6
Cognitive Approaches to the Evaluation of Healthcare Information Systems

Andre W. Kushniruk and Vimla L. Patel

Introduction

This chapter provides an overview of cognitive approaches to the evaluation of healthcare information systems. Cognitive approaches in health informatics focus on understanding the processes involved in the decision making and reasoning of healthcare workers as they interact with information systems to carry out a range of tasks. In the first part of the chapter the motivation and theoretical background to cognitive evaluation are provided. The importance of developing effective methods for understanding how systems impact on cognitive processes is discussed as well as the need for developing new approaches to system evaluation borrowing from advances in cognitive science and the study of human–computer interaction. In particular, methods emerging from the areas of usability engineering and cognitive task analysis have important implications for the improved assessment of cognition involved in complex medical tasks and the impact of information systems. Methodologies are described for considering evaluation throughout the system design and development life cycle. The chapter then illustrates how research in cognitive science can be used to drive the development of new conceptual frameworks for evaluation of healthcare information systems. Specific examples from our research will be provided, ranging from application of cognitive approaches for the laboratory analysis of user interactions with complex information systems such as electronic medical records, to the cognitive evaluation of Web-based information resources.

A wide variety of approaches have been taken in the evaluation of healthcare information systems. Many of these evaluations have focused on assessing outcomes associated with deployment and use of systems in clinical environments. These studies have typically involved measurement of dependent variables such as cost of health care, quality of care, and other outcomes [1]. Although such summative evaluation of completed healthcare information systems is necessary to ensure their effectiveness, in recent years an increasing emphasis has appeared on the in-depth study of the effects of such systems on the complex reasoning, decision making, and cognitive processes involved in health care [2–4]. Closely related to this trend is the assessment and evaluation of the impact of emerging healthcare systems on tasks and workflow in health care. The objective of many of these evaluations has been not only assessing the healthcare outcomes of completed systems, but also, as important, assessing the effects of information systems on the *process* of healthcare delivery. From a *practical* perspective, the objective of such process-oriented evaluations of systems under development is to provide iterative input into the improved design and programming of the systems *before* they are deployed. Closely related to some of the evaluation methods used for providing input to designers of healthcare systems are evaluations targeted even earlier in the systems development life cycle that are aimed at assessing the *information needs* of healthcare workers as a basis for design and development of health information systems. Indeed, as argued by Cysneiros and Kushniruk, improved methods for assessing and reasoning about system requirements in design of health information systems may be the key to delivery of improved healthcare information systems [5]. As a consequence, in this chapter we consider evaluation of healthcare information systems from a cognitive, process-centered perspective, along the entire systems development life cycle, from initial requirements gathering and assessment of user information needs, to the evaluation of completed software components and products.

Assessing Unintended Effects of Information Technology

The introduction of information technologies in health care can profoundly affect the way healthcare workers carry out tasks and provide health care. In addition, it has been shown that the introduction of health information systems can have significant *unintended* or *unexpected* effects not just on workflow but also on the decision making and reasoning of healthcare workers [3]. Evaluation approaches that employ an outcomes-based perspective, where variables of interest are identified prior to subjects interacting with systems (e.g., cost of health care, mortality rates, etc.) and then measured after interaction (e.g., a group of healthcare workers interacting with an information system), are unable to assess *unexpected* effects of an information technology that the evaluators have not expected to find. Thus, although traditional approaches to evaluating information systems involving clinical controlled trials and summative evaluation of systems are needed to ensure that systems meet *expectations* of designers, the assessment of effects of systems that are *emergent* (in that they are unexpected) requires a different kind of approach to evaluation focused around assessing the *process* of use of a system in order to discover what the effects of the system are. For example, in a series of studies we conducted of use of a computerized patient record (CPR) system, we found that the particular system under study (which promoted a high level of organization of medical data) resulted in subjects (i.e., physicians) changing the way they normally requested and processed patient information during the doctor–patient interview. Specifically, we found the physicians were strongly guided by the ordering and sequencing of information in the CPR when interviewing patients using the systems, rather than following their own "knowledge base." After experience in using the system we found that the order and organization of information within the CPR greatly affected the physicians' questioning, with experienced users of the system following what we termed "screen-driven" behavior (i.e., asking questions of patients based on the order of information presented on the computer screen) [3]. Furthermore, such unexpected effects of information technology often constitute the type of information that designers of systems find most useful for modifying and improving system design during the process of system development, described in the next section.

The Systems Development Life Cycle

In the software industry a wide range of methodologies have been developed for guiding the design and deployment of information systems [6]. The phases involved in creation and maintenance of information systems is known as the *systems development life cycle* (*SDLC*). The "traditional SDLC" (see Figure 6.1) that emerged in development of early computer applications several decades ago involves the progression through fixed "phases" (a phase consisting of a set of related activities), beginning with

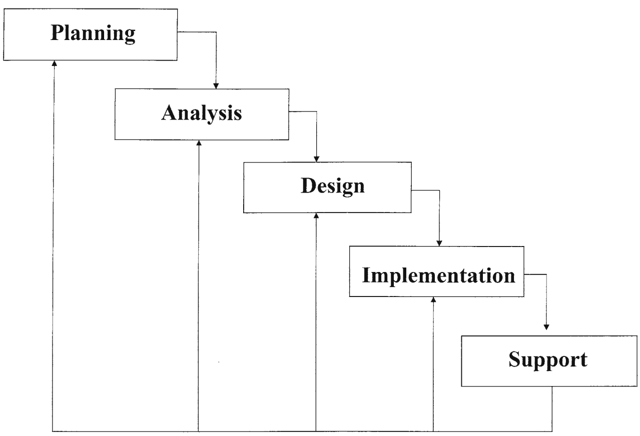

FIGURE 6.1. The classic waterfall system development life cycle.

project planning in Phase 1 and moving to *analysis and requirements gathering* in Phase 2. Once requirements for a project (both technical and user requirements) are obtained, *design* of the system is embarked on in Phase 3. Once design is finalized, in Phase 4 *implementation* (i.e., programming) of the system is undertaken. Finally, in Phase 5, the system is in place and must undergo *support and maintenance* (until it is eventually phased out, replaced, or modified by a new system, leading to a new cycle of development).

Although such an approach to system development has proven to be suitable for many software applications, ranging from applications in industries such as banking to aerospace, it has proven to be a limiting factor in the successful design and deployment of systems in many complex and highly user-centered application domains, in particular health care [2]. The emphasis of the traditional life cycle on fixed and ordered sequence of phases has had a number of drawbacks, including the following: (1) lack of flexibility in moving "back" to previous stages—in particular, if improved knowledge of user requirements would require a costly rethinking of design or implementation decisions once those phases have been passed through (i.e., it is difficult to go back to previous stages), (2) the assumption that user requirements can be adequately defined in the early analysis or requirements gathering phase, and (3) emphasis on waiting until the system is nearly complete (i.e., often during what is known as "beta testing") before conducting intensive end-user testing with a system to be deployed (again making potentially needed rethinking and redesign of major software components difficult and costly). Although such problems are typical in complex domains such as health care when attempting to apply a "traditional" approach to design and development, it should be noted that this traditional approach to software development is commonplace in the healthcare software industry today. In the context of this chapter, of particular interest is the issue of evaluation and testing of systems during the SDLC. Along these lines we will discuss the potential role of cognitive methods in improving the evaluation of systems along the various phases of the SDLC and as an important adjunct to newer approaches to systems development.

In contrast to traditional approaches to software development described earlier, in recent years, a number of software engineering methodologies have been developed that focus on deploying evaluation methods throughout the software life cycle—from the initial analysis of user needs, through the entire design process, as well as the implementation activities. Such approaches to system development are closely related to the concept of *user-centered design*, which emphasizes continued refinement and iteration in the systems development life cycle with a continual focus on evaluation with potential end users of systems at every stage of design and development [7]. As an example, the method known as *rapid prototyping* and other related approaches involve continual and iterative cycles of design and testing of software products and components prior to releasing a system.

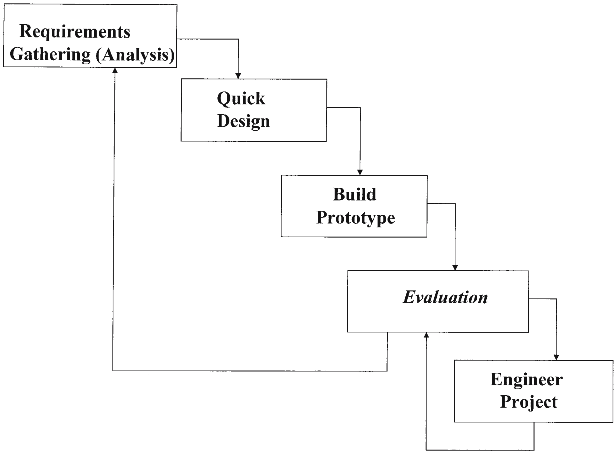

FIGURE 6.2. Rapid prototyping and the role of evaluation.

As can be seen in Figure 6.2, this approach clearly situates *evaluation* of systems as a key aspect of systems development from which decisions to modify or extend design are based. In particular, methods emerging from the fields of usability engineering and cognitive science are of particular value in providing more precise and useful assessments of information systems, particularly from the perspective of potential end users, in providing feedback to designers of systems in cycles of rapid development. This philosophy of system design is consistent with approaches to formative evaluation that have emerged in health informatics and that will form the focus of the discussion of cognitive approaches in this chapter. However, it should be noted that the methods to be described below are also of considerable value in assessing the effects of *completed* information systems on healthcare workers' decision making and reasoning in conducting summative evaluation in health informatics.

The Cognitive Continuum in Health Care, the Role of Expertise, and Cognitive Task Analysis

Prior to discussing specific cognitive methods that can be applied throughout the SDLC (as described above), we will place our work in the context of three important conceptual frameworks emerging from the study of cognitive science: (1) the *cognitive continuum* in reasoning and decision making, (2) the *expertise* continuum, and (3) a methodological framework for analyzing human–computer interactions collectively, known as *cognitive task analysis* (*CTA*). According to Hammond, cognitive processes in decision making and reasoning can be located along a *cognitive continuum*, which ranges between *intuition* and *analysis* [8]. Intuitive processing (which is characterized by recognition and quick response) is induced when experienced decision makers are faced with large amounts of information or very short time frames for responding to situations. In contrast, *analysis* is induced by tasks that involve sufficient processing time and presentation of quantitative information. Analytical processes are also associated with information processing by individuals who may lack expertise in a problem solving area, and therefore lack the ability to respond or make decisions based on their prior experiences with similar situations and recognition of similar contexts. In reality, decision making and reasoning may move along the continuum within the same problem solving context. For example, decision makers may apply intuitive (recognitional) processes in solving a part of a problem that they are familiar with or which is routine, and then shift to analytical processing when faced with problem complexity or lack of familiarity with that part of the problem. As demonstrated by Hamm, the concept of the cognitive continuum is of value in helping to explain how decisions are made in complex domains such as health care, which are characterized by complexity of information, shifting constraints, time pressure, and uncertainty of information [9]. As will be seen, knowledge of how healthcare workers move along the cognitive continuum is of considerable relevance in analyzing complex cognitive activity, providing insight into understanding cognitive processes involved in complex tasks, and more specifically, providing guidance in development of frameworks for coding and analyzing qualitative data emerging from cognitive studies. In particular, the knowledge that tasks of differing complexity and nature may dramatically affect the type of processing humans engage in has relevance for designing evaluation studies that take into account not only the expertise of decision makers but also the nature of work tasks, as will be described below.

Closely related to the cognitive continuum in the study of medical cognition is the concept of a continuum of knowledge and expertise that decision makers bring to bear in complex domains such as health care. Expertise in health care can be considered to lie along a continuum, ranging from novices (e.g., medical/nursing students) to intermediates (e.g., medical residents) to experts (e.g., accomplished physicians) [10,11]. Furthermore, the development of expertise in healthcare-related areas is characterized by transitions from novice through to intermediate and expert levels. For example, expert decision making is often characterized by what can be considered along Hammond's cognitive continuum as recognitional processes, where an expert can often quickly arrive at a solution based on analysis of current data in conjunction with his or her knowledge base of similar or related situations. Other characteristics of expert decision making and cognitive processing in general include a greater emphasis on situational analysis of complex problems prior to applying a solution, as exemplified by the research findings of Klein and associates in studying expert decision making in areas ranging from fire fighting to medicine [12].

An important methodological approach to the study of cognition that brings to bear and integrates consideration of both the complexity of decision-making tasks and the prior expertise/knowledge of decision makers is known as *cognitive task analysis* (*CTA*). CTA is a powerful methodological framework for studying and analyzing complex human cognition [13]. In addition, it has been successfully extended to methodological frameworks for studying complex decision-making and reasoning processes of users of computer systems, and as such has definite relevance and relation to methods emerging from usability engineering. The focus of CTA is on the application of scientific and analytical approaches to understanding how people process complex information, reason, and make decisions while undertaking tasks of varying levels of complexity. In contrast to a predominant paradigm in the study of decision making, which has focused on the "decision event" (some hypothesized point in time when the decision maker is supposed to weigh alternatives and arrive at a decision), CTA focuses on understanding the entire process involved in reasoning and decision making, starting with the way a subject first analyses and sizes up a problem or task, and how he or she then proceeds to acquire and process relevant information and finally come up with a decision or course of action.

One approach to CTA typically involves giving subjects (e.g., healthcare workers) specific tasks involving decision making and reasoning, and observing the process of how a decision is made or a problem is solved. This may involve asking subjects to "think aloud" as they process and work through problems in their work domain (that may be presented to them as artificial cases or alternatively as they react to real cases and situations in their work area, as will be described below in the context of situating evaluation along a continuum from experimental to naturalistic approaches). Typically, the entire session is recorded (e.g., audio and video recorded) for further analysis. In addition, CTA may involve comparison of how subjects (e.g., healthcare workers) of varying levels of expertise deal with the same cases (e.g., asking both novices and experts to process medical cases and comparing the differences in their strategies and approaches to problem solving). Currently, there are several research streams from which CTA has emerged, including the study of expertise in the study of problem solving as a basis for design of intelligent tutoring systems [13–15], cognitive engineering [16,17], and the naturalistic study of decision making in domains such as fire fighting and medicine [12].

In our work, we have applied scientific methods from cognitive task analysis (both for setting up studies and for analysis of process data) to the study of how subjects with differing levels of expertise (both in health care and in technology) interact and reason while using computer systems. In this context, the next section presents a discussion of how we have extended and integrated methods from cognitive task analysis with approaches to evaluation collectively known as usability engineering methods.

Usability Engineering Methods and Approaches to Support Cognitive Analysis in Health Informatics

Our work in the evaluation of health information systems has borrowed from research in a number of areas, including cognitive science (as described above), and also from work in the emerging area of usability engineering [18]. Usability engineering has emerged from the integration of evaluation methods used in the study of human–computer interaction (HCI) aimed at providing practical feedback into the design of computer systems and user interfaces. Usability engineering can be distinguished from traditional systems engineering approaches by emphasis on obtaining continual input or feedback from end users, or potential end users of a system, throughout the SDLC. In healthcare settings, a number of researchers have begun to apply methods adapted from usability engineering toward the design and evaluation of clinical information systems. This has included work in developing portable and low-cost methods for analyzing use of healthcare information systems, along with a focus on developing principled qualitative and quantitative methods for analyzing usability data resulting from such study [19]. Since the mid-1990s, a number of groups and laboratories involved in clinical informatics have emerged for testing and designing software applications. For example, Elkin and colleagues describe the use of a usability laboratory for testing a medical vocabulary embedded within the Unified Medical Language System [20]. Kushniruk, Patel, Cimino, and Barrows also describe the use of usability engineering methods for evaluating the design and refinement of a user interface to a CPR system and the analysis of the system's underlying medical vocabulary [21]. Coble and colleagues have described the use of usability engineering in the iterative development of clinical workstations [22]. Others have focused on these methods to deal with the "inspection" of user interfaces [23,24]. Recent work in biomedical informatics has attempted to extend the emerging trend toward usability engineering to include consideration of cognitive issues surrounding design and implementation of clinical information systems, namely cognitive engineering [24,25].

There are a number of specific methods associated with usability engineering, and foremost among these is *usability testing*. Usability testing refers to the evaluation of information systems that involves testing of participants (i.e., subjects) who are *representative* of the target user population, as they perform *representative* tasks using an information technology (e.g., physicians using a CPR system to record patient data) in a particular clinical context. During the evaluation, all user–computer interactions are typically recorded (e.g., video recordings made of all computer screens or user activities and actions). Types of evaluations using this approach can vary from formal, controlled laboratory studies of users, to less formal approaches. Principled methods for the analysis of data from such tests, which may consist of video recordings of end users as they interact with systems, can now be used as tools to aid in the analysis. These techniques generally include the collection of "think aloud" reports, involving the recording of users as they verbalize their thoughts while using a computer. Over the past decade, in the technology industry a range of commercial usability laboratories have appeared for conducting usability testing, ranging from elaborate laboratories with simulated work environments and one-way observation mirrors [26,27], to less elaborate facilities and even portable approaches to usability testing, where the recording equipment is actually taken out to field sites [28]. Many of these techniques borrow from work in the application of cognitive science to the study of human–computer interaction [19,29,30]. The practical role of usability engineering in the development life cycle of clinical information systems has also come under consideration, particularly in the context of use of rapid prototyping methodologies for the design of healthcare information systems [2,22]. Such methods differ from traditional life cycle models, where a system is developed over time using an approach involving fixed stages with limited input from users into redesign. In contrast, rapid prototyping methods typically involve the development of *prototypes* (defined as partially functioning versions of a system), which may be shown to users early in development process in order to assess their usability and functionality. If such assessment indicates that changes are needed, a further cycle of design and testing is initiated. This process continues until the system is deemed to be acceptable to users and shows the desired functionality.

The understanding of how complex information technologies can be successfully integrated into the process of human decision making and practical day-to-day use is critically important in increasing the likelihood of acceptability. Information from usability testing regarding user problems, preferences, suggestions, and work practices can be applied not only toward the end of system development (to ensure that systems are effective, efficient, and sufficiently enjoyable to achieve acceptance), but throughout the development cycle to ensure that the development process leads to effective end products. There are a number of points in the systems development life cycle (SDLC) at which usability testing may be useful in the development of new technologies. As described above, the typical SDLC is characterized by the following phases, which define major activities involved in developing software: (1) project planning, (2) analysis (involving gathering of system requirements), (3) design of the system, (4) implementation (i.e., programming), and (5) system support/maintenance [6]. There are a number of types of usability tests, based on when in the development life cycle they are applied: (1) *exploratory tests* conducted early in the systems development cycle to test preliminary design concepts using prototypes or storyboards; (2) *testing of prototypes* used during requirements gathering; (3) *assessment tests* conducted early or midway through the development cycle to provide iterative feedback into evolving design of prototypes or systems; (4) *validation tests* conducted to ensure that completed software products are acceptable regarding predefined acceptance measures; and (5) *comparison tests* conducted at any stage to compare design alternatives or possible solutions (e.g., initial screen layouts or design metaphors). From this perspective, evaluation in health informatics is seen as being essential throughout the entire life cycle of systems, not just for summative final evaluation.

Cognitive Methods Applied to the Usability Testing of Clinical Information Systems

Given the motivation for applying usability engineering in a clinical setting described earlier, in this section we describe a methodological framework for applying cognitive methods in the evaluation of healthcare information systems. The framework is based on a series of phases employed in performing usability evaluations of healthcare systems and user interfaces extending ideas from both cognitive science and usability testing [19,31,32]. Although there may be some variations in the phases, our evaluation of information systems has typically involved consideration of each of the phases.

Phase 1: Identification of Evaluation Objectives

Possible objectives for conducting evaluations can range considerably, including but not limited to the following examples: (1) assessment of system functionality and usability, (2) input into refinement of emerging prototypes, (3) identifying problems in human–computer interaction, (4) evaluating the effects of a system on decision-making processes, and (5) assessing the impact of a new information technology on clinical practice and workflow. The approach described below can be used to provide practical input into system redesign (e.g., identifying problems with human–computer interaction that need to be rectified).

Phase 2: Sample Selection and Study Design

The second phase involves the identification and selection of a sample of target subjects for the evaluation, resulting in a clearly defined *user profile* that describes the range of skills of target end users of a system. Subjects should be representative of end users of the system under study. For example, if a system is being designed for implementation for use in a particular clinical setting, subjects could consist of personnel who are representative of those who would be expected to actually use the system (e.g., if the system is designed to be used by residents and junior attending staff, it is important to select test subjects that are representative of these groups). Criteria need to be applied for classifying subjects in terms of their prior computer experience. Although there are a number of ways of categorizing users, in our work on usability we have found that considering users along the following dimensions is often useful: (1) expertise of subjects in using computers, (2) the roles of the subjects in the workplace (e.g., physicians, nurses, etc.), and (3) subjects' expertise in the domain of work the information system is targeted for. As evaluation involving cognitive analysis provides a rich source of data, a considerable amount of information may be obtained from a small number of subjects (e.g. 8 to 10 subjects in a group being studied), particularly if subjects selected are representative of target users of the system being assessed.

In addition to describing the tasks that different types of users will be expected to perform using a system, it is also important to describe as much as possible the most critical skills, knowledge, demographic information, and other relevant information about each class of users. Much of our work is an extension of the "expertise approach" [33], which involves comparison of problem solving of subjects with different levels of expertise, to the testing and evaluation of health information systems.

Number of Subjects

Prior studies have shown that carefully conducted usability studies involving as few as 8 to 10 subjects can lead to identification of up to 80% of the surface-level usability problems with an information system [18]. However, more subjects are required in order to conduct inferential statistics (e.g., 15–20 per study group).

Study Design

The study design of our evaluations borrows from approaches in experimental psychology, with a number of options for conducting practical assessments. Study designs may consist of within-group designs where individual subjects may be asked to try out different versions of a prototype system, or one or more subjects may be followed over time as they learn how to use a system. Alternatively, studies may involve between-group designs. Between-group testing might involve, for example, comparison of two different systems, with two groups of different healthcare workers using each system for conducting the same task, such as physicians or nurses looking up patient information in a CPR system. Furthermore, testing may involve use of a CPR system by two groups of subjects of the same medical designation (e.g., attending physicians), one group of which have been identified as being highly computer literate (based on a background questionnaire) and the other group with little experience with computer systems. Within-group studies may focus on longitudinal study of how healthcare workers learn to use and master clinical information systems over time, with testing occurring at specific intervals following initial training in use of a system [3]. Simpler study designs might consist of having a single group (for example, 10 to 15 physician subjects) interacting with a CPR system (with each subject carrying out the same task or set of tasks) in order to assess problems with the design of the user interface.

Phase 3: Selection of Representative Experimental Tasks and Contexts

Studies of use of systems can be situated on a continuum ranging from controlled laboratory studies (e.g., studies involving artificial conditions or tasks) to naturalistic studies of doctor–patient–computer interaction involving use of computer systems in real contexts (e.g., tasks involving subjects being asked to interview a patient while entering data into a computerized patient record system). For laboratory-based evaluations involving controlled experimental conditions, we have sometimes used written medical case descriptions, or vignettes, to be used as stimulus material (e.g., subjects may be asked to develop a diagnosis in response to presentation of a hypothetical or real medical case, while using a CPR). The development of medical cases for use in such studies (often consisting of short written descriptions) may require careful design so that the cases are realistic and representative of real-life clinical situations and elicit high-quality data about user interactions. For example, cases or scenarios can be drawn or modified from the type of cases commonly used for evaluation in medical education, or presented in medical textbooks or journals such as the *New England Journal of Medicine*. They can also be generated from real health data with the assistance of an appropriate medical expert working with the investigators.

Naturalistic studies of actual doctor–patient interactions sacrifice ability to experimentally control the study for an increase in ecological validity (e.g., collection of data on use of a system in a real clinical setting). In naturalistic studies we generally do not present subjects with artificial written cases, but rather monitor the use of systems (using recording methods to be described below) in real clinical contexts (e.g., a physician using a CPR while interviewing a patient). Regardless of the desired level of experimental control, tasks chosen for study should be representative of real uses of the information technology being evaluated.

Phase 4: Selection of Background Questionnaires

A background questionnaire may be given either before or after actual testing of a subject's interaction with a system being evaluated. This questionnaire can be used to obtain historical information about the participants that will help the evaluators to understand their behavior and performance during a test. These can include items to assess level of subjects' typical health practice, or prior experience with computer systems [34]. Some usability tests may include examination of educational systems, where the focus is on assessing how much learning takes place during the process of use of a system (e.g., a Web-based educational resource). This may involve the presentation of questionnaires or multiple-choice test items before and after testing using a system. For example, in conducting an evaluation of physicians using an educational software system on a specific topic (e.g., advances in breast cancer treatment), subjects were given a set of multiple-choice questions to assess their knowledge of that topic both before and after actually recording them interacting with the system, in order to assess the impact of their interactions with systems on their knowledge and learning.

The actual task scenarios to be used during testing also need to be developed during this phase. These may range from simple written descriptions of medical cases, to more elaborate scripts for conducting simulated doctor–patient interviews, where an experimenter plays the part of a patient while the subject interviews or interacts with the "patient" while using a technology such as a CPR system [3].

Phase 5: Selection of the Evaluation Environment

The next step is the selection of the evaluation environment (i.e., where the evaluation will take place). The physical location of the evaluation can vary considerably depending on the degree to which the study is conducted under controlled experimental conditions or in a naturalistic setting. As described in the Introduction to this chapter, a number of fixed laboratories have arisen where commercial organizations conduct testing of developing software products in domains ranging from the aerospace industry to brokerage [27]. During the 1990s there was a trend toward the development of large and expensive fixed commercial usability laboratories, which included simulated environments for testing use of systems (e.g., simulated classrooms or work environments). Such laboratories may consist of testing rooms (containing computer systems with which subjects interact) and adjoining observation rooms with one-way mirrors, for experimenters to watch subjects. However, it has been shown that many of the methods of usability engineering can be applied in a more cost-effective manner, using inexpensive and portable equipment that can be taken to actual work settings. For example, Cimino and colleagues have described the development of a portable usability laboratory for use in clinical settings [35]. For the majority of our studies we have adopted such a portable discount usability engineering approach that involves video recording of subjects in the most convenient setting possible, in some cases right in the hospital or clinic under study [21].

Phase 6: Data Collection—Video Recording and Recording of Thought Processes

Instructions given to subjects may include asking subjects to perform particular tasks using the computer system (e.g., "Please enter data into the computerized patient record system we are testing while 'thinking aloud' or verbalizing your thoughts"). In addition, instructions might involve asking a physician to conduct a doctor–patient interview while using a system, with full video recording of computer screens and concurrent audiotaping of the doctor–patient dialogue [23]. In some studies subjects may also be prompted by experimenters at key points in their interaction with a system to comment on aspects of a system or its design. For example, a study might involve comparison of two screen layouts and for each layout the experimenter might ask the user to comment on the screen's layout. In most of our studies the complete interaction of the subject, starting with the initial instructions to completion of all tasks asked of the user, is video and audio recorded (using equipment such as that detailed below).

Think-Aloud Reports

The collection of "think aloud" reports is one of the most useful techniques emerging from cognitive science. Using this approach, subjects are instructed to "think aloud" (i.e., verbalize their thoughts) as they interact with computer systems (while the computer screens are recorded). There is a principled formal method for analyzing such qualitative data. In our studies of human–computer interaction (HCI), we typically capture the computer screens using video recording (with the computer screen output to a PC–video converter and then input into a VCR) or screen capture software (e.g., the commercially available HyperCam screen recorder software) for detailed analysis of actions, such as mouse clicks and menu selections. The data collected of users' interactions typically include the video recording of all computer screens along with the corresponding audio recording of subjects' verbalizations as they use the system under study [21].

Equipment typically consists of a PC–video converter, for converting the output of computer screens to video (to go into the video-in of a VCR). This allows for recording of all computer screens to video as a user interacts with an information system. In addition, we record all subject verbalizations by using a microphone that inputs into the audio-in of the same VCR. Thus on a single videotape we can record all computer screens and user verbalizations made while a subject performs a task using the computer system under study [31].

A schematic diagram illustrating one approach to collecting video and audio recordings of user interactions with a computer system under study is given in Figure 6.3. In order to obtain video recordings of computer screens, a commercially available PC–video converter is used to convert the VGA computer display output to the video input (i.e., the video-in jack) of a standard VCR. In order to obtain concurrent audio input to the recording of the user–computer interaction we have employed a standard microphone connected to a standard audio mixer (available at most audio stores) or preamplifier, which then outputs into the audio-in jack of the same VCR being used to record computer screens (using a standard RGA cable). This approach allows for recording of user interactions both in the usability laboratory setting as well as in actual clinical settings, since the equipment required is both standard and portable. In a recent paper by Kaufman et al., the use of an inexpensive PC–video converter is described for collect-

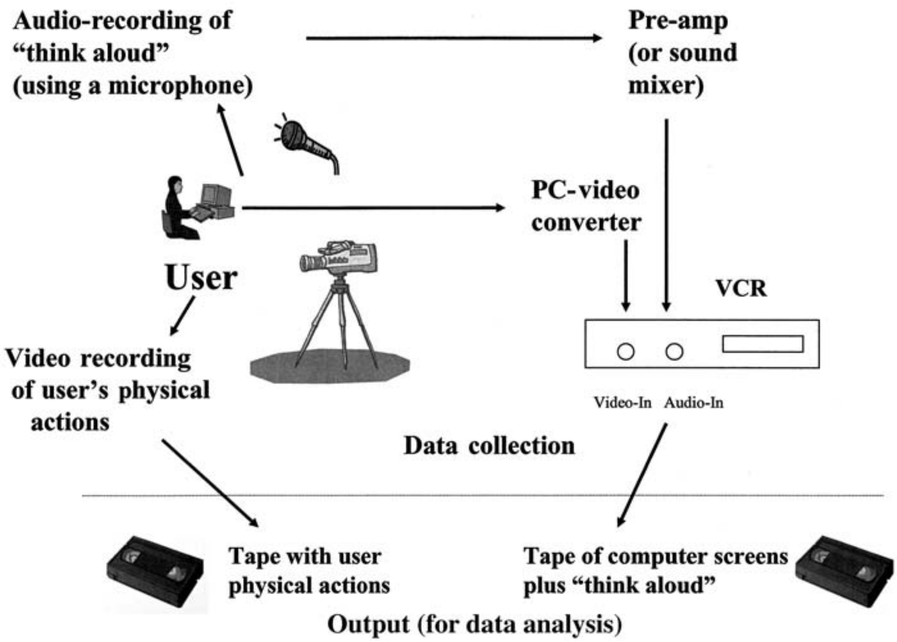

FIGURE 6.3. Video-based usability testing.

ing video data portably [36]. In that study, portable recording equipment was taken to the homes of patient subjects, where complete recordings of subjects' interaction with a diabetes management system were made. The result of this phase includes a complete video recording of user interaction with a computer system along with the audio track containing the verbalizations of subjects interacting with the system.

As indicated in Figure 6.3, video recordings of the actual users themselves (e.g., the faces and gestures of the users as they interact with systems under study) may also be obtained on a separate video recording, although for many of the types of analyses described below, the recordings of computer screens and concurrent audio may be sufficient. If recordings of the actual user are required (e.g., in a study of use of a CPR system where we may want to record how often a physician uses the system as well as physically interacts with other objects such as notes or papers on the desk) in addition to the computer screen recording, this can also be conducted in a cost-effective manner (without requiring the use of an expensive usability laboratory) by using a separate video camera and tripod directed at the user, or users, of the system (see Figure 6.3). In studies requiring unobtrusive observation of user physical interactions with the system, rooms having video cameras placed in unobtrusive locations (e.g., ceiling-mounted cameras) are ideal. In our work in hospital settings, we have on occasion conducted such recordings in rooms that are typically used for other purposes (e.g., rooms outfitted with ceiling-mounted cameras used by medical educators in evaluation of resident and student interviewing skills).

In addition to using standard video recording equipment for recording user interaction with a system, in some studies we have employed a range of software that allows for the recording of screens and audio as movie files directly on the computer being used for testing, removing the need for video cameras and VCRs for recording of the computer screens. For example, the commercially available product HyperCam allows for direct recording of the computer screens, along with audio input to the same computer via a computer microphone. However, due to storage requirements of such approaches (the resulting recordings are stored as large files that may quickly exceed storage allocation on a standard PC), in many studies we continue to employ standard video recording techniques described above, particularly when collecting data in real clinical settings, where the computer equipment and capabilities may be more limited than in the laboratory.

Phase 7: Analysis of the Process Data

The output of Phase 6 may consist of video recordings of computer screens (with an audio overlay of the subject "thinking aloud") and/or a tape of the actual user's interactions with the computer system (e.g., facial expressions, movements, gestures etc.). In many studies, the objective of the evaluation may be to analyze such data to identify problems subjects experience in using a system (e.g., a computerized patient record system or a decision-support system). The transformation of data into recommendations involves qualitative and quantitative analyses of the video-based usability data. The advantages of video recordings as a source data include the fact that videotapes of user–computer interactions provide a record of the "whole event." Furthermore, the same video recordings of user interactions can be examined from a number of theoretical perspectives and analyzed using a range of methodological approaches.

There are a variety of approaches to analyzing data on human–computer interaction from video data, ranging from informal review of the resulting taped data, to formalized and precise methods for analyzing the number and type of errors or user problems. The richness of video data requires principled methods for conducting full analysis and coding. The use of computer tools to aid the analysis of video data has greatly facilitated usability testing [19]. Computer programs are now available that interface between VCR and computer in order to facilitate video coding. A software tool we used extensively in our earlier analyses was called CVideo (Envisionology Inc.)—a program that allowed the verbal transcriptions (e.g., of subjects' "thinking aloud") to be annotated on a MacIntosh computer and linked (time-stamped) to the corresponding video sequence (using a cable that connects the Mac to the VCR while reviewing the tape of a usability testing session). In recent years a number of tools have become commercially available for assisting in the qualitative analysis of audio and video-based data (including MacShapa, Transana, and other related software tools for conducting qualitative analyses that allow for interfacing and indexing of video data). Computer-supported analysis of video data allows researchers to document video frames with textual annotations, notes, and codes on a computer, saving time in analysis, and allows for automatic indexing and retrieval of video frames and sequences. Such analyses also facilitate interrater reliability in coding and allow for coding of user actions and verbalizations.

The procedure for data analysis we employ first involves having the audio portion of the test session ("think aloud" reports) transcribed separately in a word processing file. That file then serves as a computer-based log file for entering annotations and codes that are linked or time-stamped to the corresponding video scenes [21]. However, it should be noted that for the types of analyses described below (involving application of coding schemes), computer-supported coding tools are not a requirement for conducting principled analysis of video data. The coding tool will aid in the annotation of the transcripts by linking the computer word processing file containing the transcripts to the actual video tape sequences. However, this can be also accomplished manually, that is, by watching the videotape and entering into the word processing file containing the audio transcripts the actual corresponding video counter numbers (as will be illustrated below).

Application of a Coding Scheme in Analyzing Video Data

Prior to analyzing video data, a coding scheme should be refined for use in identifying specific occurrences of user problems and aspects of cognitive processes from transcripts of the subjects' thinking aloud and interactions with a computer. Coding categories we have applied in a number of studies include the following: *information content* (e.g., whether the information system provides too much or too little information, etc.), *comprehensiveness of graphics and text* (e.g., whether a computer display is understandable to the user), *problems in navigation* (e.g., whether the user has difficulty in finding desired information or computer screen), and *overall system understandability* (e.g., understandability of icons, required computer operations, and system messages). In addition to these categories, which focus on classical aspects of HCI, one can also extend the analyses to allow for the identification of higher-level cognitive processes. For example, in some studies we code each occurrence of the generation of a diagnostic hypothesis by a subject, or request for information from a patient in the case of studies of doctor–patient interaction involving use of a CPR system.

As an illustration, to assess ease of use of computer systems, a coding system can be used as shown in Figure 6.4. The scheme shows definitions of coding categories, along with examples of coded statements made by test subjects while interacting with a system that fall under each category (an example of a coded transcript will be provided below in our discussion).

1. Examples of Categories Used to Analyze Aspects of Interaction Related Directly to the User Interface:

NAVIGATION
Coded when subject comments on basic navigation, or indicates can't move through program/interface, etc. to find or go somewhere (e.g., "How do I get back to the last screen?").

LAYOUT/SCREEN ORGANIZATION
Coded for if the subject comments on the layout or screen organization (e.g., "I find this page very cluttered").

MEANING OF LABELS
Coded for if the subject comments on the meaning of labels in the interface itself (e.g., "I don't know what this button means here, the one that says "free download" on it").

UNDERSTANDING OF SYSTEM INSTRUCTIONS/ERROR MESSAGES
Coded for if the subject comments on understanding of instructions or errors (e.g., "It says 'fatal error 404' and I don't know what to do now").

CONSISTENCY OF OPERATIONS
Coded for if the subject comments on the consistency of operatons (e.g., "How come there are two different ways to exit on the last two creens?").

OVERALL EASE OF USE
Coded for if the subject comments on the overall ease of use (e.g., "I find this system very hard to use").

RESPONSE TIME
Coded for if the subject mentions response time (e.g., "I seem to be waiting a very long time, and still there is no response from the computer").

VISIBLITY OF SYSTEM STATUS
Coded for if the subject comments on visibility of system status (e.g., "I'm not sure what the system is doing now—it seems to be hanging").

2. Examples of Categories Used to Analyze Aspects of Medical Reasoning/Decision Making Processes:

REQUEST INFORMATION
Coded when subject (e.g., physician) requests information from patient during a doctor–patient interaction (e.g., "How often do you smoke?").

CONSIDER DIAGNOSTIC HYPOTHESIS
Coded when subject considers a diagnostic hypothesis (e.g., "I think this patient has angina").

CHOOSE TREATMENT
Coded when subject chooses a medical treatment (e.g., "At this point I would put him on heparin").

Figure 6.4. Excerpts from a coding scheme for analyzing video-based data from cognitive evaluations.

The coding scheme essentially forms a manual for researchers as they watch and annotate the videotapes obtained from experimental sessions. The categories used for coding were developed from examination of categories of interactions from the HCI and cognitive literatures [37,38].

In Figure 6.5, we show the application of coding categories (from Figure 6.4) in analyzing a video log of a user's interaction with a CPR. The

00:00:00 Start of session—user starts up the system.

"I'm just starting up the system; well, what a nice office this is, here we go. This is for patient X. Well I have just tried turning the thing on. OK, here is the first screen, but it looks very poorly laid out."
LAYOUT/SCREEN ORGANIZATION—PROBLEM

00:01:15 User goes to help screen.

"How do I move to the previous screen?"
NAVIGATION—PROBLEM

"I think this patient may have diabetes. So here I go, I'm clicking on this screen about diabetes information."
CONSIDER DIAGNOSTIC HYPOTHESIS—DIABETES

00:01:23 Subject goes to diabetes guideline screen and clicks on help button.

"Now what? It looks like everything has stopped."
LACK OF INDICATION OF SYSTEM STATUS—PROBLEM

Figure 6.5. Excerpt of a coded section of a transcript of a user (a physician) interacting with a CPR.

procedure for analysis of the subjects' thinking aloud is based on the method of protocol analysis, as described in detail by Ericsson and Simon [29]. Note that the transcript of the subject's thinking aloud report is marked up with annotations from the coding scheme and that the numbers in the log file (containing the transcript) refer to the corresponding section of the videotape (i.e., the video counter number) where they occurred. Also note that codes that indicate user problems are coded as such (with the additional coding tag "PROBLEM").

We have found that up to 80% of user-interface problems with a particular clinical system can be detected with as few as 8 to 12 transcripts of subjects' interaction with the system under study, which is consistent with the literature emerging from the application of cognitive engineering methods in HCI [18].

Important advances have been made in the development of computer-based tools that aid in the detection and analysis of patterns contained in usability data. In our studies, we have developed a variety of schemes for analyzing video data in a principled manner. These allow coders to identify events of interest, such as user problems, and use of system features (preliminary schemes are typically refined and then verified). Coding schemes can include categories for user/system aspects and problems including categories for human factors issues and cognitive issues. We have developed categories that characterize at a top level the following aspects of human–computer interaction: (1) the *usefulness* of the system being tested in terms of its contents, and (2) the *ease of use* of the system or interface. The first top-level category deals with issues such as whether the system being tested provides useful, up-to-date or valuable information to a user, while the second category characterizes potential problems or issues related to the actual user interface or system design. The coding schemes we have developed are based on and extend categories that have been applied in protocol analysis in the study of medical cognition (see [37] for details). In particular, our coding schemes contain categories used to assess key aspects of medical decision making and reasoning (e.g., choice of treatment) in addition to categories used to code for aspects of usability, allowing us to relate aspects of user interfaces (and their usability) to reasoning and decision-making processes.

Phase 8: Interpretation of Findings

The data collected from usability testing can be compiled and summarized in numerous ways, depending on the goals of the evaluation. The results may summarize any number of aspects of system use, including task accuracy, user preference data, time to completion of task, frequency, and classes of problems encountered. In addition, qualitative analyses of the effects of the technology on healthcare professional reasoning and decision making can be conducted. Results of process evaluations may include a summary of types and frequency of problems that occur when subjects interact with a computer system under evaluation. If the system under study is under development, the information provided from the analysis phase should be communicated to system designers. For further investigations, the findings should be interpreted for what they mean, within the context of the theoretical framework.

Phase 9: Iterative Input into Design

After implementation of changes to a system, based on the recommendations to the programming team (for studies involving formative evaluations), evaluation may be repeated to determine how the changes now affect the system's usability. In this way, evaluation can be integrated in the process of design and development of information systems, iteratively feeding information back into their continual improvement.

Application of Cognitive Approaches to Evaluation: From Medical Informatics to Consumer Informatics—E-health and Beyond

Cognitive approaches to system evaluation can be applied throughout the life cycle of information systems, to answer a range of evaluation questions. In this section of the chapter we describe some of our experiences in applying a cognitive approach to the practical evaluation of a range of types of health information systems. In our initial work along these lines, we have applied the approach to the evaluation of educational software designed for use in continuing medical education. In one study, subjects (physicians involved in a continuing education program) were given the task of exploring a multimedia tutorial in order to improve their knowledge about the treatment of heart disease. After completing a pretest multiple-choice test to assess their prior knowledge in this area, subjects were video recorded as they interacted with the system while asked to think aloud. After completion of the task, subjects were given a follow-up questionnaire (containing the same questions) to assess if the subjects had improved their knowledge of heart disease by interacting with the system. All of the subjects' interactions with the system were video recorded. The audio portions of the sessions were transcribed and the transcripts were coded to identify problems and issues in using the system. The study approach was used both to assess learning that took place while interacting with the system, as well as to identify from the video data specific problems with the interface that needed improving (e.g., use of more meaningful icons, better navigational facilities, etc.).

Following from this initial work in applying a cognitive task analysis approach to assessing an educational program, we began a line of research into assessing the effects of emerging clinical information systems, in particular CPRs, on the decision making and reasoning of healthcare professionals [3,23]. A range of studies were conducted with the objective of evaluating the effects of introduction of a CPR system on physician decision-making and reasoning processes in a diabetes clinic. One component of this research program involved in-depth cognitive analysis of 14 subjects learning how to use and master the system over a six-month period. Subjects were video recorded as they entered information into the system (all computer screens were recorded) and subjects were also asked to think aloud while they interacted with the system. In addition, in another experimental condition, subjects were asked to interview a "simulated" patient (i.e., a research collaborator playing the part of a patient, a technique used in the evaluation of medical trainees' interviewing skills) and their interaction with both the computer system and the patient were video recorded. By both analyzing the data obtained from the experimental condition involving subjects thinking aloud, as well as analyzing the data from recording subjects interacting with simulated patients over time as they learned to use the CPR and became familiar with its capabilities, the effects of use of the system on physicians were assessed. Through analysis of both video and audio data it was found that the layout of the information on the CPR screen had a significant impact on the way the physician subjects interacted with patients and reasoned about patient cases. Specifically, it was found that as the physicians became familiar with the system they became guided by the order and organization of medical findings on the computer screen in requesting information from patients, which ultimately affected reasoning about patient cases, a pattern of interaction with patients we described as being "screen-driven." The implications of such findings of unexpected yet profound effects of CPR systems on physician information processing and reasoning have had important impact in the design of subsequent CPR user interfaces.

In a second line of studies, we applied cognitive approaches to the evaluation of emerging CPR systems at Columbia University [21]. This line of research involved the evaluation of both the user interface and the underlying medical vocabulary of a CPR system. Subjects, consisting of nine physicians, were initially asked to enter information from paper records into the CPR. Full audio transcripts of the subjects' thinking aloud were made, along with video recording of the corresponding computer screens as the subjects transferred information from paper records into the new CPR system. The approach to analysis involved annotation of the audio and video transcripts, using the method described above, to identify the frequency of categories of problems related to both the usability of the interface and the effectiveness of the underlying medical terminology to represent information about the patients' condition. Based on the analysis of the data, it was found that users found use of the system difficult due to design problems ranging from lack of consistency of the user interface (e.g., multiple and confusing ways to carry out procedures such as data entry) to problems in representing medical findings using the system. The frequency of particular usability problems was compiled and presented to the design team and consequently changes were made to the CPR based on the recommendations. Subsequent usability testing with a new set of nine different physicians (who had not used the system before) indicated that the number of problems had decreased from an average of 19 problems per user testing session prior to the suggested changes, to 1.9 problems per user session after the changes were applied. Work such as this has underlined the value and effectiveness of employing cognitive approaches to evaluation in improving the usability of healthcare information systems during the iterative process of system design and implementation.

In another line of research we have been involved in the evaluation of a number of information systems targeted to patient users of health information systems. In a recent study Kaufman and colleagues used a cognitive task analysis approach involving usability testing methods to assess use of a home-based telemedicine system for diabetes [36]. The interactions of 25 subjects, ranging in age and educational background, were recorded in their homes using portable recording equipment. In another related study the usability of an experimental text summarization system was compared to three commercial search engines [39]. This study involved having subjects (consisting of family members of patients in the hospital for cardiac surgery) pose their questions to the different search engines while thinking aloud. Based on this approach we found that although no one search engine was favored by all subjects, there were specific features of each of the systems that users invariably liked. The results have been extended to the design of new approaches to providing information to patient users, based on a reverse-engineering approach (i.e., based on the results of our analyses of search engine use by patients).

In a more recent line of research, we have extended the overall approach to evaluation of information systems described in this paper to the remote evaluation of Web-based information systems, an approach we call "televaluation" of healthcare information systems. Our first work along these lines involved the distance evaluation of a Web-based patient information system, known as PatCIS, which allows patients at home to access their own patient records over the Internet [40]. Data were collected from both in-depth study of individual users interacting with the system as well as statistics on usage of the different components (e.g., advice, review of medical information, and links to educational resources). In one set of studies, we developed an "evaluation server" that intercepts a user's request for information from a Web-based information resource and can automatically query the user for his or her impressions regarding usefulness of the information obtained from the information resource. Thus in this recent line of work we are moving toward automated evaluation of use of Web-based information resources and systems in healthcare and extending the concept of task analysis to include automated probing and tracking of users as they interact with systems remotely.

Our work has shown that cognitively based analyses of information systems can be applied throughout the systems development life cycle, and in a recent work it has been shown that the approach can be extended to the early analysis of systems requirements as a basis for systems design. Along these lines, Cysneiros and Kushniruk have recently described the development of an ontology for classifying and reasoning about cognitive aspects of use of information systems in healthcare and other domains [5]. As the issue of designing improved healthcare systems based on a better understanding of the complexities of the healthcare environment and the varied types of users becomes more widely acknowledged, cognitive approaches will likely increase in importance for evaluating systems (as well as preliminary design ideas) throughout the entire SDLC.

From Laboratory to Naturalistic Evaluations in Health Informatics: A Continuum

Approaches to evaluation of healthcare information systems using cognitive approaches can be located along a continuum of study types ranging from artificial laboratory-based studies at one end of the continuum to naturalistic studies conducted in real work settings at the other end of the continuum (as depicted in Figure 6.6).

At one end of the continuum an attempt is made to conduct studies in controlled artificial conditions. This might, for example, involve use of a

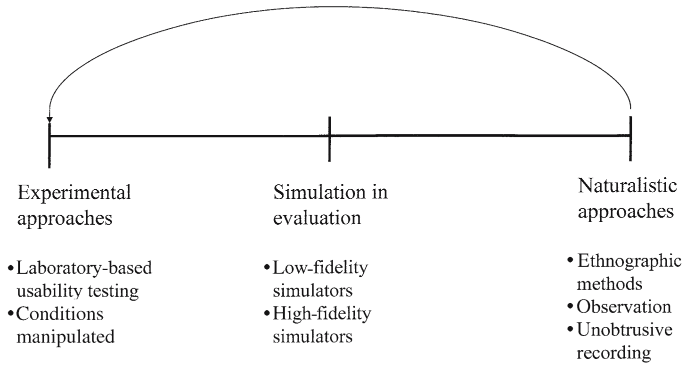

FIGURE 6.6. A continuum of approaches to evaluation of healthcare information systems.

fixed usability laboratory—often consisting of facilities designed for conducting usability testing (with built-in recording devices, ceiling-mounted video cameras, and one-way observation mirrors to view subjects interacting with systems). It should be noted that other approaches to conducting laboratory-type testing of users interacting with information systems also can be carried out using low-cost portable recording equipment. In any case, for this type of study, subjects may be given artificial medical cases as stimulus material (e.g., a written case description) and the procedure often involves subjects thinking aloud or verbalizing their thoughts (which are audio recorded) while carrying out a specific task (e.g., entering the information from the written case description into a CPR that is being evaluated). At this end of the continuum, studies may be designed that exert a higher degree of experimental control with laboratory testing of subjects interacting with the system with only one or a few variables (e.g., display format) manipulated during testing, with the test being conducted under controlled artificial conditions, either in a usability laboratory, or using portable recording equipment (as described in [31]).

Evaluations involving simulation techniques are located halfway along the continuum ranging from controlled to naturalistic approaches to assessment. Such evaluations may allow for a high degree of experimental control while also maintaining a high degree of realism in the tasks presented to subjects during testing. For example, as described above, we have recorded subjects interacting with a CPR system while interviewing a simulated patient, consisting of a research collaborator playing the part of a specific type of patient (borrowing from the concept of a "standardized patient" used for assessing medical residents in medical training). From such studies we have been able to extend our understanding of use of a CPR system from individual physicians interacting with the system to the understanding of how the computer system interacts with the physician in the context of carrying out the task of interviewing a patient in a realistic medical context. A range of other possibilities exist for carrying out evaluations using simulations, including use of high-fidelity computer-controlled mannequins that are now becoming more widely available in medical schools (for providing training to students and residents in areas such as surgery). Such simulators, not unlike their counterparts in areas such as aviation, can be used both for training and also for use in assessment of technology in carrying out work tasks in healthcare.

At the far end of the continuum shown in Figure 6.6 are naturalistic approaches to evaluation. Here user interactions with systems in real-life contexts are monitored with little or no intervention from the evaluators (e.g., recording real use of a CPR system in a doctor's office for entering and retrieving patient data). Also included at this end of the continuum would be studies described above, where use of Web-based information systems is tracked or monitored over time. It has been argued by many that such studies are necessary as results from classical controlled experimental studies may be limited in how well they generalize to real-world situations. In our work we have worked at all points along the continuum, with some of our evaluations beginning with in-depth laboratory study of use of a computer system in healthcare being followed up with collection of data from naturalistic settings. Likewise, study of use of a system using naturalistic approaches (e.g., tracking or logging of real system usage) may lead to specific research questions that may be best answered by applying experimental control and rigor (e.g., following up with laboratory testing of subjects interacting with a system to deal with specific cases using the "think aloud" method).

Cognitive approaches also can be considered in the context of where they can be applied in the systems development life cycle, as depicted in Figure 6.7. From Figure 6.7 it can be seen that approaches to evaluation that are based on ideas and principles from cognitive science and usability

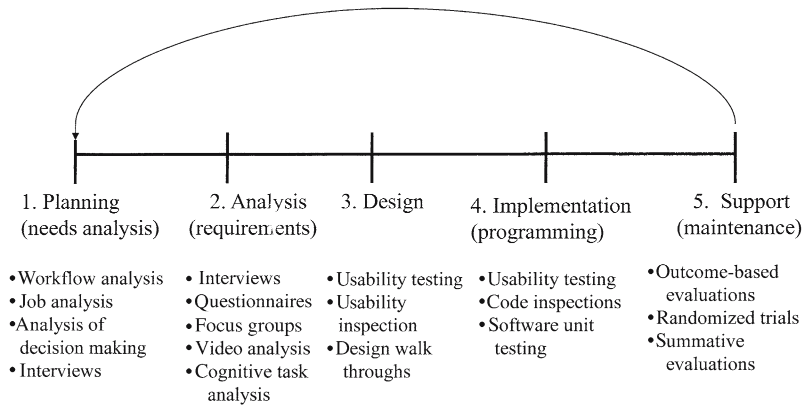

FIGURE 6.7. The systems development life cycle (SDLC) in relation to evaluation methodologies.

engineering can be applied at various points throughout the SDLC. For example, cognitive task analysis and the in-depth recording of subjects interacting with mockups of health information systems can be applied even during early stages of system design. At the other end of the continuum, approaches such as those described in this paper can be applied to assess how completed systems (at the far right of the SDLC in the figure) impact on physician reasoning and decision making in summative evaluation late in the SDLC.

Conclusions and Future Directions

In this chapter we have examined a range of techniques for evaluation of healthcare information systems that borrow from the fields of both cognitive science and usability engineering. The chapter has presented a framework for conducting evaluation at various stages throughout the systems development life cycle (SDLC) in developing healthcare information systems. In this context we consider evaluation to be closely related to system design and implementation within the process of iterative system development. A focus of our work has been on understanding and assessing the impact of new information technologies in healthcare on cognitive processes involved in reasoning, decision making, and using new technology to improve complex work activities. In recent years there has been a move in evaluation of health information systems from a nearly exclusive focus on summative evaluation of completed systems (using methods related to controlled clinical trials) to the formative evaluation of systems being developed in order to lead to their improvement. Furthermore, there has been a newly emerging focus on the analysis of the cognitive *processes* involved using information systems, as such study makes it possible to identify and assess emergent and unexpected effects of these systems on cognitive and work processes.

A challenge for future work will be to integrate data and findings from multiple evaluation approaches (e.g., methods of cognitive task analysis and methods associated with outcome-based evaluations of systems). One area where such synergy will be important is in the evaluation of information systems to ensure patient safety and to lead to design of systems that will reduce error. Recent work in this area has included study of the relationship between cognitive evaluation, using methods such as those described in this chapter, and the analysis of how medical information technology may reduce or introduce error into medical practice. Along these lines, Kushniruk, Triola, Borycki, Stein, and Kannry have demonstrated how coding of usability problems can lead to accurate predictions of actual medical errors resulting from use of medical information systems [41]. Along related lines, Zhang and colleagues have worked on developing taxonomies and frameworks for studying error in medicine based on cognitive analysis [24]. The work we are doing in evaluation is ongoing and constantly being refined as the technology we evaluate changes and advances. Important developments along these lines include work on the automated analysis of qualitative data emerging from cognitive studies [32] and work toward extending many of the approaches described in this chapter to the automated analysis of Web-based information systems from a distance, an approach we have termed "televaluation." As healthcare technology advances, the greatest challenge for evaluation of health information systems will be in understanding the effect of such systems on complex cognitive processes involved in healthcare and in finding ways to apply this understanding in creating systems that facilitate and enhance human information processing.

Additional Readings

V.L. Patel, A.W. Kushniruk, S. Yang, and J.F. Yale, Impact of a computer-based patient record system on data collection, knowledge organization and reasoning, Journal of the American Medical Informatics Association 7(6) (2000) 569–585.

V.L. Patel and D.W. Bates, editors, Special issue of the *Journal of Biomedical Informatics:* Patient safety. Journal of Biomedical Informatics 36(1/2) (2003).

V.L. Patel and D.R. Kaufman, Cognitive science and biomedical informatics, in: E.H. Shortliffe and J.J. Cimino, editors, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, 3rd ed. (Springer-Verlag, New York, in press).

A.W. Kushniruk and V.L. Patel, Cognitive and usability engineering approaches to the evaluation of clinical information systems, Journal of Biomedical Informatics 37(1) (2004) 56–76.

A.W. Kushniruk, Analysis of complex decision making processes in health care: Cognitive approaches to health informatics, Journal of Biomedical Informatics 34(5) (2001) 365–376.

A.W. Kushniruk, D.R. Kaufman, V.L. Patel, Y. Levesque, and P. Lottin, Assessment of a computerized patient record system: A cognitive approach to evaluating an emerging medical technology, MD Computing 13(5) (1996) 406–415.

References

[1] C.P. Friedman and J.C. Wyatt, *Evaluation Methods in Medical Informatics* (Springer, New York, 1997).

[2] A.W. Kushniruk, Evaluation in the design of information systems: Applications of approaches from usability engineering, Computers in Biology and Medicine 32(3) (2002) 141–149.

[3] V.L. Patel, A.W. Kushniruk, S. Yang, and J.F. Yale, Impact of a computer-based patient record system on data collection, knowledge organization and reasoning. Journal of the American Medical Informatics Association 7(6) (2000) 569–585.

[4] J. Moehr, Evaluation: Salvation or nemesis of medical informatics? Computers in Biology and Medicine (2002) 113–125.

[5] L. Cysneiros and A.W. Kushniruk, Bringing usability to the early stages of software development, in: *Proceedings of the 11th IEEE International Requirements Engineering Conference* (2003), pp. 359–361.

[6] S. McConnell, *Rapid Development: Taming Wild Software Schedules* (Microsoft Press, Redmond, WA, 1996).

[7] J. Preece, Y. Rogers, and H. Sharp, *Interaction Design: Beyond Human–Computer Interaction* (Wiley, New York, 2002).

[8] K. Hammond, Naturalistic decision making from a Brunswikian viewpoint: Its past, present, future, in: G. Klein, J. Orasanu, R. Calderwood, and C. Zsambok, editors, *Decision Making in Action: Models and Methods* (Ablex, Norwood, NJ, 1993), pp. 205–227.

[9] R. Hamm, Moment-by-moment variation in experts' analytic intuitive cognitive activity, IEEE Transactions on Systems Management Cybernation SMC-18, 5 (1988) 757–776.

[10] V.L. Patel, J.F. Arocha, and D.R. Kaufman, Diagnostic reasoning and medical expertise, in: D.L. Medin, editor, *The Psychology of Learning and Motivation: Advances in Research and Theory* (Academic Press, San Diego, 1994), (31), pp. 187–252.

[11] V.L. Patel and D.R. Kaufman, Cognitive science and biomedical informatics, in: E.H. Shortliffe and J.J. Cimino, editors, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine* 3rd ed. (Springer-Verlag, New York, in press).

[12] G.A. Klein, A recognition-primed decision (RPD) model of rapid decision making, in: G. Klein, J. Orasanu, R. Calderwood, and C. Zsambok, editors, *Decision Making in Action: Models and Methods* (Ablex, Norwood, NJ, 1993), pp. 138–147.

[13] S. Gordon and R. Gill, Cognitive task analysis, in: C. Zsambok and G. Klein, editors, *Naturalistic Decision Making* (Erlbaum, Mahwah, NJ, 1997), pp. 131–140.

[14] R. Glaser, A. Lesgold, S. Lajoie, R. Eastman, L. Greenberg, D. Logan, M. Magone, A. Weiner, R. Wolf, and L. Yengo, Cognitive task analysis to enhance technical skills training and assessment (Report, Learning Research and Development Center) (University of Pittsburgh, 1985).

[15] B. Means and S. Gott, Cognitive task analysis as a basis for tutor development: Articulating abstract knowledge representations, in: J. Psotka, L. Massey, and S. Mutter, editors, *Intelligent Tutoring Systems: Lessons Learned* (Hillsdale, New Jersey, 1988).

[16] J. Rasmussen, A. Pejtersen, and L. Goodstein, *Cognitive Systems Engineering* (Wiley, New York, 1994).

[17] K.J. Vicente, *Cognitive Work Analysis: Toward Safe, Productive and Healthy Computer-Based Work* (Lawrence Erlbaum, New York, 1999).

[18] J. Nielsen, *Usability Engineering* (Academic Press, New York, 1993).

[19] A. Kushniruk and V. Patel, Cognitive computer-based video analysis: Its application in assessing the usability of medical systems, Medinfo (1995) 1566–1569.

[20] P.L. Elkin, D.N. Mohr, M.S. Tuttle, W.G. Cole, G.E. Atkin, K. Keck, T.B. Fisk, B.H. Kaihoi, K.E. Lee, M.C. Higgins, H. Suermondt, N. Olson, P.L. Claus, P.C. Carpenter, and C. Chute, Standardized problem list generation, utilizing the Mayo Cannonical Vocabulary embedded within the Unified Medical Language System, in: *Proceedings of the AMIA Symposium* (1997), pp. 500–504.

[21] A.W. Kushniruk, V.L. Patel, J.J. Cimino, and R.A. Barrows, Cognitive evaluation of the user interface and vocabulary of an outpatient information system, in: *Proceedings of the AMIA Symposium* (1996), pp. 22–26.
[22] J.M. Coble, J. Karat, M. Orland, and M. Kahn, Iterative usability testing: Ensuring a usable clinical workstation, in: *Proceedings of the AMIA Symposium* (1997), pp. 744–748.
[23] A.W. Kushniruk, D.R. Kaufman, V.L. Patel, Y. Levesque, and P. Lottin, Assessment of a computerized patient record system: A cognitive approach to evaluating an emerging medical technology. MD Computing 13(5) (1996) 406–415.
[24] J. Zhang, T.R. Johnson, V.L. Patel, D.L. Paige, and T. Kubose, Using usability heuristics to evaluate patient safety of medical devices, Journal of Biomedical Informatics 36(1–2) (2003) 23–24.
[25] J. Horsky, D.R. Kaufman, M.L. Oppenheim, and V.L. Patel, A framework for analyzing the cognitive complexity of computer-assisted clinical ordering, Journal of Biomedical Informatics 36(1–2) (2003) 4–22.
[26] J. Rubin, *Handbook of Usability Testing*: *How to Plan*, *Design and Conduct Effective Tests* (Wiley, New York, 1994).
[27] M.E. Wiklund, *Usability in Practice* (Academic Press, New York, 1994).
[28] A.W. Kushniruk and V.L. Patel, Cognitive evaluation of decision making processes and assessment of information technology in medicine, International Journal of Medical Informatics 51 (1998) 83–90.
[29] K. Ericsson and H. Simon, *Protocol Analysis*: *Verbal Reports as Data* (Academic Press, Cambridge, MA, 1993).
[30] J.M. Carroll, *Human-Computer Interaction in the New Millennium* (Addison-Wesley, New York, 2002).
[31] A.W. Kushniruk, V.L. Patel, and J.J. Cimino, Usability testing in medical informatics: Cognitive approaches to evaluation of information systems and user interfaces, in: *Proceedings of the AMIA Symposium* (1997), pp. 218–222.
[32] A.W. Kushniruk and V.L. Patel, Cognitive and usability engineering approaches to the evaluation of clinical information systems, Journal of Biomedical Informatics 37(1) (2004) 56–76.
[33] K. Ericsson and J. Smith, *Toward a General Theory of Expertise: Prospects and Limits* (Cambridge University Press, New York, 1991).
[34] A.W. Kushniruk, C. Patel, V.L. Patel, and J.J. Cimino, "Televaluation" of information systems: An integrative approach to design and evaluation of Web-based systems, International Journal of Medical Informatics 61(1) (2001) 45–70.
[35] J.J. Cimino, V.L. Patel, and A.W. Kushniruk, Studying the human-computer-terminology interface, Journal of the American Medical Informatics Association 8(2) (2001) 163–173.
[36] D.R. Kaufman, V.L. Patel, J.L. Starren, P.C. Morin, C. Hilliman, J. Pevzner, R.S. Weinstock, R. Goland, and S. Shea, Usability in the real world: Assessing medical information technologies in patients' homes, Journal of Biomedical Informatics 36(1–2) (2003) 45–60.
[37] A.W. Kushniruk, Analysis of complex decision making processes in healthcare: Cognitive approaches to health informatics, Journal of Biomedical Informatics 34(5) (2001) 365–376.
[38] B. Shneiderman, Designing the User Interface, 4th ed. (Addison-Wesley, New York, 2003).

[39] A.W. Kushniruk, M.Y. Kan, K. McKeown, J. Klavans, D. Jordan, M. LaFlamme, and V.L. Patel, Usability evaluation of an experimental text summarization system and three search engines: Implications for the reengineering of health-care interfaces, in: *Proceedings of the AMIA Symposium* (2002), pp. 420–424.
[40] J.J. Cimino, V.L. Patel, and A.W. Kushniruk, The patient clinical information system (PatCIS): Technical solutions for and experience with giving patients access to their electronic medical records, International Journal of Medical Informatics 68(1–3) (2002) 113–127.
[41] A. Kushniruk, M. Triola, E. Borycki, B. Stein, and J. Kannry, Technology induced error and usability: The relationship between usability problems and prescription errors when using a handheld application. International Journal of Medical Informatics, in press.