What is Meant by Educational Evaluation and Research?

4.1 Introduction
Now that we’ve described the LEPO framework, let’s examine the process of evaluation in more detail. In section 1.4, we summarized various evaluation scenarios in higher education, concluding that our focus in this book is on evaluating the effectiveness of e-learning, but not evaluating strategic initiatives, selecting products or resources, evaluating teaching quality, nor evaluating the sustainability of broader e-learning initiatives.

This chapter provides a brief overview of the literature on evaluation, with a summary of program evaluation, and some definitions that will be helpful in the presentation of the following arguments. Similarly, we review various approaches to evaluation, which then support our subsequent and more specific discussion of e-learning evaluation across the e-learning life cycle. This analysis will establish that there is clearly a research component to studies that consider the effectiveness of e-learning. This leads us into a discussion of how ‘research’ is interpreted differently in different disciplines, and an analysis of the broad types of research which are appropriate for e-learning. This sets the scene for further discussion of this issue in Chapter 6.

4.2 Educational Evaluation
As we noted in Chapter 2, human beings naturally ask questions about how valuable their activities are. We can think of evaluation as a process of considerably sharpening this natural activity of checking on our ongoing work. Chapter 1 introduced the idea that the term ‘evaluation’ is used in different ways in different contexts. This chapter takes that discussion a little further by seeking to unpack the characteristics of evaluation in an education context, in order to identify clearly the scope of this work, and discuss overlaps with other areas of scholarship.
4
What is Meant by Educational Evaluation and Research?

4.1 Introduction
Now that we’ve described the LEPO framework, let’s examine the process of evaluation in more detail. In section 1.4, we summarized various evaluation scenarios in higher education, concluding that our focus in this book is on evaluating the effectiveness of e-learning, but not evaluating strategic initiatives, selecting products or resources, evaluating teaching quality, nor evaluating the sustainability of broader e-learning initiatives.

This chapter provides a brief overview of the literature on evaluation, with a summary of program evaluation, and some definitions that will be helpful in the presentation of the following arguments. Similarly, we review various approaches to evaluation, which then support our subsequent and more specific discussion of e-learning evaluation across the e-learning life cycle. This analysis will establish that there is clearly a research component to studies that consider the effectiveness of e-learning. This leads us into a discussion of how ‘research’ is interpreted differently in different disciplines, and an analysis of the broad types of research which are appropriate for e-learning. This sets the scene for further discussion of this issue in Chapter 6.

4.2 Educational Evaluation
As we noted in Chapter 2, human beings naturally ask questions about how valuable their activities are. We can think of evaluation as a process of considerably sharpening this natural activity of checking on our ongoing work. Chapter 1 introduced the idea that the term ‘evaluation’ is used in different ways in different contexts. This chapter takes that discussion a little further by seeking to unpack the characteristics of evaluation in an education context, in order to identify clearly the scope of this work, and discuss overlaps with other areas of scholarship.
4

What is Meant by Educational Evaluation and Research?

4.1 Introduction

Now that we’ve described the LEPO framework, let’s examine the process of evaluation in more detail. In section 1.4, we summarized various evaluation scenarios in higher education, concluding that our focus in this book is on evaluating the effectiveness of e-learning, but not evaluating strategic initiatives, selecting products or resources, evaluating teaching quality, nor evaluating the sustainability of broader e-learning initiatives.

This chapter provides a brief overview of the literature on evaluation, with a summary of program evaluation, and some definitions that will be helpful in the presentation of the following arguments. Similarly, we review various approaches to evaluation, which then support our subsequent and more specific discussion of e-learning evaluation across the e-learning life cycle. This analysis will establish that there is clearly a research component to studies that consider the effectiveness of e-learning. This leads us into a discussion of how ‘research’ is interpreted differently in different disciplines, and an analysis of the broad types of research which are appropriate for e-learning. This sets the scene for further discussion of this issue in Chapter 6.

4.2 Educational Evaluation

As we noted in Chapter 2, human beings naturally ask questions about how valuable their activities are. We can think of evaluation as a process of considerably sharpening this natural activity of checking on our ongoing work. Chapter 1 introduced the idea that the term ‘evaluation’ is used in different ways in different contexts. This chapter takes that discussion a little further by seeking to unpack the characteristics of evaluation in an education context, in order to identify clearly the scope of this work, and discuss overlaps with other areas of scholarship.
4.2.1 Defining Educational Evaluation

There are numerous definitions of ‘evaluation’ in the literature and it can be difficult to find a universally accepted view of what it is. Regardless, most definitions of evaluation encompass notions of making judgements about the value or worth of something. For example, the Joint Committee on Standards for Educational Evaluation defined evaluation as the “systematic investigation of the worth or merit of an object” (1994). With such a definition there is an implication that judgements are made against a pre-defined set of standards or criteria. A drawback of this assumption is that unanticipated issues which arise from any enterprise are in danger of being overlooked if evaluation is based on preordained criteria. More recent definitions have gone beyond ‘merit or worth’ to encompass “production of knowledge based on systematic enquiry to assist decision-making” (Owens, 2006, p. 18.) about whatever is being evaluated. A further definition that we find useful because it recognizes the stages of a project is: “Evaluation is any activity that throughout the planning and delivery of innovative programmes enables those involved to learn and make judgements about the starting assumptions, implementation processes and outcomes of the innovation concerned” (Stern, 1990, cited in Jackson, 1998).

Evaluation is an applied form of social research, “the primary purpose of which is not to discover new knowledge, as is the case with basic research, but to study the effectiveness with which existing knowledge is used to inform and guide practical action” (Clarke & Dawson, 1999, p. 2). Evaluation employs research techniques as a means of systematically generating the necessary information, and uses similar criteria to judge the quality of the evidence. However, it differs from other forms of social research in its audience, goals and its practical orientation (Clarke & Dawson, 1999; Trochim, 2006), as well as in making judgements about the value of what is being evaluated.

One of the more common evaluation approaches is called program evaluation which has a substantial literature base, particularly in the USA (cf. Clarke & Dawson, 1999; Guba & Lincoln, 1989; Owen, 2006; Patton, 1990, 1997; Payne, 1994; Worthen, Sanders & Fitzpatrick, 1997). Program evaluation is typically used to systematically evaluate initiatives – or programs – that are large-scale activities designed to elicit broad changes in society. For example, a program evaluation may be used to determine the value and worth of initiatives in public health or the criminal justice system, and the example of an institution gathering evidence about the implementation its e-learning strategic plan provided in section 1.4 is a case of program evaluation within education.

While our explanation of program evaluation could suggest that most educational evaluations will fit within this approach, this is not the case, as not all educational innovations can be seen as ‘programs’. For example, many e-learning initiatives are more specific in their focus and are better considered as relatively discrete projects rather than broad educational programs. Projects
are temporary activities, with fixed end-dates and budgets and specific goals to produce a single deliverable (Baume, Martin & Yorke, 2002). Projects can vary in scale. Developments in interactive learning systems may involve substantial amounts of funding over years, while some learning environments produced with generic learning tools may be developed by individual practitioners with minimal funding and a lack of formality. Even though program and project evaluations may use similar methods, the questions asked of ongoing programs are likely to be different from those asked about individual educational innovation projects.

4.2.2 Types of Evaluation

Many different judgements can be made in evaluation, which has led to a range of evaluation types and models, each with different goals. We will not discuss these in detail here and interested readers are directed to standard textbooks on this topic (for example, Worthen et al., 1997). However, we think it would be useful to introduce briefly some general types of evaluation that will underpin the development of our arguments in later chapters.

A long-standing distinction in evaluation has been between formative and summative evaluation (Worthen et al., 1997). The fundamental goal of formative evaluation is to provide information to refine and improve the object being studied. Formative evaluation looks for weaknesses in the design of an innovation and in the processes associated with its implementation. In an e-learning context, this information can assist the ongoing design, development and implementation process. Summative evaluations focus on the outcomes of the innovation (project) and seek to determine a project’s value and worth by determining “its overall effectiveness or impact” (Clarke & Dawson, 1999, p. 8). Evaluators have found the characterizing of these two types of evaluation helpful in the context of the e-learning life cycle, as one type can be aligned with the design and development of innovation (formative) and the other can aligned with an innovation after it has been implemented and used (summative).

However, this simple dichotomy has been found to be inadequate for the complexity of modern evaluation practice and the multifaceted nature of e-learning. A number of researchers have developed taxonomies that delineate different types of evaluation of e-learning (Alexander & Hedberg, 1994; Bain, 1999; Draper, Brown, Henderson & McAteer, 1996; Gunn, 1999; Phillips et al., 2000; Reeves, 1989; Reeves & Hedberg, 2003; Reeves & Lent, 1984). The nature of these taxonomies is illustrated by some examples. Reeves (1989) suggested six levels of evaluation which include project documentation, assessment of the worth of project objectives, formative evaluation, immediate effectiveness evaluation, impact evaluation and cost effectiveness. Draper et al. (1996) proposed that evaluation of educational technology could have a formative, summative, illuminative or integrative focus as well as fulfilling the purpose of
quality audit, assessment or assurance. Finally, in an update of Reeves’ earlier work, Reeves and Hedberg (2003) articulated a systems-based approach to the evaluation of interactive learning. They proposed an evaluation framework that includes review, needs assessment, formative-, effectiveness- and impact-evaluation.

An important caveat about any overarching evaluation framework is that the division between the types of evaluation it describes is somewhat artificial. That is, the evaluation types are not mutually exclusive, either conceptually or methodologically. A clear and often-cited example of this is that a single evaluation study can fulfil both formative and summative purposes. A practitioner may carry out an investigation that has the dual aims of improving the educational design of an e-learning environment and determining its effectiveness. This notwithstanding, it is still helpful to see each type of evaluation as conceptually distinct, fulfilling different purposes and responding to different evaluation needs.

4.3 e-Learning Evaluation

In this book, we are primarily interested in the effectiveness of e-learning artefacts and learning environments that have been designed using e-learning artefacts. This interest extends across the e-learning life cycle introduced in Chapter 1 and will be illustrated by four scenarios which reflect different phases of the e-learning life cycle. The object being evaluated is the e-learning environment, where this has been designed to facilitate a mixture of learning processes and outcomes.

The following paragraphs will explore these scenarios in more detail, building on concepts established earlier. In summary, these concepts are:

- An e-learning artefact has one of three forms: an interactive learning system; a generic learning tool; or a learning object – each with similar but slightly different characteristics.
- A learning environment is built from one or more e-learning artefacts, and can be contextualized by the designed learning tasks and the desired learning outcomes.
- Learners engage with the learning environment and learning tasks as part of their learning processes.
- Learners can demonstrate learning outcomes.

Table 4.1 outlines four evaluation scenarios which reflect four phases of the life cycle of an e-learning project, from initial exploration, through development, to a mature e-learning environment. We present these scenarios to help structure a discussion of how evaluation can be used across the e-learning life cycle. These scenarios will be revisited at different stages of the book, notably in Chapter 6 and Chapter 8.
quality audit, assessment or assurance. Finally, in an update of Reeves’ earlier work, Reeves and Hedberg (2003) articulated a systems-based approach to the evaluation of interactive learning. They proposed an evaluation framework that includes review, needs assessment, formative-, effectiveness- and impact-evaluation.

An important caveat about any overarching evaluation framework is that the division between the types of evaluation it describes is somewhat artificial. That is, the evaluation types are not mutually exclusive, either conceptually or methodologically. A clear and often-cited example of this is that a single evaluation study can fulfil both formative and summative purposes. A practitioner may carry out an investigation that has the dual aims of improving the educational design of an e-learning environment and determining its effectiveness. This notwithstanding, it is still helpful to see each type of evaluation as conceptually distinct, fulfilling different purposes and responding to different evaluation needs.

4.3 e-Learning Evaluation

In this book, we are primarily interested in the effectiveness of e-learning artefacts and learning environments that have been designed using e-learning artefacts. This interest extends across the e-learning life cycle introduced in Chapter 1 and will be illustrated by four scenarios which reflect different phases of the e-learning life cycle. The object being evaluated is the e-learning environment, where this has been designed to facilitate a mixture of learning processes and outcomes.

The following paragraphs will explore these scenarios in more detail, building on concepts established earlier. In summary, these concepts are:

- An e-learning artefact has one of three forms: an interactive learning system; a generic learning tool; or a learning object – each with similar but slightly different characteristics.
- A learning environment is built from one or more e-learning artefacts, and can be contextualized by the designed learning tasks and the desired learning outcomes.
- Learners engage with the learning environment and learning tasks as part of their learning processes.
- Learners can demonstrate learning outcomes.

Table 4.1 outlines four evaluation scenarios which reflect four phases of the life cycle of an e-learning project, from initial exploration, through development, to a mature e-learning environment. We present these scenarios to help structure a discussion of how evaluation can be used across the e-learning life cycle. These scenarios will be revisited at different stages of the book, notably in Chapter 6 and Chapter 8.
Scenario A refers to the situation where a new technology with educational potential comes on to the market (a relevant example at the time of writing was micro-blogging, e.g. Twitter). It is appropriate for a teacher to evaluate how he or she might use this technology for teaching and how learners might use this for learning. This will involve trialling, exploration and observation, asking, ‘How does this new technology work?’ and ‘Is this tool worth investigating further?’ Initial exploration may be followed by a period of experimentation, asking, ‘How can I use it? Can I use it to do x?’

Scenario B corresponds to the evaluation of the conduct of a project to develop an e-learning artefact from scratch (section 1.6). It may follow from satisfactory outcomes from scenario A, but it may also arise from the identification of a teaching and learning problem that an existing e-learning artefact might be able to address. That is, it may be a learning-driven, rather than technology-driven, activity. Evaluation of scenario B is closely aligned with the development of an e-learning artefact. As such, it maps on to Stern’s (1990) definition of evaluation, which involves making judgements about the starting assumptions, implementation processes and outcomes of the innovation.

There are two aspects to this scenario, evaluating the conduct of the project itself and evaluating the artefact that is the project deliverable. The former,

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
<th>Questions</th>
<th>Judgement/decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Evaluating the potential of a new technology for use in learning and teaching</td>
<td>How does this new technology work? How could it be used in learning and teaching?</td>
<td>Potential use</td>
</tr>
<tr>
<td>B</td>
<td>Evaluating the development of an e-learning artefact</td>
<td>How effective was the conduct of the project? Does the e-learning artefact function technically as designed?</td>
<td>Project quality</td>
</tr>
<tr>
<td>C</td>
<td>Evaluating ways to improve a designed e-learning environment</td>
<td>Does the e-learning environment function as designed? How can it be improved?</td>
<td>Improvement Formative</td>
</tr>
<tr>
<td>D</td>
<td>Evaluating the effectiveness of an e-learning environment</td>
<td>What do learners learn from the e-learning environment? How do learners learn from the e-learning environment?</td>
<td>Impact Summative</td>
</tr>
</tbody>
</table>
which we call project-management evaluation, is concerned with areas such as the quality of the project-management processes, the project documentation and the communication between team members. (We will explore the evaluation of the conduct of an e-learning project more fully in Chapter 10.) The latter is often associated with usability testing and considers questions such as ‘Does the e-learning artefact function as designed and intended?’ and ‘Can learners interact with and navigate around as they need to?’ If a project is well designed and implemented, project-management evaluation will confirm this and the project is likely to produce usable deliverables. However, these two types of evaluation will not directly provide information about the quality of these deliverables, nor their effectiveness in a learning environment. These are the focus of later cycles of evaluation.

As described in section 1.6, e-learning artefacts, whether developed from scratch, sourced from elsewhere or derived from generic learning tools, need to be thoughtfully embedded into learning environments. Scenario C describes a phase in which the integration of the e-learning artefact within a particular learning environment is judged to be sufficiently mature to evaluate in real life in classrooms or virtual environments. It then becomes appropriate to evaluate whether the e-learning environment functions as designed. The goal is to improve the learning environment through formative evaluation, and evaluation centres, not on the technology per se, but on the way learners use the environment to learn. After sufficient cycles of improvement, the evaluation should lead to a decision that the learning environment functions well and can be used with confidence.

The final scenario (D) in Table 4.1 involves evaluating the effectiveness of a mature e-learning environment, which is known – to a greater or lesser extent – to function as it was designed. With this type of evaluation we focus on learners’ learning processes and outcomes, and make judgements about the e-learning environment based on these. As well as contributing to knowledge about the effectiveness of e-learning environments, this type of evaluation will almost certainly identify further ways that the learning environment can be improved, leading to further cycles of development. For example, of the 16 studies reported in Phillips (2002b), while half were more ‘summative’ in nature, every project reported formative results, identifying ways in which the e-learning environment could be improved. In one project (Daniel, Lockwood, Stewart & McLoughlin, 2002), which sought to evaluate summatively the effectiveness of three CD-Roms funded by large grants over several years, the results were unexpectedly formative. The CD-Roms were not used by learners in the way intended, and the development team had to rethink their design, particularly the way in which the CD-Roms were integrated with other teaching activities.

This example reinforces the earlier point about the interdependency of evaluation phases. We can see from the example above that a nominally
summative evaluation can highlight areas for improvement in future cycles of development. Furthermore, it may sometimes be reasonable to combine the evaluation of learning in an e-learning environment with an evaluation of the technical or usability aspects of an e-learning artefact (scenarios B and C). It is also perfectly reasonable to evaluate both improvement and effectiveness (scenarios C and D) at the same time.

While the labelling and ordering of the scenarios in Table 4.1 may suggest that the phases of evaluation should be carried out in a sequential order, this is not the case. There is no need for a lock-step approach to evaluation across the e-learning life cycle. For example, it is relatively common to start an evaluation study such as the one described in scenario C without scenarios A and B being part of a preceding study. This occurs when external evaluators come in to study an e-learning environment developed by a committed teacher. Finally, there is no implication that all four of these phases need be carried out every time a technology-based implementation is considered in a teaching and learning context. In other words, evaluation of e-learning should not be seen as a linear process, but should be viewed as being iterative and context-dependent.

4.3.1 Evaluation, Research and Evaluation Research

The final scenario described above (D) raises a further issue – how an investigation that sets out to ‘evaluate’ an e-learning artefact or environment may, in fact, gather evidence about how and why an e-learning environment works – that is, how it affects learning. An evaluation of the effectiveness of an e-learning environment may quite easily shed light on how learners engaged with the designed learning processes to achieve their results, or why some learners achieved at different levels, or how some learners used the learning environment to achieve a deeper understanding. While any of these findings could be seen as the outcomes of an evaluation study, they could equally be seen as legitimate outcomes of an educational research investigation. So the question becomes: Where does educational evaluation stop and educational research begin?

This is a somewhat vexed issue as some authors have argued that providing an explanation of outcomes is not within the remit of evaluation (see Glass & Worthen, 1971). However, others, such as Oliver, Harvey, Conole and Jones (2007), have argued that evaluation and research studies can use similar methods to arrive at similar outcomes, but they can be distinguished based on how those outcomes are used:

The relationship between evaluation and research more generally remains contested. Evaluation can, in fact, contribute to research as well as providing feedback for changing teaching and learning practice. Both processes use the same methods and study the same things. However, one way to distinguish them is to consider how findings are used. If they are
Theory

interpreted by an immediate, local audience and used to support decision making, the study was probably an evaluation; if findings are interpreted in terms of theories and presented as a contribution to knowledge, it was probably research. (Oliver, Harvey et al., 2007, p. 203)

We argue here that studies of the effectiveness of e-learning involve a mixture of evaluation and research, where:

- evaluation is gathering information to help make judgements about the value and worth of an e-learning artefact or environment that can inform decision-making;
- research is gathering information to assist our understanding of how people learn using an e-learning artefact or environment.

For simplicity, we give this the overarching label of evaluation research. Our choice resonates with Patton’s (1990) definition:

When one examines and judges accomplishments and effectiveness, one is engaged in evaluation. When this examination of effectiveness is conducted systematically and empirically through careful data collection and thoughtful analysis, one is engaged in evaluation research. (p. 11)

However, like many constructs in the social sciences, the term ‘evaluation research’ has been used with various connotations and meanings. In the late 1960s and early 1970s it was used to distinguish evaluation that employs rigorous, experimental social-science research methodology from other forms of evaluation (Worthen et al., 1997). No such interpretation is present in our use of the term.

We use the term evaluation research simply to capture the idea that investigations of e-learning will often involve a mix of evaluation and research activities that can be applied throughout the e-learning life cycle. When it comes to e-learning investigations, there is typically an ebb and flow between making judgements about the e-learning environment and developing a greater understanding of learning in that environment. Furthermore, Chapter 2 indicated that varying levels of rigour can be applied to evaluation activities, while rigour is required in publishable research (for more, see Chapters 5 and 6).

Table 4.2 shows how evaluation research can occur across the e-learning life cycle by demonstrating how each of the scenarios summarized in Table 4.1 can be broken down to support both evaluative and research-based inquiries. Each of the questions in the third column of Table 4.1 is treated separately in Table 4.2, sometimes with both an evaluation and a research component. For example, scenarios B2 and B3 correspond to the second question of scenario B in Table 4.1, with evaluation and research activities respectively. Table 4.2
also explicitly indicates the elements of the LEPO framework which are most relevant at each phase of the e-learning life cycle. That is, in scenario B we are mostly concerned with the learning environment; in scenario C we are concerned with the learning environment and learning processes; and in scenario D we are interested in both learning outcomes and learning processes.

Table 4.2 graphically indicates the balance between evaluation and research for each scenario. Rather than portraying particular evaluation-research activities as purely evaluative or purely research-based, the graphics imply

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Focus</th>
<th>Activity</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Exploration</td>
<td>Evaluation Research</td>
<td>Exploring potential</td>
</tr>
<tr>
<td>B1</td>
<td>Learning environment</td>
<td>Evaluation Research</td>
<td>Judgements about the project management</td>
</tr>
<tr>
<td>B2</td>
<td>Learning environment</td>
<td>Evaluation Research</td>
<td>Judgements about the usability of the designed e-learning artefact in the context of the particular learning environment</td>
</tr>
<tr>
<td>B3</td>
<td>Learning environment</td>
<td>Evaluation Research</td>
<td>Understanding the characteristics of e-learning environments which effectively facilitate learning processes and learning outcomes</td>
</tr>
<tr>
<td>C1</td>
<td>Learning environment and learning process</td>
<td>Evaluation Research</td>
<td>Judgements, derived from actual use, about the way the learning environment was designed and how it could be improved</td>
</tr>
<tr>
<td>C2</td>
<td>Learning environment and learning process</td>
<td>Evaluation Research</td>
<td>Seeking deeper understanding about ways that learners use and interact with the e-learning environment</td>
</tr>
<tr>
<td>D1</td>
<td>Learning process and learning outcome</td>
<td>Evaluation Research</td>
<td>Judgements about whether the e-learning environment works. How effective was it in facilitating its desired outcomes?</td>
</tr>
<tr>
<td>D2</td>
<td>Learning process and learning outcome</td>
<td>Evaluation Research</td>
<td>Seeking deeper understanding about how and why the learning environment engaged particular learning processes and led to particular learning outcomes</td>
</tr>
</tbody>
</table>
that there is a variable mixture of each. In other words, we make judgements about the aspects of the learning environment that we more or less understand, and we research the aspects of the learning environment that we do not understand.

For example, in scenario A, we are studying the potential of a new technology for use in teaching and learning involving a mix of evaluation and research. We need both to understand how the technology works and to decide how we can use it, if at all. There is also a level of overlap between scenarios D1 and D2. The approaches taken to determine how effective a learning environment was and understanding how it was effective are very similar. We will unpack these distinctions further in Chapter 8.

However, many aspects of e-learning are relatively well understood. For example, online discussion forums are now an entrenched learning technology and studies over a number of years have provided insight into their effective use (Harasim, Hiltz, Teles & Turoff, 1995; Palloff & Pratt, 1999; Salmon, 2003). However, there are always additional questions that emerge when ‘tried and tested’ technologies are used in a different context, for example, with learners from a different cultural and/or linguistic background. What might be evaluation in one context may have a stronger research focus in another. So, while in many contexts online discussions are well understood, there is an emerging evaluation-research area in the use of online discussions with Asian learners (McNaught, 2011; McNaught, Cheng & Lam, 2006).

4.4 The Nature of Research

We have argued above that there is a research component to investigating the effectiveness of e-learning artefacts and environments. However, to progress our discussion we need to clarify what we mean by ‘research’. At some level, a person who cannot find their car keys and investigates where they might be is conducting research. Like ‘evaluation’ the term ‘research’ is used in a variety of ways in academic discourse but, despite this, its meaning is rarely questioned and often it is regarded as universally or even implicitly understood. At most universities an institutional bureaucracy supports research, and academic staff are rewarded for their ‘strength’ in research. But what exactly is this research? There are many contexts in which research takes place and many different ways in which research is conducted but, unless this diversity of understanding is recognized, it is difficult to have a meaningful dialogue about it. This is what we are attempting here.

Broadly speaking, ‘research’ is akin to what Shulman (1988) termed disciplined inquiry, which can be distinguished “from other sources of opinion or belief [and] is conducted and reported in such a way that the argument can be painstakingly examined” (Cronbach & Suppes, 1969, p. 15). Shulman (1988) viewed academic research as that which is disciplined, systematic, explicit and ethical, whose “data, arguments and reasoning [are] capable of withstanding
that there is a variable mixture of each. In other words, we make judgements about the aspects of the learning environment that we more or less understand, and we research the aspects of the learning environment that we do not understand.

For example, in scenario A, we are studying the potential of a new technology for use in teaching and learning involving a mix of evaluation and research. We need both to understand how the technology works and to decide how we can use it, if at all. There is also a level of overlap between scenarios D1 and D2. The approaches taken to determine how effective a learning environment was and understanding how it was effective are very similar. We will unpack these distinctions further in Chapter 8.

However, many aspects of e-learning are relatively well understood. For example, online discussion forums are now an entrenched learning technology and studies over a number of years have provided insight into their effective use (Harasim, Hiltz, Teles & Turoff, 1995; Palloff & Pratt, 1999; Salmon, 2003). However, there are always additional questions that emerge when ‘tried and tested’ technologies are used in a different context, for example, with learners from a different cultural and/or linguistic background. What might be evaluation in one context may have a stronger research focus in another. So, while in many contexts online discussions are well understood, there is an emerging evaluation–research area in the use of online discussions with Asian learners (McNaught, 2011; McNaught, Cheng & Lam, 2006).

4.4 The Nature of Research

We have argued above that there is a research component to investigating the effectiveness of e-learning artefacts and environments. However, to progress our discussion we need to clarify what we mean by ‘research’. At some level, a person who cannot find their car keys and investigates where they might be is conducting research. Like ‘evaluation’ the term ‘research’ is used in a variety of ways in academic discourse but, despite this, its meaning is rarely questioned and often it is regarded as universally or even implicitly understood. At most universities an institutional bureaucracy supports research, and academic staff are rewarded for their ‘strength’ in research. But what exactly is this research? There are many contexts in which research takes place and many different ways in which research is conducted but, unless this diversity of understanding is recognized, it is difficult to have a meaningful dialogue about it. This is what we are attempting here.

Broadly speaking, ‘research’ is akin to what Shulman (1988) termed disciplined inquiry, which can be distinguished “from other sources of opinion or belief [and] is conducted and reported in such a way that the argument can be painstakingly examined” (Cronbach & Suppes, 1969, p. 15). Shulman (1988) viewed academic research as that which is disciplined, systematic, explicit and ethical, whose “data, arguments and reasoning [are] capable of withstand...
What is Meant by Educational Evaluation and Research?  •  53

careful scrutiny by another member of the scientific community” (1988, p. 5). For the purposes of this discussion, we will use the definition of research cited by the Higher Education Funding Council for England:

‘Research’ … is to be understood as original investigation undertaken in order to gain knowledge and understanding. It includes work of direct relevance to the needs of commerce, industry, and to the public and voluntary sectors; scholarship; the invention and generation of ideas, images, performances, artefacts including design, where these lead to new or substantially improved insights; and the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction. (Research Assessment Exercise, 2008)

Research is undertaken primarily for two reasons:

1. To satisfy human curiosity – that is, to come to a greater or more profound understanding of the world in which we live.
2. To solve problems – that is, to learn to manipulate the world we live in to obtain desirable outcomes.

The multiple purposes of research are discussed further in section 4.5.

In order to achieve high standards in academic research, all researchers and research teams need to draw on a range of generic skills and capabilities including critical-thinking and problem-solving skills, integrity, honesty and reflection throughout the research process, creativity and openness to new ideas, and communication and project-management skills. However, research approaches differ across disciplines, and the following section considers this in order to inform the development of a robust approach to e-learning evaluation research.

4.4.1 Disciplinary Research Differences

The area of ‘e-learning’ is an interdisciplinary field. Both practice and evaluation research associated with e-learning can draw on a range of discipline areas such as education, sociology, psychology, information technology, computer science, information systems and design, to name just a few. While e-learning is developing as a discipline area, often individual academics have come to e-learning from established disciplines, and have tended to bring with them their particular, discipline-specific research traditions. Shulman (1988) suggested that:

What distinguishes disciplines from one another is the manner in which they formulate their questions, how they define the content of their domain and organize that content conceptually, and the principles of discovery and verification that constitute the ground rules for creating and testing knowledge in their fields. (Shulman, 1988, p. 5)
Given that it is likely that researchers, evaluators and practitioners in the area of e-learning will often reflect multiple disciplines, and that some of the research traditions of these disciplines will have different ground rules, it is worth considering the characteristics of various disciplinary research approaches, to see which can be appropriately applied to e-learning evaluation research.

The way in which disciplinary differences and research approaches can be distinguished has been the subject of much research itself. For example, Becher (1989), in his work on academic tribes and territories, identified disciplinary differences between types of academic work. Jones, Zenios and Griffiths (2004) subsequently synthesized Becher’s textual descriptions and distinguished disciplines across two dimensions: pure versus applied, and hard versus soft. Alternatively, Burkhardt (2006) discussed four research traditions: humanities, sciences, engineering and the arts, which map well to the dimensions proposed by Jones et al. (2004). The four general disciplinary research approaches are presented in Table 4.3, described in terms of their aims, products, characteristics, concerns and outcomes.

From the classifications in Table 4.3, it should be clear that some disciplines (e.g. physics) are referred to as ‘hard’ disciplines that conduct ‘pure’ research and aim to create fundamental knowledge and understanding. Research that comes from this tradition typically uses ‘hard’ data and values the reliability of numbers, and generalizability of findings. Research can be seen as ‘blue sky’ with less of an emphasis on the application or practical utility of the results. Other disciplines, such as those found within the humanities, also seek knowledge and understanding and can also engage in ‘blue sky’ research, but are typically ‘softer’ and less deterministic in their approach. An anthropologist or sociologist might seek to explore and uncover the behavioural and cultural characteristics of a particular community, with no concern for causality, and attempting to discount any preconceptions about that behaviour. Hard–applied disciplines, such as engineering, focus on the practical outcomes of the research process. They are often directed at developing an understanding in order to enhance the way in which something in the world operates. The products or outcomes of this type of research are typically tools and processes that are fit for purpose. For example, an engineer or computer scientist might create and implement a solution to a problem and test its reliability and efficiency. Here, while there is still some sense of discovery, there is an overriding interest in knowing whether the innovation works and why.

Finally, there is the soft–applied quadrant in which educational research is often positioned. Research in this category is often initiated to resolve practical problems or improve professional practice. While researchers such as Burkhardt (2006) claim that, as in the humanities, educational research tends to focus more on critical commentary than on hard evidence, the practical reality is more pluralistic than this. A review of ‘educational’ journals would reveal that some educational research focuses on questions and employs techniques that
Table 4.3 Disciplinary differences in academic research (derived from Jones et al. (2004) and Burkhardt (2006))

<table>
<thead>
<tr>
<th>Pure</th>
<th>Applied</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hard</strong></td>
<td><strong>Applied sciences, e.g. engineering</strong></td>
</tr>
<tr>
<td><strong>Aim</strong></td>
<td>• knowledge and fundamental understanding</td>
</tr>
<tr>
<td></td>
<td>• practical impact</td>
</tr>
<tr>
<td></td>
<td>• helping the world work better</td>
</tr>
<tr>
<td><strong>Product</strong></td>
<td>• assertions supported by evidence-based arguments</td>
</tr>
<tr>
<td></td>
<td>• tools and/or processes which are fit for purpose</td>
</tr>
<tr>
<td><strong>Characteristics</strong></td>
<td>• cumulative</td>
</tr>
<tr>
<td></td>
<td>• atomistic (crystalline/tree-like)</td>
</tr>
<tr>
<td><strong>Concerned with</strong></td>
<td>• universals</td>
</tr>
<tr>
<td></td>
<td>• quantities</td>
</tr>
<tr>
<td></td>
<td>• simplification</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>• discovery</td>
</tr>
<tr>
<td></td>
<td>• explanation</td>
</tr>
</tbody>
</table>

| **Soft** | **Humanities and pure social sciences, e.g. anthropology, history** |
| | **Applied social sciences, e.g. education and the arts** |
| **Aim** | • knowledge and understanding |
| | • practical impact |
| | • helping the world work better |
| **Product** | • critical commentary |
| | • assertions not backed up by (empirical) evidence |
| | • protocols |
| | • procedures |
| | • practical guides |
| **Characteristics** | • reiterative |
| | • holistic (organic/river-like) |
| **Concerned with** | • particulars |
| | • qualities |
| | • complication |
| **Outcomes** | • understanding |
| | • interpretation |

are closely aligned with the applied sciences (i.e. hard-applied) while other forms of educational research are more sociological and are clearly aligned with a pure social-science tradition (i.e. pure-soft). There are no easy categorizations here: educational research is diverse and contested; the discipline of education contains a number of viewpoints about what constitutes valid research. Just as education is difficult to assign to a single disciplinary focus, investigations into e-learning – effectively a subset of educational research – are
equally difficult to categorize. We have argued above that investigations of the effectiveness of e-learning can have both an evaluative and a research focus. Because e-learning evaluation research is concerned with practical applications in education as well as with enhancement of professional practice, it can clearly be classified as soft–applied research. However, many e-learning investigations contain a ‘hard’ element with an ‘applied’ research focus. Alternatively, some social-science research in e-learning seeks fundamental understanding in complex learning communities, institutions or cultures. So, while disciplinary differences are a useful reminder that research – what it is, how it is done – is often aligned with disciplinary traditions, it may be useful to consider other non-disciplinary ways of categorizing research. This is the focus of the next section, which looks at two non-disciplinary ways of categorizing research. Together with what we have just discussed, they provide us with several lenses through which we can view e-learning evaluation research.

4.5 Research Classifications

4.5.1 Pasteur’s Quadrant

In the second half of the 20th century, the accepted way of classifying research was to distinguish between pure (or basic) research and applied research. As indicated in Table 4.3, the primary aim of pure research was to seek fundamental understandings, while applied research was more focused on solving practical problems. Stokes (1997) criticized the pre-eminence of pure research, and argued that the pure–applied distinction was too narrow and did not consider how different types of research might be used. Based on his critique, he proposed a two-dimensional model for classifying research that he called Pasteur’s Quadrant (see Table 4.4).

One dimension of Pasteur’s Quadrant classified research in terms of the degree to which it reflected a quest for fundamental understanding. Stokes introduced a second dimension to the research classification, which was ‘consideration of use’. By using this simple two-by-two classification and providing historical examples from the work of eminent researchers, Stokes (1997) challenged two

Table 4.4 Pasteur’s Quadrant

<table>
<thead>
<tr>
<th>Quest for fundamental understanding</th>
<th>Consideration of use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Pure basic research (Bohr)</td>
<td>1.</td>
</tr>
<tr>
<td>Use-inspired basic research (Pasteur)</td>
<td>2.</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Pure applied research (Edison)</td>
<td>3.</td>
</tr>
<tr>
<td></td>
<td>4.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© Phillips, Rob; McNaught, Carmel; Kennedy, Gregor, Mar 22, 2012, Evaluating e-Learning : Guiding Research and Practice
equally difficult to categorize. We have argued above that investigations of the effectiveness of e-learning can have both an evaluative and a research focus. Because e-learning evaluation research is concerned with practical applications in education as well as with enhancement of professional practice, it can clearly be classified as soft–applied research. However, many e-learning investigations contain a ‘hard’ element with an ‘applied’ research focus. Alternatively, some social-science research in e-learning seeks fundamental understanding in complex learning communities, institutions or cultures. So, while disciplinary differences are a useful reminder that research – what it is, how it is done – is often aligned with disciplinary traditions, it may be useful to consider other non-disciplinary ways of categorizing research. This is the focus of the next section, which looks at two non-disciplinary ways of categorizing research. Together with what we have just discussed, they provide us with several lenses through which we can view e-learning evaluation research.

4.5 Research Classifications

4.5.1 Pasteur’s Quadrant

In the second half of the 20th century, the accepted way of classifying research was to distinguish between pure (or basic) research and applied research. As indicated in Table 4.3, the primary aim of pure research was to seek fundamental understandings, while applied research was more focused on solving practical problems. Stokes (1997) criticized the pre-eminence of pure research, and argued that the pure–applied distinction was too narrow and did not consider how different types of research might be used. Based on his critique, he proposed a two-dimensional model for classifying research that he called Pasteur’s Quadrant (see Table 4.4).

One dimension of Pasteur’s Quadrant classified research in terms of the degree to which it reflected a quest for fundamental understanding. Stokes introduced a second dimension to the research classification, which was ‘consideration of use’. By using this simple two-by-two classification and providing historical examples from the work of eminent researchers, Stokes (1997) challenged two
assumptions of the traditional pure–applied dichotomy: (i) that pure research sought fundamental understanding while applied research did not; and (ii) that basic research always preceded applied research.

Stokes provided several examples from the career of microbiologist Louis Pasteur to debunk these assumptions and demonstrate that his work was inspired by both consideration of use and a quest for fundamental understanding (Quadrant 2 in Table 4.4). Similarly, Stokes argued that Neils Bohr’s work on atomic structure had no consideration of use, but sought fundamental understanding (Quadrant 1 in Table 4.4). On the other hand, Thomas Edison’s inventions applied existing understanding to develop new tools and techniques to solve real-world problems, without any intention of extending that understanding (Quadrant 4).

One might question whether research in Quadrant 3, which has no consideration for use and no quest for fundamental understanding, has any value, or, indeed, is academic research at all. Stokes (1997) characterized this Quadrant as “research that systematically explores particular phenomena without having in view either general explanatory objectives or any applied use to which the results will be put” (italics in original; p. 74). The dedicated activities of amateur astronomers and birdwatchers fit into this category, and this type of ‘research’ may ultimately prove to be of considerable value to other researchers, in other Quadrants, at a later stage.

Reeves and Hedberg (Reeves, 1999; Reeves & Hedberg, 2003) have been vocal critics of much e-learning research, claiming that it is located in this third Quadrant and, as such, has neither quest for understanding nor consideration of use. They have argued for e-learning research to be more use-inspired and socially responsible. In section 4.3.1, we argued that evaluation research into e-learning involves a mixture of judgements (evaluation) and understanding (research). We consider both of these activities to be inspired by use (in Stokes’ terms) because we are studying a learning environment as it is used by learners and, through the process of evaluation research, we are aiming to understand better how technology can be applied and used. Furthermore, the research component of evaluation research is clearly intent on a quest for fundamental understanding, while the evaluation component is not.

4.5.2 Boyer’s Four Scholarships

One final lens through which to look at academic research is Ernest Boyer’s (1990) four different types of scholarship. This analysis was also based on a critique of the post-war emphasis on pure research, which Boyer (1990) called the scholarship of discovery. In addition to the scholarship of discovery, Boyer proposed three other types of scholarship: the scholarship of integration, the scholarship of application and the scholarship of teaching.

The scholarship of discovery is perhaps the most immediately recognizable of Boyer’s classifications, and is aligned with traditional notions of research –
that is, research that pursues new knowledge and fundamental understanding. The scholarship of discovery, therefore, is clearly high on Stokes’ (1997) dimension of ‘quest for fundamental understanding’, and, as it is silent about ‘consideration of use’, fits squarely in Quadrant 1 (Table 4.4). The scholarship of integration involves connecting knowledge and discovery into larger patterns and contexts. This includes interdisciplinary and multidisciplinary research, at the “boundaries where fields converge” (p. 19). As Boyer (1990) explained:

The distinction we are drawing between ‘discovery’ and ‘integration’ can be best understood, perhaps, by the questions posed. Those engaged in discovery ask ‘What is to be known, what is yet to be found?’ Those engaged in integration ask ‘What do the findings mean? Is it possible to interpret what’s been discovered in ways that provide a larger, more comprehensive understanding?’ (p. 19)

The scholarship of application involves engagement in problems that affect individuals, institutions and society, and asks questions such as: ‘How can knowledge be responsibly applied to consequential problems? How can it be relevant to society?’ This scholarship type is directly analogous to Stokes’ (1997) ‘consideration of use’ dimension. Like Stokes’ consideration of use, Boyer’s scholarship of application acknowledges that real-world problems can define an agenda for research. Boyer argues that the scholarship of application moves “from theory to practice, and from practice back to theory, which in fact makes theory … more authentic” (Boyer, 1996, p. 17).

The scholarship of teaching is perhaps the hardest to conceptualize. It is not scholarship about teaching, but the scholarship of teaching. Hutchings and Shulman (1999) distinguished between good teaching and the scholarship of teaching in that the latter: gathers evidence; is informed by current ideas about the field, and teaching in that field; and invites “peer collaboration and review” (p. 13). Furthermore, the scholarship of teaching is public, “open to critique and evaluation and in a form that others can build on” (p. 13); and it involves inquiry into “issues of student learning” (p. 13). The scholarship of teaching goes beyond improving individual classroom practice to advancing teaching practice in general, the focus of various forms of action inquiry (Kemmis & McTaggart, 2000; McNiff, Lomax & Whitehead, 2003).

Boyer’s four scholarships have some overlaps with Stokes’ schema. The scholarship of discovery and the scholarship of application are directly analogous to Stokes’ ‘fundamental understanding’ and ‘consideration of use’ components, respectively, which are conceived of as separate dimensions by Stokes (1997). On the other hand, the scholarship of integration cuts across Stokes’ boundaries, and the scholarship of teaching builds on the other scholarships. As Hutchings and Shulman (1999) explained, scholarship of teaching “is a special case of the scholarship of application and engagement,
and frequently entails the discovery of new findings and principles” (p. 15). Boyer’s (1990) scholarship of teaching is an eminently suitable lens through which to view evaluation research into e-learning.

4.6 Examples of Evaluation Research into e-Learning

Collectively, we have managed and contributed to many teams developing and researching innovative e-learning applications over many years. This section describes some of these projects and maps them against the general research approaches outlined in previous sections, to demonstrate the diversity of approaches to e-learning evaluation research. These examples cover a significant time period, partly because issues in e-learning evaluation research are enduring, and particularly to indicate the diversity of valid approaches.

Example 1: Virtual reality techniques for veterinary diagnostic imaging

A particular teaching and learning problem led to the development of the e-learning solution which was the focus of this research. A course in veterinary diagnostic imaging was offered at Masters level by distance education. In a face-to-face situation, diagnostic imaging skills are learned in small groups, through gesturing at radiographs around a ‘light box’, but the traditional, print-based distance-education model was not well suited to the desired learning outcome. First, the packages of hard-copy radiographs sent to learners were bulky and expensive to post. Second, learners were often unable to identify the relevant regions of X-rays, and were unable to receive timely feedback on diagnoses they made.

These issues were addressed by several technological components, progressively developed in three cycles over several years.

1. The QuickTime VR three-dimensional virtual-reality technique was applied to large, two-dimensional X-ray images. This enabled learners to view the whole X-ray on screen, and then zoom in on fine detail within the images with minimal loss of diagnostic clarity, as well as seeing annotations embedded in the images (Phillips, Pospisil & Richardson, 2001).

2. Microsoft Word macros were developed to enable academics to design their own diagnostic self-tests. Learners can enter their answer, click a button and receive a detailed expert answer immediately (Phillips, 2002a).

3. A prototype digital whiteboard was developed to enable text-based discussion and annotation of a shared image (Phillips, Scott & Richardson, 2003).

An evaluation was carried out during each cycle, and the focus was largely on the e-learning artefact, and how it could be improved. However, the research also addressed aspects of the learning environment (accessibility issues and
and frequently entails the discovery of new findings and principles” (p. 15). Boyer’s (1990) scholarship of teaching is an eminently suitable lens through which to view evaluation research into e-learning.

4.6 Examples of Evaluation Research into e-Learning
Collectively, we have managed and contributed to many teams developing and researching innovative e-learning applications over many years. This section describes some of these projects and maps them against the general research approaches outlined in previous sections, to demonstrate the diversity of approaches to e-learning evaluation research. These examples cover a significant time period, partly because issues in e-learning evaluation research are enduring, and particularly to indicate the diversity of valid approaches.

Example 1: Virtual reality techniques for veterinary diagnostic imaging
A particular teaching and learning problem led to the development of the e-learning solution which was the focus of this research. A course in veterinary diagnostic imaging was offered at Masters level by distance education. In a face-to-face situation, diagnostic imaging skills are learned in small groups, through gesturing at radiographs around a ‘light box’, but the traditional, print-based distance-education model was not well suited to the desired learning outcome. First, the packages of hard-copy radiographs sent to learners were bulky and expensive to post. Second, learners were often unable to identify the relevant regions of X-rays, and were unable to receive timely feedback on diagnoses they made.

These issues were addressed by several technological components, progressively developed in three cycles over several years.

1. The QuickTime VR three-dimensional virtual-reality technique was applied to large, two-dimensional X-ray images. This enabled learners to view the whole X-ray on screen, and then zoom in on fine detail within the images with minimal loss of diagnostic clarity, as well as seeing annotations embedded in the images (Phillips, Pospisil & Richardson, 2001).

2. Microsoft Word macros were developed to enable academics to design their own diagnostic self-tests. Learners can enter their answer, click a button and receive a detailed expert answer immediately (Phillips, 2002a).

3. A prototype digital whiteboard was developed to enable text-based discussion and annotation of a shared image (Phillips, Scott & Richardson, 2003).

An evaluation was carried out during each cycle, and the focus was largely on the e-learning artefact, and how it could be improved. However, the research also addressed aspects of the learning environment (accessibility issues and
the broader course context), and whether the innovations were addressing the learning process as intended.

A number of sources of data were used to evaluate the usability and effectiveness of the e-learning environment. These included design specifications, expert feedback, learner observation, lecturer feedback, online usage statistics and an email survey of learner experiences.

The rounds of evaluation identified bugs and interface issues which were resolved in subsequent years. Learners found the virtual X-rays were very useful. In the initial round, the quality wasn’t quite as high as hard copies, and learners preferred to practise on the virtual images and be assessed on the hard copies. A higher-resolution scanner was used in subsequent years. However, the flexibility of the new approach was a clear benefit. Practising professionals could study for a higher qualification in their own time at their location of choice, without being tied to the piles of hard copies at home.

In terms of the research classifications described in previous sections, this work was a mixture of hard- and soft-applied research, with a strong concern for use, but little quest for fundamental understanding. In Boyer’s terms this work involved the scholarship of teaching, with an emphasis on the scholarship of integration, due to the linking of existing ideas in new ways.

Example 2: A longitudinal study of the use of e-cases in veterinary microbiology

The development of e-cases in veterinary microbiology began in the early 1990s at an Australian university. The e-cases were designed for undergraduate learners and integrated into a revised learning design with an expanded set of desired learning outcomes (information management, practical diagnosis of bacterial and fungal animal diseases, team work and communication skills) and more appropriate assessment strategies. Data was collected over a period of seven years and some data collection is ongoing. A brief summary of the data and interpretations is in Table 4.5. The underlying technology is now web-based but a similar learning design with the same e-cases is still in use 20 years later. The overall outcome of the project was that changes in institutional climate worked together with the project’s outcomes to influence faculty policy for the design of the veterinary program to include a greater number of problem-based courses with learning designs similar to the veterinary microbiology course. This project highlights the need to think at an institutional or system level in order to achieve lasting benefits (Laurillard, 1999; see also Chapter 11). Further detail can be found in McNaught, Whithear and Browning (1994; 1999) and McNaught, Burd, Whithear, Prescott and Browning (2003).

This was soft-applied research or, in Stokes’ terms, a mixture of use-inspired basic research and pure applied research, through a clear quest for understanding together with a pragmatic focus on operational matters. In Boyer’s classification, this work involved the scholarship of teaching, with some emphasis on the scholarship of application (innovative e-learning artefact).
What is Meant by Educational Evaluation and Research?  

Example 3: Collaborative writing exercises using a wiki with large classes

This e-learning project was part of a larger study that explored staff and learners’ experiences with technology in higher education, and the use of Web 2.0 technologies in teaching and learning (Kennedy, Dalgarno, Bennett et al., 2009). The project involved the design and implementation of a technology-based collaborative writing exercise using a wiki with a large class of first-year psychology learners. Laboratory classes containing between 20 and 30 learners were asked to create collaboratively a wiki-based submission on a key area being covered in the psychology course (motion detection by the visual system). Learners and teachers were provided with extensive technical, administrative and pedagogical support, and learners were given guidance on the basic concepts they should cover in their submission. Learners were encouraged to create scholarly summaries of the key concepts of motion detection using descriptions, reflections, quotes, images, web links and diagrams.

This e-learning environment was one of eight e-learning implementation projects that were evaluated through a cross-case comparison, broadly based on Reeves and Hedberg’s (2003) notion of effectiveness evaluation. Quantitative and qualitative data were collected from staff and learners in the areas of knowledge, skills and attitudes. Staff members involved in the learning activity were interviewed on a number of occasions across the semester, and a sample of learners completed an online questionnaire and contributed to focus-group discussions at the conclusion of the exercise. The outcomes of this evaluation

Table 4.5 Outline of the longitudinal study of the use of e-cases in veterinary microbiology

<table>
<thead>
<tr>
<th>Data</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners’ impressions of the learning environment and course design – annual survey (ongoing) and focus groups over three years</td>
<td>Initial reservations but views changed quite quickly to being positive as learners adapted to what was their first experience of e-learning</td>
</tr>
<tr>
<td>Learners’ perceptions of their learning outcomes – annual survey (ongoing)</td>
<td>Generally positive but feedback has allowed adjustments, e.g. to the length of the e-cases</td>
</tr>
<tr>
<td>Assessment data on learning outcomes – analysis of the cognitive demand (Bloom, 1956) of examination items pre- and post-innovation and learners’ overall performance (two years)</td>
<td>Evidence that learners do well on assessment which was more demanding than previous assessment designs</td>
</tr>
<tr>
<td>Interviews with cohort 2 of the learners two years later when they were in clinical experience</td>
<td>Learning benefits did not last because of the isolated nature of the experience</td>
</tr>
</tbody>
</table>

Example 3: Collaborative writing exercises using a wiki with large classes

This e-learning project was part of a larger study that explored staff and learners’ experiences with technology in higher education, and the use of Web 2.0 technologies in teaching and learning (Kennedy, Dalgarno, Bennett et al., 2009). The project involved the design and implementation of a technology-based collaborative writing exercise using a wiki with a large class of first-year psychology learners. Laboratory classes containing between 20 and 30 learners were asked to create collaboratively a wiki-based submission on a key area being covered in the psychology course (motion detection by the visual system). Learners and teachers were provided with extensive technical, administrative and pedagogical support, and learners were given guidance on the basic concepts they should cover in their submission. Learners were encouraged to create scholarly summaries of the key concepts of motion detection using descriptions, reflections, quotes, images, web links and diagrams.

This e-learning environment was one of eight e-learning implementation projects that were evaluated through a cross-case comparison, broadly based on Reeves and Hedberg’s (2003) notion of effectiveness evaluation. Quantitative and qualitative data were collected from staff and learners in the areas of knowledge, skills and attitudes. Staff members involved in the learning activity were interviewed on a number of occasions across the semester, and a sample of learners completed an online questionnaire and contributed to focus-group discussions at the conclusion of the exercise. The outcomes of this evaluation
were that the e-learning environment was effective (e.g. high task relevancy, encouraged both staff and learners to share and reflect); and identified where learners encountered difficulties (e.g. many learners found it hard to negotiate and manage collaboration and contribution) (cf. Kennedy et al., 2009).

In addition to the use of interviews and questionnaires as tools in this evaluation, metrics that were automatically generated by the wiki were also harvested for a study that sought to investigate the degree to which learners behaved cooperatively and collaboratively in developing their contributions. Electronic measures of the number, timing and scope of contributions made by the learners in the cohort and the content of their wiki comments were the subject of detailed analysis. This analysis showed that while a great number of learners were making a contribution to the activity, evidence of sustained online collaboration was limited (Judd, Kennedy & Cropper, 2010).

Like the previous example, this work can be characterized as soft–applied research in disciplinary terms, but as use-inspired basic research according to Stokes, because the aim was to arrive at evaluation–research outcomes that could inform practice. In Boyer’s terms, this work was mainly the scholarship of application with an element of the scholarship of teaching.

4.7 Summary

This chapter discussed the nature of evaluation in the context of this book, noting the characteristics of educational evaluation and the similarities and distinctions between program and project evaluation. We presented various evaluation approaches and models which we applied to four e-learning evaluation scenarios related to the e-learning life cycle.

We established that studies of the effectiveness of e-learning involve a variable mixture of both evaluation and research, and we have called this ‘evaluation research’. We then turned our attention to the interpretation of ‘research’, and discussed how the term is interpreted differently in different disciplines. We presented three models which characterize research in different dimensions: the disciplinary distinctions of Jones et al. (2004) between hard and soft, and pure and applied research; Stokes’ (1997) consideration of use, and quest for fundamental understanding; and Boyer’s (1990) scholarships of discovery, integration, application and teaching. These acted as lenses through which we can view e-learning evaluation research.

In e-learning evaluation research, the evaluation component involves making judgements about the usability and usefulness of an e-learning environment, while the research component involves a search for fundamental understanding. These two components resonate well with Stokes’ (1997) consideration of use, and quest for fundamental understanding. We have also established that e-learning evaluation research can be multidisciplinary, combining a mixture of soft and hard, and applied and pure characteristics. Finally, we established that evaluation research into e-learning is a form of Boyer’s (1990) scholarship...
were that the e-learning environment was effective (e.g. high task relevancy, encouraged both staff and learners to share and reflect); and identified where learners encountered difficulties (e.g. many learners found it hard to negotiate and manage collaboration and contribution) (cf. Kennedy et al., 2009).

In addition to the use of interviews and questionnaires as tools in this evaluation, metrics that were automatically generated by the wiki were also harvested for a study that sought to investigate the degree to which learners behaved cooperatively and collaboratively in developing their contributions. Electronic measures of the number, timing and scope of contributions made by the learners in the cohort and the content of their wiki comments were the subject of detailed analysis. This analysis showed that while a great number of learners were making a contribution to the activity, evidence of sustained online collaboration was limited (Judd, Kennedy & Cropper, 2010).

Like the previous example, this work can be characterized as soft–applied research in disciplinary terms, but as use-inspired basic research according to Stokes, because the aim was to arrive at evaluation–research outcomes that could inform practice. In Boyer’s terms, this work was mainly the scholarship of application with an element of the scholarship of teaching.

4.7 Summary

This chapter discussed the nature of evaluation in the context of this book, noting the characteristics of educational evaluation and the similarities and distinctions between program and project evaluation. We presented various evaluation approaches and models which we applied to four e-learning evaluation scenarios related to the e-learning life cycle.

We established that studies of the effectiveness of e-learning involve a variable mixture of both evaluation and research, and we have called this ‘evaluation research’. We then turned our attention to the interpretation of ‘research’, and discussed how the term is interpreted differently in different disciplines. We presented three models which characterize research in different dimensions: the disciplinary distinctions of Jones et al. (2004) between hard and soft, and pure and applied research; Stokes’ (1997) consideration of use, and quest for fundamental understanding; and Boyer’s (1990) scholarships of discovery, integration, application and teaching. These acted as lenses through which we can view e-learning evaluation research.

In e-learning evaluation research, the evaluation component involves making judgements about the usability and usefulness of an e-learning environment, while the research component involves a search for fundamental understanding. These two components resonate well with Stokes’ (1997) consideration of use, and quest for fundamental understanding. We have also established that e-learning evaluation research can be multidisciplinary, combining a mixture of soft and hard, and applied and pure characteristics. Finally, we established that evaluation research into e-learning is a form of Boyer’s (1990) scholarship.
of teaching, which, in turn, encompasses elements of the scholarship of application, the scholarship of integration and the scholarship of discovery.

In the remainder of this book we will use the terms ‘practitioners’, ‘evaluators’, ‘researchers’ and ‘investigators’ interchangeably, using the term which seems most appropriate to the context where it is used.

We have argued in this chapter that there is a need to look beyond disciplinary assumptions when conducting e-learning evaluation research. At the same time, we do not claim that there is one right approach to evaluation research into e-learning. The examples in the previous section demonstrate that a diversity of approaches is feasible. This makes it important to clarify the context of any evaluation research and justify the approach taken. Different approaches may also be appropriate at different phases of the e-learning life cycle and for different contexts, but those approaches need to be justifiable, rather than applied in an unthinking manner. Chapter 5 explores these issues in a more theoretical way.
5 Research Paradigms and Methodologies

5.1 Introduction

Chapter 4 established that there are many types of research, and there are disciplinary differences between approaches to research. It further established that e-learning evaluation research involves a mixture of making judgements (evaluation) and seeking understanding (research), and that this resonates well with Stokes’ (1997) two dimensions of ‘consideration of use’ and ‘quest for fundamental understanding’, as well as with Boyer’s (1990) notion of the scholarship of teaching.

Chapter 4 also presented broad arguments that there are different approaches to e-learning evaluation research, depending on the stage of the e-learning life cycle that is being investigated. This chapter seeks to provide a more theoretical view of the philosophies and concepts which underpin paradigms of research and the research methodologies which may be associated with them. This is done in order to help readers to recognize their own preferred approaches to research or investigation, and to guide them towards the most appropriate approach for their situation. This chapter addresses issues relevant to educational research in general (with a ‘leaning’ towards e-learning). Chapter 6 will build on this chapter to explicitly discuss research approaches which are appropriate for e-learning.

We recognize, as Reeves (2000b) did, that “the research goals held by any given [educational] researcher are influenced by many factors including the epistemological views of the investigator, his/her research training, and the dominant research paradigms within his/her line of inquiry” (p. 22). One aim of this chapter is to encourage readers, whether novice or experienced researchers, to take a broader, reflective view of their scholarly practice, so that they can contribute to the improvement of research into e-learning in all its forms. We commence by discussing the definition of the research problem (the phenomenon, research goals and overarching research questions), and then various epistemological and paradigmatic aspects of research, before discussing methodologies.