]	3 [1.4]	2 [1.1]	1 [0.4]	1 [0.5]

[M = male; F = female; other abbreviations see table 1
* in this example 5/8 : 17/27 = 1.0; an index of 1.0 indicates that the sample group (percentage of males that did not reply) has the same percentage as in the group as a whole (percentage of all males); numbers > 1 indicate 'overrepresentation' and <1 'underrepresentation']. The highest and lowest values in each column are highlighted (bold).]

Although not significant (p >0.05 Fisher exact test), a number of trends were also observed (see **tables 2** and **3**). For example 7 out of 20 subject matter experts (7/20) declined to participate outright, while none of the e-learning (0/7) did so. A number of reasons were given for this with work-related issues predominating (3/7) followed by stating not being an expert (2/7) and health reasons (1/7). Three of the seven 'decliners' provided names of alternative candidates, most of whom had already been included in the list (one of them was added and replaced an expert that had not been contacted yet; 'ST' in table 1).

Many did not reply and it was decided to send out a reminder. Still, in the subject matter group 7/20 never replied in any form whatsoever versus 1/7 in the e-learning group. Non-response behaviour was not significantly linked to gender, level of acquaintance, seniority or mother institution. However, there were some interesting trends which were not statistically significant since overall numbers were too small (**table 2**). For example females appeared less likely to decline outright (index below 1 in table 2). Invitees unknown to the investigator were more likely to decline or not reply at all. On the other hand colleagues from the mother institution (SU) were more likely to commit whereas invitees from other institutions tended not to reply. If there was a response at all, the decision to decline was taken within one week of contact (7/0). The decision to commit was also often taken within the first week (7/12) but took considerably longer in some cases (**table 3**). Actual feedback was received from 8 experts at this stage after a time of 3-18 weeks after initial contact. Experts providing feedback were more likely to be from the home institution and known to the investigator.

Table 3: Time to respond of study participants

	no reply	declined	committed	feedback received
1 week		7	7	
1-4 weeks			3	2
5-8 weeks			1	
9-12 weeks				4
13-16 weeks			1	1
16-20 weeks				1
no reply after 20 weeks	8			

Some other interesting issues: There was some confusion by the experts what was expected from them. Two of the e-learning experts pointed out that they were not subject matter experts (they were never asked to give feedback on subject matter). Two of the subject matter experts expressed a concern about being unfamiliar with web-conferencing and one was not sure about the format of the feedback (they were told to provide feedback in writing on the feedback form provided). Only one expert, a subject matter expert requested additional information (in this case on assessment practices which was provided). Ethical concerns emerged as well: One e-learning expert refused to divulge her ID number (consent form submitted personally instead) and one subject matter expert requested a copy of the consent form not only signed by him but also the investigator (this was done, as envisaged in the consent form).

4.3. E-learning expert feedback

E-learning experts were contacted as described above. One of them expressed excitement about the possibility to give open-ended feedback rather than having to stick to 'those boring checklists.' Those who responded positively were given access to an older mirror site of the course (see 3.6). This had to do with new University regulations for learner enrolment introduced at the beginning of 2014 which made it

near-impossible at the time to enrol outside experts on the actual course. This fact drew criticism from some of the evaluators.

Five feedbacks were received from e-learning experts in the end. Two of them (EF and EG) took part in a focus group meeting as suggested in the original research proposal. Because it proved difficult to accommodate all five experts in one meeting, it was eventually decided to give the option of a written feedback, using the same feedback form as the subject matter experts. Three of the evaluators (EB, EE, ED) took that option. During the focus group meeting the independent moderator initiated discussions with two questions: 'In your view, what constitutes a good e-learning course?' and 'Based on your framework, how do you rate the course to be evaluated here?' In order to level the playing field, the same two questions were emailed to those planning written feedback as well. The recorded audio from the focus group meeting was then transcribed by the investigator. For the written feedbacks there was no need for transcription.

All the responses were then analysed by the investigator. Key statements were highlighted in the text and multiple suggestions for possible codes and higher categories were marked next to the statements. This process was repeated a number of times until some general themes seemed to emerge which worked for all the expert responses. Overall the number and the quality of responses given did not appear to differ between focus-group and written feedbacks and it was decided to analyse them together rather than separately.

Key statements were then shortened and these are presented below as bullet lists (without codes). The amount of key statements made ranged typically from 5 to nine. One of the reviewers provided a written feedback which was much more detailed with 29 key statements overall, more than all the others combined.

Focus group feedback:

EF (8 key statements)

- Good general and chapter layout is logical
- Good instructions (to learners)
- Too much text
- Text small, difficult to read
- Use more graphics
- Navigation buttons not working
- Make pdfs available as simple download
- Include audio (read)

EG (7 key statements)

- Presented as pdfs; not really e-learning
- Too much text
- Use more graphics and animation
- Create new on-line materials: include graphics and animation (engaging), include quizzes (grasping of concepts, revision) 'interactive' [use more than one modality]
- Assessment with immediate feedback
- Navigation buttons difficult to use
- Content – good general impression

Written feedback:

EB (9 key statements)

- Content was 'chunked', systematic, structured
- Good instructions to learners
- Good that pdfs can be downloaded
- Like callouts
- Materials are copyrighted, why not OER?
- Self-study questions not found
- Clinical case studies: add audio, video; more 'engaging'
- Not sure about course blogs; news forum
- How does certification work (tracking, tests)

EE (5 key statements)

- Structure simple, layout logical
- Combine content and assessment
- Navigation buttons unnecessary
- Include discussion and Q&A; interactive communication between participants, 'interactive'
- Content just provides information and may nor facilitate learning

ED (29 key statements)

- Audience is defined (for clinicians)
- Define audience better and earlier (also scientists?)
- Students need to be guided to the correct course
- Made clear: self-paced on-line course
- Duration and cost to user unclear
- Works within institutional LMS
- Students hopefully comfortable with the LMS
- Not sure about the news forum
- Navigation buttons don't work
- No links to other on-line sources
- Recommended readings are identified
- Content overview is clear; optimal sequence suggested
- Some objectives unclear / need to be added
- Not sure about which source materials were used?
- Good: course not designed to convey a prescribed set of knowledge
- Good general layout and grouping of chapters
- Fix some headings / sub-headings
- Good: course evaluation through student feedback
- Far too much text
- Pedagogy too 'instructivist'
- Learning response too passive (click – read)
- Call-outs are nice
- Good use of examples
- Activities are tracked
- Assessment strategy, certification process made clear
- Assessment - address 'peeking'
- Not sure about the certification of completion
- Copyright is clear
- Why not open-source?

Some of the comments appeared of more general nature (later split into 'learners' and 'content' in **table 4**). Many addressed usability or the learning process. There were few comments on legal aspects as well, such as copyright or the possibility to offer course materials as open educational resource (OER). Responses could also be classified into 'good' and 'improve'. The coding and the higher categories were discussed with an independent researcher not involved in the study and her suggestions were taken into consideration. Here are some of these statements in their original form:

1. **Layout**: "The structure is simple and the division of the content and assessments makes sense [EE]."

2. **Instructions to learners**: "I like the intro on how to use … [that] type of info. [EB]"

3. **Callouts**: "I think the call-out box is a good idea [ED]"

4. **Navigation buttons**: "I found it difficult and at times, unnecessary to use the navigation buttons [EE]"

5. **Text**: "FAR too much text on these slides – REALLY hard to read. [ED]"

6. **Learning response**: "Seems to be very passive – click –read - click – read – click read [ED]"

7. **Use various types of media**: "…I thought It could be more engaging and less text based as this lent itself to a video/ or even an image … with voice over [EB]"

8. **Combine content / questions and give immediate feedback**: "… every couple of slides throw in a question; break up the flow … and … reinforce a concept … but then … give the answer or give feedback immediately. [EG]"

9. **Open educational resource**: "Why have you chosen to copyright this material? I think you could get great profile and cudos [sic] for making it an OER. [EB}"

10. **General**: "… to me it's quite clear that he's a, an expert in the field, and I immediately feel quite comfortable that the content is sound …but I don't find it very appealing as something to learn online only [EG]."

4.4. Subject matter expert feedback

It was alluded to above already that feedback from subject matter experts was much more difficult to obtain. The response was in writing using the feedback form provided. At the time of writing only three feedbacks were received (SN, SS, SK). A further two (SG and ST) are expected to be received later and cannot be included here. The coding of subject matter feedback was done separately and after analysing the responses from the e-learning group. The instructions were to comment both on the course overall as well as on one selected chapter. Where chapter content was commented on, either suggestions for improvement were made or certain items were identified as presumably incorrect and in need of being corrected. Again, the key statements were shortened and are presented below as a bullet list.

SN (Infectious diseases chapter) (9 key statements)

- Suitable for all postgraduate clinical and research students and scholars in infectious diseases
- Covers important aspects
- Well-structured
- Simple language
- Use of graphics
- Replace more text with figures
- Shorten the text
- Content: Lytic cycle [needs correction]
- Content: % human genome viral [needs correction]

SS (Infectious diseases chapter) (10 key statements)

- The chapter is informative
- Good background; links physiology and immunology to pathology
- Content: Reactivation / reinfection / recurrent infection [suggestion]
- Content: Immunopathology is … [suggestion]
- Content: Diagnosis of TB [suggestion]
- Content: Remove bullet
- Content: Molecular mimicry [needs correction]
- Content: Herpes zoster [needs correction]
- Content: LPS [needs correction]
- Spelling: Tuberculosis [needs correction]

SK (Hypersensitivity) (8 key statements)

- Appropriate on postgraduate student level
- Covers subject matter comprehensively
- Easy to understand language
- Text style not reader friendly
- Use more illustrations
- Cannot go back
- No self-assessment in the text
- More clinical cases and examples

ST (only received after write-up; not included here)

Similar to the e-learning group, 8-10 key statements were made by each expert. Interestingly, some of the evaluators focused almost exclusively on course content (SS), while one made no specific reference to content at all (SK). Remarkably, where two evaluators evaluated the same chapters, different and non-overlapping suggestions for improvement/ correction were made. Here are some of these statements with suggestive categories:

1. **Audience**: "It is a well-structured very valuable chapter… It would improve the understanding of most scholars or students in … clinical science or medical research … of infectious diseases. "

2. **Appropriateness**: "…the chapter is appropriate for a postgraduate student level and covers the subject matter comprehensively…"

3. **Content**: "Shingles (Herpes zoster) does not really cause 'impaired function' … it is just a painful rash, but patients are still 'functional'."

4. **Learning**: "More illustrations and cartoons would improve the content … Clinical cases and examples will augment and explain the factual text."

5. **Assessment**: "There are no self-assessment questions in the text."

4.5. Cumulative feedback from experts

After the first subject matter feedbacks were in it became clear that there was some degree of overlap between both groups. A further category 'content' and a sub-category 'audience' now emerged. All evaluator feedbacks were now re-analysed to identify common themes. These were discussed with an independent researcher (see above). Based on her comments the feedbacks were arranged in two simple comments 'good' or 'improve.' Other higher categories were 'content,' 'learners,' 'usability,' 'learning,' and 'legal.' Key statements were simplified to accommodate responses from multiple evaluators. An overview of this is given in **table 4**. It appears that there is a roughly equal amount of positive and negative remarks. The table visualises to what extent evaluator comments merge on certain statements and it appears that some saturation has now occurred. The table also shows the overlap between comments from e-leaning and subject matter experts. Some comments appear contradictory. A further analysis of the findings can be found in the discussion 5.2.2.

Table 4: Summary of responses from both e-learning and subject matter experts

	Good	Improve
CONTENT		
Adequacy	▪ Good background suitable for defined audience (SN, SK, SS)	
Authenticity	▪ Good overall impression (EG)	▪ Unclear which sources the course is based on (ED) ▪ Improve some content (SS) ▪ Correct some content (SN, SS) ▪ Correct some spelling or grammar (SS)
Length	▪ Important aspects covered; informative (SN, SS) ▪ Comprehensive (SK)	

Table 4 continued

LEARNERS		
Audience	▪ Audience is defined (ED, SN, SS)	▪ Define audience earlier and better (ED)
Delineation of the course	▪ Made clear: self-paced on-line course (ED) ▪ Learner instructions, objectives good overall (EB, EF) ▪ Recommended readings identified (ED)	▪ Duration, cost to user etc. unclear (ED) ▪ Some objectives unclear / need to be added (ED) ▪ No links to other on-line sources (ED)
USABILITY		
LMS*	▪ Works within the institutional LMS (ED)	▪ Make sure students comfortable with the LMS (ED) ▪ Make sure students guided to the correct course (ED) ▪ Uncertainty what the course blog and news forum etc. are for (EB, ED)
Course	▪ Simple, logical general layout (ED, EE, EF) ▪ PDFs are downloadable (EB)	▪ Some headings / sub-headings need fixing (ED) ▪ PDFs should be more easily downloadable (EF) ▪ Navigation buttons not working / unnecessary (ED, EE, EF, EG, SK)
LEARNING		
Social	▪ Students have opportunity for feedback (ED)	▪ Enable communication between learners (EE)
Cognitive	▪ Not designed to convey a prescribed set of knowledge (ED) ▪ Materials well-structured (SN) ▪ Text was chunked (EB) ▪ Simple language (SK, SN) ▪ Callouts are nice (EB, ED) ▪ Use of graphics (SN) ▪ Use of examples (ED)	▪ Too much text (ED, EF, EG, SN) ▪ Style and font hard to read (EF, SK) ▪ Passive – just content/ reading (ED, EE, EG) ▪ Too 'instructivist' (ED) ▪ Use more graphics (EF, SK, SN) ▪ Use animation, audio (EB, EF, EG) ▪ Use more examples, clinical cases (SK)
Assessment and certification	▪ Activities are tracked (ED) ▪ Overall assessment strategy, certification process clear (ED)	▪ Combine content with questions and give immediate feedback (EE, EG, SK) ▪ Self-study questions not found (EB) ▪ Tracking unclear (EB) ▪ Certification details unclear (EB, ED) ▪ Address student cheating (ED)
LEGAL		
Copyright & OER	▪ Copyright clear (ED)	▪ Materials should be OER (EB, ED)

e-learning experts: EB, ED, EE, EF, EG / subject matter experts: SN, SS, SK
* LMS = learning management system

4.6. Summary of chapter 4

Two groups of potential evaluators, subject matter experts and e-learning experts were approached within mostly academic institutions of South Africa. Their expertise was confirmed by an existing publication record. The subject matter group contained a significantly higher amount of males. The response behaviour from both groups was dissimilar. While five out of seven e-learning experts provided feedback, only three out of twenty did so in the subject matter group, although two more are committed to still deliver feedback. Overall there was a trend for respondents from the own institution to provide feedback.

The originally envisaged focus group approach for the e-learning group did not work well for logistic reasons and most of the experts provided written feedback in the end.

Although the responses from the two groups overlapped somewhat, subject matter experts commented largely on course content. They confirmed that course materials were adequate for the target group of learners. E-learning experts appreciated the clear layout and the delineation of the course to learners but identified issues around usability and the facilitation of learning.

The recommendations from both groups of experts were then accumulated and grouped into higher categories. This revealed that certain items were identified by multiple evaluators, indicating a degree of saturation. Their cumulative input is shown in table 4 and provides a summary of the feedback obtained.

Chapter 5: Discussion

The discussion starts off with a scrutiny of the research process. It then proceeds with an analysis of the results. While chapter 4 above presented expert feedback uncommented, this chapter interprets the suggestions, especially with respect to whether and how they could be implemented. It is finally argued that the findings are not only applicable here but relevant for other settings. In conjunction with suggestions from the literature recommendations for others can be made.

5.1. Discussion of the research process - Trustworthiness of findings

One key criterion of good qualitative research, according to Nieuwenhuis (2007b), is the 'trustworthiness' of data analysis, findings and conclusions. To achieve this he lists verifying raw data, keeping an account of research decisions taken, using multiple coders, stakeholder checking, verification and validation of findings, controlling for bias, choosing quotes carefully, maintaining confidentiality and anonymity and stating limitations upfront (see above). He also includes avoiding of generalisation and suggests to rather seek insight into participant's perspectives, experiences, attitudes, and behaviours. In Denscombe's (2010) view it is the 'verifiability' of findings, which makes or breaks good qualitative research. This encompasses validity, reliability, generalizability (external validity), objectivity (absence of bias). In order for the findings to be credible or valid, the investigators need to demonstrate that their data are accurate and appropriate. Reliability or dependability entails that other researchers would reach similar findings using the same approach. By proxy the investigators need to give a full account of methods, analysis and decision-making. Transferability or generalizability according to Denscombe (2010) should answer to the question as to what extent findings are likely to exist elsewhere. This issue of objectivity or confirmability concerns the extent to which the findings are free from investigator bias. Some of these important issues will be discussed next.

5.1.1. Sampling

A first possible source of error and bias could be the sampling process. According to Fraenkel and Wallen (2009) there are some threats to internal validity related to sampling, importantly selection bias and loss of subjects.

As outlined above, a 'stratified purposive' sampling approach was taken which also included an element of snowball sampling. However, can a purposive approach ever be free of bias and can it lead to generalizable results with value in other settings? Some authors think not. Sivo and Saunders (2006) state that non-random samples 'result in sampling error'. According to Radhakrishna and Doamekpor (2008) when using non-random samples findings cannot be extrapolated to the population. Other authors express a more positive view. Fraenkel and Wallen (2009) compare random and non-random/ purposive sampling. They explain that non-probability approaches require an element of discretion or choice on the

35

part of the investigator while still retaining the aim of generating a representative sample or at least an explorative sample. Denscombe (2010) concurs. He says that in purposive sampling the sample is 'hand-picked' for the research on the basis of relevance to the issue or the theory being investigated. This requires knowledge or experience about the topic on the side of the researcher. He describes this as a pragmatic approach often used in social research when samples are small while still leading to either representative or exploratory samples.

In the current study some element of random sampling and snowball sampling were introduced to limit investigator bias. However, some element of non-random sampling was hardly avoidable. As outlined in 2.8 and 2.9, many of the terms defining the expert groups envisaged were ill-defined and proved to be of little use on search engines such as google. This also meant that the selection of experts could not easily be handed over to an independent researcher, thus avoiding selection or exclusion bias on the part of the investigator. Finding possible candidates required a good deal of inside knowledge of the respective fields and the academic environment of South Africa from which the evaluators were to be selected. A selection of names of departments from which the experts were eventually picked highlights this. Subject matter experts came amongst others from Medical Virology, the Division or the Department of Immunology, Paediatric Immunology & Rheumatology, or the National Health and Laboratory Service. Note that the term 'Clinical Immunology' does not appear in any of these and that 'Clinical Immunology' does not exist as a clinical speciality or sub-speciality in South Africa as explained above. Choosing suitable e-learning experts required a distinction between more technically-orientated IT divisions and more academic divisions with educational expertise. Again a brief list of names: Centre for Innovative Educational and Communication Technologies, Centre for Educational Technology, Educational Technology Unit, Division of Nephrology, or SURMEPI e-teaching and learning. For both 'Clinical Immunology' as well as 'e-learning' the total population of all possible experts was not defined.

The eventual samples chosen can therefore be described as explorative at best and claims of representativeness cannot be made. One has to question though whether representativeness is a strict requirement to conduct research through expert review. As outlined above (see 2.6) other expert characteristics such as 'expertise' appear more important. It will be argued below (see 5.1.2) that experts should also be free of personal bias. Lastly, following the logic of Nielsen (1992) there are limited number of important issues to be identified by evaluation and given a sufficient number of evaluators, be they representative or not, these will be revealed (see 2.11). One must also bear in mind that peer review or peer observation of teaching often don't rely on all the possible expertise available out there but mostly make use of in-house expertise, yielding good results in principle.

Sampling bias may also be introduced by insufficient sample size resulting in sampling error (Sivo and Saunders, 2006). Maree and Pietersen (2007) explain that both theoretical and practical considerations may determine the sample size chosen for a study. For pragmatic reasons a number of five

experts in both the subject matter and the e-learning group was chosen bringing the combined planned total to ten (see 2.11).

5.1.2. Data collection

There are other threats to validity which can occur after the sampling process, during the collection of data. Bias may arise from an error in the data collection.

Error may be introduced through a poor choice of methods and badly standardised instruments, thereby affecting the reliability of results (Maree and Pietersen, 2007). Ideally, more than one method should be used allowing triangulation of findings (Denscombe, 2010).

As pointed out above, the choice of methods for this study was based on mostly practical and financial considerations. Some brief pilot testing was done within the module 'Educational Research for Change' during the first year of the MPhil HSE to test formats such as the blank sheet questionnaire for subject matter experts and expert interview and observation (walk-throughs) for technical experts. Comparing multiple methods in the main study, say focus group feedback plus blank-sheet written feedback within each expert group would have been beyond the scope of a practical MPhil project. Obtaining comments from two times five experts proved to be hard enough in the end.

In the end a decision was made to use written feedback for the subject matter expert group and focus-group feedback for the e-learning expert group. The focus-group approach was chosen for the e-learning group with the intention that interaction within the group might prompt participants to reply and that concepts might be developed together (Denscombe, 2010). As it turned out, focus group meetings were difficult to organise because study participants were not willing to compromise their busy schedules. Written feedback, allowing for completion of comments in their own time, were more easily accepted. In the end, the focus group meeting consisted of only two participants which hardly allowed for interaction and responses were similar in nature and extent as those received in writing. Written feedback rather than focus group meetings were considered for the subject matter group right from the start for two reasons. Firstly, they each had to focus on a particular course chapter allowing for overlap only in the general comments for the course overall. It was also feared that they would be uncomfortable using technology for feedback, which may have created trepidations in its own right.

In the context of the current study it is important to explore whether the responses given by the experts may be biased in any way. House (1976, cited in Goldie, 2006) states that evaluation sponsors, participants and audiences share responsibilities. He lists possible fallacies during the evaluation process such as clientism, i.e. evaluators trying to please the client. The latter point may be particularly pertinent in a higher education setting where those being evaluated may be called upon later to evaluate other programmes and it becomes a quid pro quo ('I don't tell on you if you don't tell on me'). Also Swinglehurst et al. (2008) bring up the fact that there may be certain 'no-go areas' associated with colleagues observing each other's teaching. The overall result is response bias, a type of bias where the subject gives responses

37

that they think that the investigator wants to hear. Being part of research may in itself have an effect upon the behaviour and the attitudes of subjects may change. This is known as Hawthorne effect (Fraenkel and Wallen, 2009).

As mentioned above, peers supplying feedback in this study were often from the home institution rather than from outside which may suggest some bias. However, the results do not indicate that 'friends' supplied more positive than negative answers overall when compared to outside experts or that they shied away from particular issues as the responses were quite evenly spread in all regards. Participants were informed that while anonymity was not guaranteed their confidentiality would be assured. This was due to the investigator himself having to maintain email correspondence. It was therefore quite clear that the investigator would know which comment came from particular experts which may have hampered their replies. If more funding had been available, an independent investigator could have been paid to take over this task handing over anonymised feedback forms and transcripts to the investigator, thereby allaying possible apprehensions in study participants.

Researchers may also introduce bias themselves. For example, too much pressure may be applied to the subjects, forcing them to reply and thereby skewing their response behaviour. There is also the danger of 'leading', when the instructor is involved in the data gathering process. Denscombe (2010) describes this in the context of interviewing. The personal identity, self-presentation and personal involvement of the interviewer are important factors which may overall result in an 'interviewer effect.' A good interviewer should be practised in using prompts, be non-judgemental and should avoid asking leading questions.

In this study, participants were well-informed about the study and clearly made aware of their right to withdraw. Reminders were sent to non-respondents once and if a participant chose to ignore a reminder after giving consent initially, no further pressure was applied. It was decided to rather 'err on the side of too much ethics than too little.' Interviewer (investigator) bias in the focus group meeting was minimised through the use of an independent moderator. For the written feedback only background information was provided leaving the actual feedback form free of any further questions or guidance altogether. Overall then, participants provided their responses free of pressure and with minimal involvement of the investigator.

5.1.3. Analysis

Challenges to validity can also occur at the analysis stage. The University of Minnesota Center for Teaching and Learning (2013) states that there is a possible bias relating to the observer's own beliefs, the temptation to see what one wants to see, when analysing the data. This is particularly pertinent when an inductive approach to coding, as in this study, is taken. According to Nieuwenhuis (2007b) researchers should keep notes of all research decisions taken. They should also enhance the credibility of findings by

giving others (stakeholders, people with specific interest in the research) access to research findings and include them in the process of reaching interpretations and conclusions.

In the current study records of coding from various stages done by the investigator were kept and discussed with a volunteer with experience in educational research but not involved in this study. Her comments and suggestions were taken into consideration.

One problem was particular to the study. As indicated in the method section, a considerable amount of information about the experts involved in the study could be obtained from departmental websites and literature searches. However only in certain cases actual qualifications (degrees obtained etc.) were available - only 'staff' and 'HOD' was eventually used. The relation to the instructor was added subjectively by the investigator himself. With hindsight, both the qualifications and the perceived relationship to the course instructor could have been answered by the experts themselves in the consent form and would then have been available at least for those who responded.

5.1.4. What has been learnt from the methodological approach taken?

Many of the methodological choices made here, were made more for practical and financial considerations. While this was clearly declared in above (see 1.8), given more time and money, other instruments or a combination of instruments may have been chosen. Also independent researchers may have been included to a greater extent to interact with experts and analyse data. Furthermore, certain aspects of quality assurance have been excluded here either because they have been explored previously (student feedback) or simply to assure the coherence of the research project in the view of the investigator. Consequently, the research question was phrased quite narrowly. This may have limited the breadth of the project.

With hindsight, quite practical methodological improvements could have been made in particular with a view on improving response rates. Firstly, subject and e-learning experts received the same information materials in an attempt to make them aware of the study as a whole. This however caused some degree of confusion as to what expert group they found themselves in and what was expected from them. With hindsight, these two aspects of the study could have been more clearly separated. Secondly, the experts were contacted (initially) in bulk and by email in an attempt to treat them all the same whether unknown or known to the investigator. It may have proved more successful to contact each expert individually. It may have also helped to activate the 'confirmation of reading' function in the email message (this was not done for concerns of coercing potential candidates). Thirdly, it had to be made clear that no form of payment could be made to potential experts. Given some limited funding some kind of incentive could have been implemented. It is assumed that open questions may have been viewed positively as pre-coded questions are often perceived as frustrating, deterring from answering (Denscombe, 2010). In the email responses from the experts this open-ended approach was welcome by

one of the e-learning experts, but also questioned by one of the subject matter experts. Overall, a more differentiated approach between the two expert groups may have been helpful.

It is felt that overall reasonable steps were taken to arrive at trustworthy and credible findings.

5.2. Discussion of results

5.2.1. Response rates and response behaviour

A very interesting and unexpected finding in this study was the fact that response behaviour differed between the two expert groups. At this stage of analysis five out of seven e-learning experts provided feedback while only three out of twenty Clinical Immunology experts did so. Should two more Immunology experts still submit, which is expected, the difference seen will no longer be statistically significant. What could have caused this non-response and why are the two groups different?

Participants may be generally apprehensive about the concept of peer evaluation, which might be seen as a staff performance management tool (Swinglehurst et al., 2008; Schultz and Latif, 2006). A clue for apparent non-response may also come from studies on participation in surveys. Groves, Cialdini and Couper (1992) as well as Denscombe (2010) list various factors ranging from generalised categories such as societal factors, a helping tendency or emotional state to issues relating to the specific survey such as unattractiveness, lack of reciprocation or a cost in time and energy. Creighton (2003) discusses the 'resistance' personality type in another context and list factors such as feelings of inadequacy and not being convinced about the value of an implementation.

While all the factors mentioned above may provide an explanation for non-response behaviour in this study overall, they do not explain why Clinical Immunology experts and e-learning experts apparently behaved differently. The only significant difference between the groups was an overrepresentation of males in the immunology group when compared to the e-learning expert group. However, 'maleness' was not related to non-response behaviour overall and may have been a confounding factor. One factor may be the attitude towards participating in a study evaluating teaching using an interpretivist approach and a sense of possible reciprocation. It can be speculated that the Clinical Immunologist group through their work may have never had exposure to qualitative research. The value of such research may thus be questionable to them. The e-learning group on the other hand contained a number of qualified educators who, while checking for their publication record, proved to have experience in quantitative research themselves. They may have a more positive attitude towards social research and teaching in general and may well be thinking of evaluative studies themselves, expecting collaboration with others.

The low number in the subject matter group resulted from a large rate of non-response. There are certain biases associated with non-response which are distinct from low sample size (Fowler, 2009). Shadish, Cook and Campbell (2002, cited by Sivo and Saunders, 2006) argue that nonresponse error is akin to selection bias in experiments. Denscombe (2010) distinguishes between non-response through refusal

and non-response stemming from non-contact. Marcus and Schütz (2005) distinguish complete nonresponse, careless and selective item omission, and follow-up nonresponse. But which rate of non-response is considered too low? Unfortunately there is no hard and fast answer and even a low response rate may not necessarily be a predictor of nonresponse error (Fowler, 2009). Non-response behaviour may cause error when non-respondents share particular features and through not responding skew the findings in a systematic and relevant fashion (Denscombe, 2010), Sivo and Saunders, 2006). Marcus and Schütz (2005) find that nonresponse biases may have significant implications for representativeness in surveys. Also Radhakrishna and Doamekpor (2008) warn that non-response may impact on the generalizability of findings. Only if there is no difference between early-, late-, and non-responders, can one generalize findings to the population. If the differences are unknown, one may generalize but this will impact on external validity. Dooley and Lindner (2003) recommend comparing early to late respondents, and comparing respondents to non-respondents as an approach to address non-response error.

The fact that only three subject matter experts had submitted their feedback as of end August was in fact a big drawback. At this stage of the analysis the findings from this group likely still lack overall validity. On the other hand, it was clear from the beginning that not all course materials could be made subject to evaluation. Even a simple coverage of one expert per course chapter would have required 16 subject matter experts altogether. It was decided to choose a number of five, similar to the e-learning group, to get at least an idea about the appropriateness of the content and the type of shortcomings that might be identified.

5.2.2. Recommendations by the experts for course improvement

The study described here attempted to answer the question of whether an open-ended, interpretivist approach to expert evaluation would provide suggestions for course improvement for both content and instructional practises of a selected e-learning course, a short-course in Clinical Immunology.

One aspect of the study was to look at finding answers to a local problem. In that respect the study was descriptive, and situated (see 1.7). So, what kind of suggestions were made by the experts and how may this affect this specific course offering in the future?

For a summary of expert comments, the reader is referred back to table 4 (4.5). The table includes collective statements from all the experts and many of the themes emerged when all the feedbacks were analysed cumulatively and in this respect, many conclusions could be drawn that would not have been possible by just analysing expert feedbacks individually. What follows next is a summary of this collective expert feedback.

The subject matter experts did indeed confirm that materials were adequate for the intended audience, which are postgraduate students such as registrars or Medical scientists in training. They also found that the subject matter was presented in adequate detail. Within the actual chapter content they made suggestions and pointed out presumed errors. Two of the experts looked at the same chapter

(infectious diseases) and it was interesting that the shortcomings that were identified differed between the two. This highlighted the fact that a simple coverage with one evaluator per chapter would miss many possible points of improvement. The take-home message for others planning to review course content is that an exhaustive coverage of all content through expert review is very likely unmanageable and that other forms of review might be considered. Overall we were glad to hear that the experts confirmed that the course materials were pitched correctly. The few 'mistakes' that were identified will be investigated.

One recommendation on content came from an e-learning expert who was uncertain on which 'canon of knowledge' the course was based on. On the other hand, some of the suggestions from subject matter experts covered course delivery rather than content and overlapped with those made by the e-learning experts and will be discussed there. This is not surprising, since both expert groups have experience in education. It also illustrates the fact that principles of good teaching largely overlap between e-learning and traditional learning (see 2.2).

In contrast to the subject matter group, the e-learning experts were not asked to look at particular course chapters in detail. They rather explored the course as a whole at will. Some prominent themes emerged from their replies, which are shown in table 4.

A great number of comments were made on the usability of the course. Some of this was related to the institutional learning management system (LMS) and is therefore separated in table 4. It was appreciated that the course operates on the institutional LMS but it was also pointed out that learners need to be guided to the correct course and need to be comfortable in using the LMS. This is currently clarified by an email from the course instructor to the learner. Evaluators then started pressing all the buttons available on the LMS, such as the news forum. Many of these are not used for course instruction. Learners will be possibly as confused as the evaluators were and it will be made clear to the students in future which elements of the LMS are used and which not or only on occasion.

A large number of comments were also made on the usability of the course itself. It was pointed out that the general layout is simple and logical although some headings or sub-headings may need to be looked at. It was also appreciated that learning materials are downloadable, however one expert suggested to merge all pdfs into one big document allowing for a single download. Most evaluators had issues with the navigation icons used currently in the documents and none of them actually liked them. These icons were added to allow learners to freely navigate within the document possibly skipping whole sections they are not interested in. Although the learners themselves never complained about the navigation buttons it will have to be confirmed whether they experience similar problems.

A huge amount of comments were made on the learning process. One interesting aspect emerged during coding. Some of the evaluators suggested to make the course more interactive, but it emerged from the context that they understood interactivity differently. Wang, Woo and Zhao (2009) explain that interactivity is a concept of constructivism. While socio-constructivists believe that knowledge is collaboratively constructed, cognitive constructivists believe that students construct knowledge

individually. Interactivity in the first sense requires enabling learners to interact with each other, in the second sense it requires enabling learners to actively interact with course materials. The two require different e-learning approaches and are therefore separated in table 4.

It was criticised that learner-learner interaction is completely absent from the course. This is in fact true and necessitated by the fact that only one student is active on the course at any given time. Real social interactivity can therefore not easily been facilitated in this specific case.

Many comments were made relating to cognitive constructivism and some weaknesses were pointed out. It was strongly suggested to use not only more graphics, but also other modalities of knowledge transfer such as audio and animations. Even a particular programme was pointed out to make this possible. The teaching style overall was perceived as 'instructivist' and 'not making the most of the potential of the e-learning platform.'

A number of comments were also made on assessment, in particular the proposition to combine content with questions and giving immediate feedback. Together with the suggestion to include other modalities such as audio as well as more graphics, it was recommended to include a number of small interactive and voiced-over self-study files. These could reinforce key concepts, pause for reflection and ask questions which could be automatically answered.

Overall, the feedback from the e-learning experts was rather negative. One of them summed it up by saying:

"…it shouldn't be a case of merely taking …something which should be identical in a paper-based version and making it accessible on-line. That to me would be… not making the most of the potential of the e-learning platform."

5.2.3. What has been learnt from the results?

After the expert replies are now in, there is of course a range of issues which have been considered by the course convener translating to a range of changes to be made to the course in the near future. Most importantly, two interactive, voiced over modules are currently being developed which will actively engage students by asking questions, while also providing immediate feedback. Participation in these activities could be enforced and traced. These will not replace the existing materials, the content of which was deemed by the evaluators of high quality. The formal assessment with delayed instructor feedback could be phased out completely.

It is also envisaged to introduce a forum to the course. Students could then leave small bio-sketches of themselves and reflect on which aspects of the course were most pertinent to their clinical practice. In this way at least some non-synchronous interaction between learners could occur.

Apart from the actual expert feedback, there have also been important lessons regarding the open-ended interpretivist approach for course evaluation by experts. At the onset of this study it was not clear whether this would be appropriate. The theoretical underpinnings were clearly declared

above, as demanded by Rees and Monrouxe (2010). The interpretivist approach was chosen for the current study because quality in e-learning education was variably defined by different authors resulting in a lack of agreement on what kind of frameworks to use in the design and evaluation of e-learning courses (see 2.4). Notwithstanding this viewpoint, a number of frameworks have indeed been defined by numerous authors and expert committees and one may argue that quality standards could be extracted from the published literature to fit a particular context. This alternative view might still envisage expert review but it would drastically affect the nature of the research suggested here. It would likely be positivist, ask a closed-type research question (does the course conform to defined standards/guidelines) and use quantitative research methods (Ringsted et al., 2011). Some of the assumptions made in the current study follow on the interpretivist approach which was chosen (see 1.4) and had an impact on the choice of methods (quantitative) and analysis (inductive). More practically, there was also a concern whether splitting the experts into subject matter and e-learning would work and whether a number of five experts per group would be sufficient.

When looking at the e-learning group alone it is evident that a number of problems were identified in the course. Some of these were pointed out by a single evaluator. Others were recognised by up to four experts. This may indicate that a degree of data saturation was indeed reached using a number of five experts in total and that adding more would only marginally increase the number of additional problems to be found. It is noteworthy that one evaluator contributed most of the suggestions which indicates that a number of items could have been missed with a different group of only five evaluators. Overall, it appeared that a total number of five evaluators worked well in the context of the open-ended approach taken in this study and the saturation was thus similar to the one observed in a more structured usability approach (see 2.11). Interestingly, some of the issues were identified by both expert groups and suggestions partially overlapped.

One of the most exciting and unexpected findings that emerged was that the suggestions by the evaluators, when grouped according to categories as done in table 4 actually provided something which could have been a good framework for course evaluation tailored to the requirements of this particular course, had a positivist approach been taken. This framework is different from the one originally envisaged by the investigator. Whichever framework would have been chosen, it would have forced evaluators to follow a pre-set path only to confirm the investigator's own pre-conceptions and may have limited additional responses by the experts or missed certain points. The constructivist approach chosen for the current study instead resulted in the development of a suitable framework from the bottom up provided by the experts in a group effort.

5.3 Relevance in other settings

5.3.1. Is this research?

One of the stated goals of the current project is to make course improvements based on expert feedback. But is this research or just an ordinary process of ensuring quality? Many authors regard research aimed at course improvement as 'poor' research. Jansen (2007) for example regards research questions around planning, needs analysis, or solution-seeking poor research questions. Also Ringsted et al. (2011)) state that research should not be about solving local problems or finding solutions. However Jansen (2007) points out that 'applied research' in contexts such as computer studies or community development is an exception if it is designed to provide meaningful data to improve a situation.

The project suggested here also attempted to take evaluation further into the arena of evaluative research. Suchman (1967) as well as Powell (2006) define evaluative research similarly as the utilization of standard social research methods and techniques for the purpose of making an evaluation. Pratt (2006) points out that 'pure' evaluation and evaluative research share methods and approaches and they both add knew knowledge and make a meaningful contribution. However, their audiences and their main objectives are different. Evaluative research aims to enhance knowledge and understanding through publication to a scholarly community whereas an evaluation report is aimed to inform or influence decision makers.

Based on these criteria, the current project is evaluative research. It aims to contribute to the pool of knowledge in the area of expert evaluation of e-learning courses and make suggestions which are aimed at the scientific community in the fields of medical education and information technology.

5.3.2. What type of experts might work elsewhere?

For the current study it was initially envisaged to identify only one type of expert, knowledgeable in both immunology and with some expertise on the use of technology in teaching, i.e. a dual experts. There was however only one expert of this calibre known to the investigator within the borders of the Western Cape region of South Africa, and this plan was abandoned early on. For the current study therefore two types of expert were chosen, Clinical Immunology and e-learning experts and the reasoning behind this was given in 2.7. This decision was obviously tailored to the specific task at hand, which was evaluating an online course in immunology. Overall, it appeared that gathering feedback from two different expert groups worked well in this study with each commenting on different aspects of the course. But what may work for others contemplating some form of expert or peer assessment of a teaching activity?

Some may consider reviewing the content of courses and the experts asked in this study confirmed that content is indeed one important part. In this sense 'Clinical Immunology' can be replaced with whatever subject matter is being covered in a course. Identifying suitable subject experts may be

easy if the subject matter is well defined and clear career paths exist or, as the example of the current study shows, this may be more complex. While subject matter expertise can doubtlessly contribute to course improvement, the first question instructors need to ask is whether expert review should be used or alternative approaches rather (see 2.1). A decision must also be made whether the review of course materials should take the form of 'simple' evaluation or evaluative research. If content review does not offer much that is of interest beyond the stakeholders involved in a course it is not very suitable for the purpose of educational research which aims at an outside audience (see 5.3.1).

An educational offering is of course more than just content delivery. The evaluation of teaching practises (as opposed to content) may require various other forms of expertise. The current study followed suggestions from computer sciences where usability experts review technological entities through heuristic evaluation. It was found here that 'e-learning experts,' as defined in 2.8, were well-suited to fill this role in an educational context. The same likely applies to many educational offerings which operate on a course level and make use of e-learning. However Swinglehurst et al. (2008) describe a less formal peer observation of teaching-approach for online environments in which virtually any fellow educator, usually a colleague, may provide constructive feedback.

Other types of expertise may however be considered where the purpose and audience are different. This will determine whether one type of expert will suffice or whether experts with various types of expertise should be called upon.

Schultz and Latif (2006) bring up the point that peer raters should also have expertise in adult learning and curricular design. Wyllie (2011) describes a situation where assessment in an online environment was of particular importance and suggests that reviewers did not have to be from the same discipline, but that the experience of teaching within a Blended Learning Environment was essential for this purpose. Some suggest to include also administrators in both traditional (Brown and Ward-Griffin, 1994) and e-learning settings (Kidney et al., 2007) especially when various stakeholders (even adversaries as suggested by Worthen et al. (1997) need to find agreement on quality criteria. Also Brent and Felder (2004) suggest to move peer review of teaching to a departmental or faculty level.

Peer evaluation works well in an environment that embraces a culture of collective evaluation. Ultimately peer evaluation operates within a culture where colleagues engage their peers to improve the quality of their work and increase their productivity on various possible levels of formality (Wiegers, 2002). However, evaluators do not necessarily need to be fellow staff members. The usability experts alluded to above, are often recruited from outside to perform heuristic evaluation. They are typically supplied with detailed checklists of accepted usability principles allowing them to identify problems which intended end users may also experience. Some distinguish from this an 'expert review.' Here the evaluators are better qualified and already know and inherently understand the heuristics. Because of this, reviewers do not need to be supplied with checklists. As a result, the expert review tends to be less formal

(www.usability.gov) allowing more open-ended approaches. Expert review may even employ the recruitment of professional evaluation agencies (Averch, 2004).

The current study was a compromise. In order to allow for an open-ended, interpretivist approach, experts had to be sufficiently qualified to give feedback on the course without receiving any further details on what evaluation criteria to use. Experts were either from the own institution, but not necessarily close colleagues. They were also fellow academics from other institutions. Part of the agreement of taking part in this study was an agreement that there was no expectation of payment. This exchange of academic give and take affords the opportunity of having access to highly qualified evaluators while still being able to do evaluative research without a huge research budget.

5.3.3. Structured or open approach?

The planned purpose of expert course evaluation not only determines what kind of expertise is needed but also the way the evaluation is to be conducted. For the current study a very open-ended interpretive approach was taken. Would this work, for example, as an appraisal process to be used by faculty to make personnel decisions?

Brent and Felder (2004) as well as the University of Minnesota Center for Teaching and Learning (2013) point out that peer review can take possible forms: summative (to provide data to be used in personnel decisions or awards) and formative (to improve teaching). Also York's Senate Committee on Teaching and Learning (2002) warn that it is crucial to keep summative evaluation and formative assessment strictly apart and that informal peer observation of teaching in particular has limitations for summative purposes. The University of Minnesota Center for Teaching and Learning (2013) state that summative peer observation can be used to evaluate teaching effectiveness for the use of merit, promotion, and/or tenure decisions. This however requires a planned and systematic approach to reduce bias and unreliability. They suggest checklists, rating scales, and open-ended narratives as suitable instruments. Also Brent and Felder (2004) suggest developing class observation and course material rating forms. They go further suggesting the formation of a departmental peer review committee (a cadre of raters). Summative evaluation may of course go beyond peer evaluation and may include measuring the performance outcomes of students experiencing the program (Tuckman, 1999) or student feedback. Peer rating should be but one of several sources of information to be included in a teaching portfolio (Brown and Ward-Griffin, 1994).

An open-ended, interpretivist approach is generally not recommended in the literature for any summative purposes. But what could be its place? Berk (2005) recommends that peer observation data should be used for formative rather than for summative decisions. Both the University of Minnesota Center for Teaching and Learning (2013) and Brent and Felder (2004) suggest modifying the summative procedure for formative peer review. More open-ended methods are acceptable in this context. For example simple note taking about what is taking place during the class can be a valuable prompt for

discussion between evaluator and instructor (University of Minnesota Center for Teaching and Learning, 2013). Swinglehurst et al. (2008) suggest an adaptation of a traditional classroom observation for use in the online environment.

The open-ended approach described here provided suggestions that were neither obtained by student feedback nor by self-reflection previously. As this study has shown, an open-ended, interpretivist approach worked well in a context of formative course evaluation. Rather than predetermining a path of evaluation with the use of a structured feedback form an open-ended approach was taken. Being given as little guidance as possible, they took the opportunity to come up with their own suggestions. This approach may well work in the context of evaluating other e-learning courses as well both during more explorative course development and, as here, for the evaluation of established courses, where there is an element of uncertainty about the effectiveness of the programme (Averch, 2004).

5.3.4. What has been learnt about the relevance of this research in other settings?

A crucial issue in any study is whether it is able to move beyond local issues. One aspect of this is the obligation to create theory, thereby generating solutions that have educational significance and wider applicability (McMillan, 2010). Rees and Monrouxe (2010) argue that clarity on the theoretical perspectives behind the research aids in the transferability and generalizability of findings, meaning that other situations and contexts within which the results are likely to be relevant and applicable can be identified. In the context of evaluation research, Pratt (2006) states that evaluation reports benefit from the development of theory, contributing to a science of education.

While the current study aimed to make very practical recommendations for a particular course, it was conducted as research, aimed at an outside audience and thereby moving beyond local issues. It specifically confirmed the appropriateness of an open-ended approach to expert evaluation. Based on the results here and on information from published literature it appeared that a combination of five subject matter and five e-learning experts may work well for other e-learning courses where formative feedback is sought. It is felt that making use of peers in the evaluation process not only diminishes cost but also encourages collaboration with experts in the field. This may be particular pertinent in an African setting where there is a benefit to working together on providing a much-needed contribution to local African solutions in particular on the use of information technology in medical education (Greysen et al., 2011). This includes the courage, as shown here, to report openly on challenges faced in developing countries (Conradie and Roodt, 2004) while implementing as much quality e-learning as possible within a given possibly non-optimal context.

5.4. Summary of chapter 5

The discussion interpreted the evaluative research of expert feedback on an on-line immunology course with respect to implementing suggestions from the peers for the particular course in question but also to make suggestions for the evaluation of other e-learning courses.

While some weaknesses and limitations were pointed out, the methodological approach was deemed sound overall yielding trustworthy results. The interpretivist, open-ended approach chosen did indeed provide local answers for course improvement. Also the choice of five subject and five e-learning experts appeared to work as described in the context of usability testing. One drawback of the study was the fact that only three subject matter experts provided feedback in time for write-up. The evaluation was conducted as evaluative research and together with information from published sources recommendations for others could be made: An interpretivist, open-ended approach using five subject and five e-learning experts, while not allowing an exhaustive review of all course materials, will likely give valuable information on quality for formative assessment of other e-learning courses. For a summary of the main discussion points please also refer to chapter 6 'conclusions.'

6. Conclusions

This study asked whether an open-ended, interpretivist approach to expert evaluation could provide suggestions for improvement for an on-line course in Clinical Immunology. The answer was affirmative. Two groups of experts, subject matter experts and e-learning experts were approached. Both groups made valuable suggestions. Subject matter experts indicated that the course materials were of good quality and adequate on a postgraduate level. E-learning experts expressed concern about the ability of the course to facilitate learning and also identified some usability issues.

Despite some methodological limitations, the findings were deemed trustworthy as biases during the sampling process, data collection and analysis were limited. A marked difference in the response rate between subject matter and e-learning experts was noted and it can be speculated that the resistance in the subject matter expert group could have been due to apprehensions about qualitative research and teaching. Eventually most of the expert feedback was obtained from colleagues from the same institution, which may indicate some response bias.

A first step will of course be the implementation of some of the changes suggested by the experts in this study. It may well be worth conducting this in the form of research, with the aim of informing the community of instructors in developing countries with an example of how to put Information and communications technologies into practice under circumstances that are more challenging than those in developed countries.

The investigation attempted to extend evaluation into the field of evaluative research and thus to go beyond description and make recommendations for course evaluation in other settings. The choice of subject matter plus e-learning experts here was somewhat arbitrary but may be helpful in similar situations of evaluating on-line courses where dual expertise in not readily available. Feedback from five subject matter experts, while not allowing an exhaustive review of all course materials, should give a good idea about the appropriateness for the intended target audience. The feedback from e-learning experts should identify the most salient issues around facilitating of learning and usability.

The open-ended interpretivist approach may work well for courses that are still in development or where an amount of uncertainty about teaching effectiveness exists. Evaluators require a good deal of expertise on their part going beyond what is needed for typical structured peer observation of teaching approaches because there is little guidance. However, rather than imposing evaluative criteria on the experts, the open-ended approach allowed them here to develop their own framework and this may work in other scenarios too.

Should any further expert review for the current course be undertaken in the future, this framework may serve as a guide. It will have to be explored how procedures can be simplified to find ways to sustain expert evaluation of the course as a routine quality measure.

Another important future step will be publishing the findings of the current study in a peer-reviewed journal. This will create an awareness of a number of issues. Firstly it may inspire others to conduct evaluative research aiming at an outside audience as well rather than leaving it at an in-house improvement of teaching quality. Secondly it may encourage the information and communication technology field to embrace more open-ended approaches of research and evaluation. Thirdly it is hoped that research such as this will help others to take a positive view of quality assurance and creating a culture of co-operative evaluation. Finally it is hoped that this will encourage collaboration with peer experts in the field. As pointed out above, this may be especially important in an African setting where there a particular need to work together on providing much-needed contributions to local African solutions on the use of information technology in medical education.

References

Ally M (2004) Foundations of educational theory for online learning. Chapter 1 in: Anderson T, and Elloumi F (eds.) *Theory and practice of online learning*. Athabasca University online book.

Anderson J, and McCormick R (2006). A common framework for e-learning quality. In: McCluskey A (ed.) *Policy and Innovation in Education. Quality Criteria*. European Schoolnet. Brussels.

Anderson T (2004) Toward a theory of online learning. Chapter 2 in: Anderson T, and Elloumi F (eds.) *Theory and practice of online learning*. Athabasca University online book.

Anderson T, and Elloumi F (2004) *Theory and practice of online learning*. Athabasca University online book.

Ashton RH (1986). Combining the judgments of experts: How many and which ones? *Organizational Behavior and Human Decision Processes* 38, 405–414.

Averch H (2004) Using expert judgment. Chapter 10 in: Wholey J, Hatry H, and Newcomer K (eds.) *Handbook of Practical Program Evaluation* (2nd ed.). Jossey-Bass. San Francisco. USA.

Bell M (2005) *Peer observation partnerships in higher education*. HERDSA Guide. HERDSA, NSW, Milperra, Australia.

Berk RA (2005) Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education* 17(1), 48-62.

Biswas S, Basu PN, and Chowdhury T (2013). A feedback system based self-evaluation framework for quality assurance of multi-modal digital distance education. *International Journal of E-Education, E-Business, E-Management and E-Learning* 3(1), 57-61.

Bloch KJ (1994) What is a clinical immunologist? *Pediatric Clinics of North America* 41(4), 591-596.

Bower M (2008) Affordance analysis - matching learning tasks with learning technologies. *Educational Media International* 45(1), 3-15.

Brent R, and Felder R (2004) A protocol for peer review of teaching. *Proceedings of the American Society for Engineering Education Annual Conference and Exposition*, Session 3530.

Brooke J (1996) SUS- A quick and dirty usability scale. In: Jordan PW, Weerdmeester B, Thomas A, and McClelland IL (eds.). *Usability Evaluation in Industry*. Taylor and Francis. London.

Broom A, and Willis E (2007) Competing paradigms and health research. Chapter 2 in: *Researching health: Qualitative, quantitative and mixed methods*. Saks M, and Allsop J (eds.) Sage. London.

Brown B, and Ward-Griffin C (1994) The use of peer evaluation in promoting nursing faculty teaching effectiveness: a review of the literature. *Nurse Education Today* 14(4), 299–305.

Bunniss S, and Kelly DR (2010) Research paradigms in medical education research. *Medical Education* 44, 358-366.

Castro J, and Teniente E (2005) *Proceedings of the CAiSE'05 Workshops* – Volume I. FEUP, Porto, Portugal.

Center for Research on Learning and Teaching (2014) Guidelines for evaluating teaching. Center for Research on Learning and Teaching (CRLT) University of Michigan. Available at: http://www.crlt.umich.edu/tstrategies/guidelines.php

Chao C-J, and Salvendy G (1994) Percentage of procedural knowledge acquired as a function of the number of experts from whom knowledge is acquired for diagnosis, debugging, and interpretation tasks. *International Journal of Human-Computer Interaction* 6(3), 221-233.

Chua A, and Lam W (2007) Quality assurance in online education: The Universitas 21 Global approach. *British Journal of Educational Technology* 38(1), 133–152.

Conradie DP, and Roodt J (2004) Realities versus ideals with regard to e-learning in South Africa. *Online educa Berlin, Germany: 10th international conference on technology supported learning and training*. Book of abstracts.

Creighton T (2003) Resisters and saboteurs. In: *The Principal as Technology Leader*. Corwin Press. Thousand Oaks, USA.

Davids MR, Chikte UME, and Halperin ML (2011) Development and evaluation of a multimedia e-learning resource for electrolyte and acid-base disorders. *Advances in Physiology Education* (25), 295–306.

De Villiers MR (2005) Interpretative research models for informatics: Action research, grounded theory, and the family of design- and development research. *Alternation* 12(2), 10–52.

De Villiers MR (2007) An action research approach to the design, development and evaluation of an interactive e-learning tutorial in a cognitive domain. *Journal of Information Technology Education* 6, 455-479.

Denscombe M (2010) *The good research guide for small scale social research projects* (4th ed.). Open University Press, McGraw-Hill Education. Maidenhead.

Department of Sustainability and Environment (2005) *Effective engagement: building relationships with community and other stakeholders. Book 3. The engagement toolkit*. The Community Engagement Network Resource and Regional Services Division. Melbourne, Australia.

Dooley LM, and Lindner JR (2003) The handling of nonresponse error. *Human Resource Development Quarterly* 14(1), 99-110.

Ellaway R, and Masters K (2008) AMEE Guide 32: E-learning in medical education Part 1: Learning, teaching and assessment. *Medical Teacher* 30(5), 455–73.

Felder R, and Brent R (2002) Random thoughts… How to evaluate teaching. *Chemical Engineering Education* 38(3), 200–202.

Fernandez A, Insfran E, and Abrahão S (2011). Usability evaluation methods for the web: A systematic mapping study. *Information and Software Technology* 53(8), 789–817.

Fowler FJ (2009) *Survey research methods* (4th ed.) Sage.

Fraenkel JR, and Wallen NE (2009) *How to design and evaluate research in education* (7th ed.). McGraw-Hill Higher Education.

Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, Fineberg H, Garcia P, Ke Y, Kelley P, Kistnasamy B, Meleis A, Naylor D, Pablos-Mendez A, Reddy S, Scrimshaw S, Sepulveda J, Serwadda D, and Zurayk H (2010) Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet* 376 (9756), 1923–58.

Goldenkoff R (2004) Using focus groups. Chapter 12 in: Wholey JS, Hatry HP, and Newcomer KE (eds.) *Handbook of practical program evaluation* (2nd ed.). Jossey-Bass. San Francisco.

Goldie J (2006) AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher* 28(3), 210–24.

Grant J (2006) Principles of curriculum design. In: Swanwick T (ed.) *Understanding medical education*. Association for the Study of Medical Education (ASME). Edinburgh.

Gravett S, and Geyser H (2003) *Teaching and learning in higher education*, 1st edition. Van Schaik Publishers, Pretoria.

Greysen SR, Dovlo D, Olapade-Olaopa EO, Jacobs M, Sewankambo N, and Mullan F (2011). Medical education in sub-Saharan Africa: a literature review. *Medical Education* 45(10), 973–86.

Groves R, Cialdini R, and Couper M (1992) Understanding the decision to participate in a survey. *Public Opinion Quarterly* 56, 475-495.

Harden RM, and Crosby JR (2000). AMEE Guide No 20: The good teacher is more than a lecturer-the twelve roles of the teacher. *Medical Teacher* 22(4), 334-347.

Higher Education Quality Committee (2004) *Criteria for programme accreditation*. Council on Higher Education (CHE). Available at: http://www.che.ac.za/media_and_publications/frameworks-criteria/criteria-programme-accreditation

Horn L (2011) *Guidelines for new researchers*. Stellenbosch University, Faculty of Medicine and Health Sciences, Research Development and Support. Available at: http://sun025.sun.ac.za/portal/page/portal/Health_Sciences/English/Centres%20and%20Institutions/Research_Development_Support/Ethics/Application_package/Research%20Booklet%2014%2003%202011.pdf

House ER (1976) Justice in evaluation. In: Glass GV (ed.) *Evaluation Studies Review Annual*, Vol. 1. Sage. Cited in: Goldie J (2006) AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher* 28(3), 210-224.

ICT Task Team (2013) *Strategy for the use of ICT in learning and teaching at Stellenbosch University*. Available at: http://www.sun.ac.za/english/Documents/Rector/ICT%20in%20Teaching%20and%20Learning.pdf

Jansen JD (2007) The research question. Chapter 1 in: Maree K (ed.) *First Steps in Research*. Van Schaik. Pretoria.

Jara M, Mohamad F, and Cranmer S (2008) Evaluation of e-learning courses. *WLE Centre, Occasional Paper* 4, Institute of Education, University of London.

Jordan PW, Weerdmeester B, Thomas A, and McClelland IL (1996). *Usability Evaluation in Industry*. Taylor and Francis. London.

Jung I (2011) The dimensions of e-learning quality: from the learner's perspective. *Educational Technology Research and Development* 59(4), 445–464.

Kern DE, Thomas PA, Howard DM, and Bass EB (1998) *Curriculum development for medical education. A six-step approach.* The Johns Hopkins University Press. Baltimore.

Kidney G, Cummings L, and Boehm A (2007) Toward a quality assurance approach to e-learning courses. *International Journal on E Learning* 6(1), 17–30.

Lambert PH, Metzger H, and Myamoto T (1993) Clinical Immunology: Guidelines for its organization, training, and certification. Relationships with allergology and other medical disciplines: A WHO/IUIS/ IAACI Report. *Clinical Immunology and Immunopathology* 69(2), 234–243.

Liebrich W, and Esser M (2014) Establishing an online short-course in clinical immunology. Sixth National Conference of the South African Association of Health Educationalists (SAAHE) 'Information to Transformation', Umhlanga, 27 - 29 June 2013. *African Journal of Health Professions Education* 6(1), 69-113.

Marcus B, and Schütz A (2005) Who are the people reluctant to participate in research? Personality correlates of four different types of nonresponse as inferred from self-and observer ratings. *Journal of Personality* 73(4), 959-84.

Maree K (2007) *First steps in research*. 1st edition. Van Schaik Publishers. Pretoria.

Maree K, and Pietersen J (2007) Surveys and the use of questionnaires. Chapter 9 in: Maree K (ed.) *First Steps in Research*. Van Schaik, Pretoria.

Maree K, and van der Westhuizen D (2007) Planning a research proposal. Chapter 3 in: Maree K (ed.) *First steps in research*. 1st edition. Van Schaik Publishers. Pretoria.

McCluskey A (2006) *Policy and Innovation in Education. Quality Criteria*. European Schoolnet. Brussels.

McMillan W (2010) Moving beyond description: Research that helps improve teaching and learning. *African Journal of Health Professions Education* 2(1) 3-7.

Medical and dental professions board committee for medical science (2010) *Policy regarding intern medical scientists and training of intern medical scientists*. Health Professions Council of South Africa (HPCSA) Document MSC A.

Mehrens WA (1991) *Measurement and evaluation in education and psychology*. 4th ed. Holt, Rinehart and Winston. Cited in: Goldie J (2006) AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher* 28(3), 210-224.

Muraskin L (1997) *Understanding evaluation: The way to better prevention programs*. Westat. Cited in: Goldie J (2006) AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher* 28(3), 210-224.

Newble D, and Cannon R (1994) A handbook for medical teachers. 3rd ed. Kluwer Academic Publishers.

Newcomer KE, Hatry HP and Wholey JS (2004) Meeting the need for practical evaluation approaches: An introduction. In: Wholey JS, Hatry HP, and Newcomer KE (editors) *Handbook of practical program evaluation*. 2nd edition. Jossey-Bass. San Francisco

Nielsen J (1992). Finding usability problems through heuristic evaluation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* - CHI '92, 373–380.

Nielsen J (1994) Heuristic evaluations. In: Nielsen J, and Mack RL (eds.) *Usability inspection methods*. John Wiley and Sons. Cited in: Ssemugabi S (2006) Usability evaluation of a web-based e-learning application: a study of two evaluation methods. Masters of Science in Information systems thesis. University of South Africa.

Nieuwenhuis J (2007a) Qualitative research designs and data gathering. Chapter 5 in: Maree K (ed.) *First Steps in Research*. Van Schaik, Pretoria.

Nieuwenhuis J (2007b) Analysing qualitative data. Chapter 6 in: Maree K (ed.) *First Steps in Research*. Van Schaik, Pretoria.

Patton MQ (2002) *Qualitative research and evaluation methods*. 3rd ed. Sage Publications, Thousand Oaks.

Pawlowski JM (2003) The European Quality Observatory (EQO): Structuring quality approaches for e-learning. *Proceedings 3rd IEEE International Conference on Advanced Technologies* 209–213.

Peterson P, Baker E, and McGaw B (2010) *International Encyclopedia of Education* (3rd ed.). Elsevier.

Posner GJ (2004) Concepts of curriculum and purposes of curriculum study. Chapter 1 in: *Analyzing the curriculum* (3rd ed.). McGraw-Hill. Boston.

Powell RR (2006) Evaluation research: An overview. *Library Trends* 55(1), 102-120.

Pratt N (2006) Evaluation research in education. Originally prepared by Silver H, Faculty of Education, University of Plymouth, 2004. Available at: http://www.edu.plymouth.ac.uk/resined/evaluation/

Preece J, Rogers Y, and Sharp H (2002) *Interaction design: Beyond human computer interaction* (1st ed.). Wiley.

Radhakrishna R, and Doamekpor P (2008) Strategies for generalizing findings in survey research. *Journal of Extension* 46(2), Article Number 2TOT1.

Recker J (2005). Conceptual model evaluation. Towards more paradigmatic rigor. In Castro J, and Teniente E (eds.): *Proceedings of the CAiSE'05 Workshops* – Volume I. FEUP, Porto, Portugal, pp. 569-580.

Rees CE, and Monrouxe LV (2010). Theory in medical education research: how do we get there? *Medical Education* 44(4), 334–9.

Ringsted C, Hodges B, and Scherpbier A (2011) 'The research compass': An introduction to research in medical education: AMEE Guide No. 56. *Medical Teacher* 33, 695–709.

Saks M and Allsop J (2007) *Researching health: Qualitative, quantitative and mixed methods*. Sage. London.

Schultz KK and, Latif D (2006) The planning and implementation of a faculty peer review teaching project. *American Journal of Pharmaceutical Education* 70(2), 32.

Senate Committee on Teaching and Learning (2002) The teaching assessment and evaluation guide. Published by the Senate Committee on Teaching and Learning (SCOTL), York University. Available at: www.yorku.ca/secretariat/senate/committees/scotl/

Shadish WR, Cook TD, and Campbell DT (2002) Experimental and quasiexperimental designs for generalized causal inference. In: Sivo S, and Saunders C (2006) How low should you go? Low response rates and the validity of inference in IS questionnaire research. *Journal of the Association for Information Systems* 7(6), 351–414.

Shearer WT (2002) Recognition of clinical immunology as a distinct medical subspecialty: importance for the practice of allergy. *Journal of Allergy and Clinical Immunology* 110(4), 567-70.

Siemens G, and Tittenberger P (2009) *Handbook of emerging technologies for learning*. Learning Technologies Center, University of Manitoba, Canada, electronic book.

Sivo S and, Saunders C (2006) How low should you go? Low response rates and the validity of inference in IS questionnaire research. *Journal of the Association for Information Systems* 7(6), 351–414.

Ssemugabi S (2006) Usability evaluation of a web-based e-learning application: a study of two evaluation methods. Masters of Science in Information systems thesis. University of South Africa.

Stellenbosch University (2012) Learning and teaching policy. Available at: http://www0.sun.ac.za/ctl/wp-content/uploads/2013/02/Learning-and-teaching-policy1.pdf

Suchman EA (1967) *Evaluative research: Principles and practice in public service and social action programs*. 1st ed. Sage. New York.

Swanwick T (2006) *Understanding medical education*. Association for the Study of Medical Education (ASME). Edinburgh.

Swinglehurst D, Russell J, and Greenhalgh T (2008) Peer observation of teaching in the online environment: an action research approach. *Journal of Computer Assisted Learning* 24(5), 383–393.

Tavangarian D, Leypold M, Nölting K, and Röser M (2004) Is e-learning the solution for individual learning? *Journal of e-learning* 2(2), 273-279.

Taylor C, and White S (2000). Making knowledge. Chapter 2 in: Practising reflexivity in health and welfare. Buckingham: Open University Press.

Tuckman BW (1999) Conducting evaluation studies. Chapter 13 In: *Conducting Educational Research* (5th ed.). Harcourt Brace College Publishers. USA.

University of Minnesota Center for Teaching and Learning (2013) Peer observation guidelines and recommendations. Available at: http://www1.umn.edu/ohr/teachlearn/resources/peer/guidelines/

Van der Westhuizen D (2003) The design and development of a web-based learning environment. Chapter 10 in: Gravett S, and Geyser H (eds.) *Teaching and learning in higher education*. 1st ed. Van Schaik Publishers. Pretoria.

Van Ort S, Noyes A, and Longman A (1986) Developing and implementing a model for evaluating teaching effectiveness. *Journal of Nursing Education* 22(8), 324-328. Cited in: Brown B, and Ward-Griffin C (1994) The use of peer evaluation in promoting nursing faculty teaching effectiveness: a review of the literature. *Nurse Education Today* 14(4), 299–305.

Wang Q, Woo H L, and Zhao J (2009) Investigating critical thinking and knowledge construction in an interactive learning environment. *Interactive Learning Environments* 17(1), 95–104.

Wholey JS, Hatry HP, and Newcomer KE (2004) *Handbook of practical program evaluation* (2nd ed.). Jossey-Bass. San Francisco.

Wiegers KE (2002) A little help from your friends. Chapter 2 in: *Peer Reviews in Software: A Practical Guide*. Addison-Wesley.

Williams DD, and Graham CR (2010) Evaluating e-learning. Pp 530–538 in: Peterson P, Baker E, and McGaw B (eds.). *International Encyclopedia of Education* (3rd ed.). Elsevier.

Witteck T, Most B, Kienast S, and Eilks I (2007) A lesson plan on 'methods of separating matter' based on the Learning Company Approach – A motivating frame for self-regulated and open lab-work in introductory secondary chemistry lessons. *Chemistry Education Research and Practice* 8(2), 108-119.

Woolrych A, and Cockton G (1986) Why and when five test users aren't enough. Organizational Behavior and Human Decision Processes 38, 405-414. Re-print in: *Proceedings of IHM-HCI 2001 Conference*.

World Federation for Medical Education (2003) Basic medical education WMFE global standards for quality improvement. WMFE Office: University of Copenhagen, Denmark.

Worthen BL, Sanders JR, and Fitzpatrick JL (1997) Program Evaluation: Alternative Approaches and Practical Guidelines. 2nd ed. Longman. Cited in: Goldie J (2006) AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher* 28(3), 210-224.

Wyllie A (2011) Eager 'weavers': designing assessment for an online environment. *Nurse education in practice* 11(2), 99-103.

Websites, visited August 2014:

British Society for Immunology. Clinical careers in immunology: https://www.immunology.org/page.aspx?pid=1273

Health Research Ethics Committee (HREC) of the Faculty of Medicine and Health Sciences at Stellenbosch University: http://sun025.sun.ac.za/portal/page/portal/Health_Sciences/English/ Centres%20and%20Institutions/Research_Development_Support/Ethics

Medical Informatics. Discipline of Information Systems & Technology. University of Kwazulu-Natal: http://is.ukzn.ac.za/Courses/medicalinformatics.aspx

Merriam Webster dictionary: www.merriam-webster.com/dictionary/expert

Recruiting usability test participants. Usability.gov: http://www.usability.gov/how-to-and-tools/methods/recruiting-usability-test-participants.html

Short-courses Division. Stellenbosch University. Clinical Immunology short course - An introduction to the immune system for clinicians: http://shortcourses.sun.ac.za/courses/3615.html

Social Science Statistics. Fisher exact test calculator: http://www.socscistatistics.com/tests/fisher/Default2.aspx

University of Edinburgh, MSc in Digital Education: http://online.education.ed.ac.uk/

Appendices

Appendix 1: Materials provided to experts on contact

Email:

Dear colleagues,
I am Walter Liebrich from the Faculty of Medicine and Health Sciences at Stellenbosch University.
Some of you I know well, others not at all. Your names have been drawn at random to participate in an expert evaluation of an on-line short project for a Master in Philosophy in Health Science Education.
Please find attached an electronic consent form and some information about the project.
Please email the consent form back to me. I will then organise access to the moodle-course for you to initiate the process. You will have a ch MS Lync; Skype if the majority of you prefer this).
Should you not wish to participate, please indicate this to me by email soon and suggest an alternative e-learning expert.

ïind regards!
ᵗer Liebrich

The protocol synopsis:

Protocol Synopsis, Dr Walter Liebrich

Title
Expert evaluation of an online course in Clinical Immunology

Introduction, motivation and literature overview
We are offering an online course "Clinical Immunology short course: An introduction to the immune system for clinicians". While student feedback and self-reflection by the course co-ordinators proved helpful in re-designing the course, one aspect of course assessment had ᵒot been explored yet: an evaluation by independent experts. This will be the focus of the ᵉsted study.

…

The electronic participant information leaflet and consent form:

TITLE OF THE RESEARCH PROJECT: Expert evaluation of an online cours...
Clinical Immunology

REFERENCE NUMBER: S13/11/232

PRINCIPAL INVESTIGATOR: Walter Liebrich
Medical Microbiology and FMHS e-learning at Stellenbosch University (SU)

ADDRESS:
e-learning office
Faculty of Medicine and Health Sciences (FMHS)
PO Box 19063
Tygerberg 7505

CONTACT NUMBER: 021-938-9190
CONTACT E-MAIL: liebrich@sun.ac.za

My name is Walter Liebrich. I am offering a registered online short-course in clinical
...munology. I would like to invite you to participate in a research project that aims to
...tigate expert evaluation of this short course. This research also forms part of my
...h assignment within the Master of Philosophy in Health Science Education.

Feedback forms for subject matter and e-learning experts:

FEEDBACK FORM FOR SUBJECT-MATTER EXPERTS

TITLE OF THE RESEARCH PROJECT: Expert evaluation of an online course in Clinical Immunology

PRINCIPAL INVESTIGATOR: Dr Walter Liebrich

PARTICIPANT DETAILS		
First name:	Initials:	Surname:

Background:

In 2005 a one-month practical immunology bench rotation was implemented in the Immunology
...nit of the National Health Laboratory Service (NHLS) Tygerberg / Stellenbosch University Division of
...ical Microbiology, aimed initially at Pathology Registrars. There is typically one rotator at any

Appendix 2: Coding approach

Printouts of transcripts; key statements marked; initial codes and categories suggested:

Key statements simplified and entered into tables along with suggestive codes:

Collective key statements from both expert groups transferred into an overview table; an independent researcher was brought in who commented on the coding strategy and the suggested categories: