Open Research Online

The Open University's repository of research publications
and other research outputs

A multidimensional evaluation framework for personal learning environments

Book Section

For guidance on citations see FAQs.

Version: Version of Record

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.1007/978-3-319-02399-1_3

oro.open.ac.uk

A Multidimensional Evaluation Framework for Personal Learning Environments

Effie Lai-Chong Law and Fridolin Wild

Abstract Evaluating highly dynamic and heterogeneous Personal Learning Environments (PLEs) is extremely challenging. Components of PLEs are selected and configured by individual users based on their personal preferences, needs, and goals. Moreover, the systems usually evolve over time based on contextual opportunities and constraints. As such dynamic systems have no predefined configurations and user interfaces, traditional evaluation methods often fall short or are even inappropriate. Obviously, a host of factors influence the extent to which a PLE successfully supports a learner to achieve specific learning outcomes. We categorize such factors along four major dimensions: technological, organizational, psycho-pedagogical, and social. Each dimension is informed by relevant theoretical models (e.g., Information System Success Model, Community of Practice, self-regulated learning) and subsumes a set of metrics that can be assessed with a range of approaches. Among others, usability and user experience play an indispensable role in acceptance and diffusion of the innovative technologies exemplified by PLEs. Traditional quantitative and qualitative methods such as questionnaire and interview should be deployed alongside emergent ones such as learning analytics (e.g., context-aware metadata) and narrative-based methods. Crucial for maximal validity of the evaluation is the triangulation of empirical findings with multi-perspective (end-users, developers, and researchers), mixed-method (qualitative, quantitative) data sources. The framework utilizes a cyclic process to integrate findings across cases with a cross-case analysis in order to gain deeper insights into the intriguing questions of how and why PLEs work.

Keywords Evaluation • Multi-method • Usability • User experience • Community of practice • Self-regulated learning • Diffusion of innovation • Cross-case analysis • Automated monitoring

E.L.-C. Law (✉)
University of Leicester, Leicester, UK
e-mail: elaw@mcs.le.ac.uk

F. Wild
The Open University, Buckinghamshire, UK
e-mail: f.wild@open.ac.uk

Introduction

Among others, a critical success factor in technology-enhanced learning is the personalization of learning experience. As emphatically pointed out in the Leuven/Louvain-la-Neuve Communiqué of the Bologna Process 2020, "student-centered learning requires empowering individual learners, new approaches to teaching and learning, effective support and guidance structures and a curriculum focused more clearly on the learner" (p.3). Personalization is also a key issue for implementing mechanisms to foster and increase activities in informal and lifelong learning networks. This implies a need for new technology-enhanced learning models that start from the learners and satisfy their unique needs in order to achieve a personalized learning experience for everyone.

Recent discussions about technologies for learning have shifted from institution-managed learning management systems (LMS) to user-controlled social software for learning. Indeed, the advent of Web 2.0 technologies has phenomenally transformed the way in which users consume, communicate, collaborate, and create information and knowledge on the Web. These technologies have underpinned the emergent notion of *Personal Learning Environment* (PLE), which is characterized by qualities such as personalization, openness, responsiveness, flexibility, shareability, interactivity, and sociability. PLEs can be perceived as both a technology and a pedagogical approach (Attwell 2007; van Harmelen 2006; Johnson et al. 2011; Johnson and Liber 2008; Schaffert and Kalz 2009) that aim to empower students to be in charge of their own learning by selecting tools and resources to create, organize, and package learning content, thereby meeting their personal needs and goals (McLoughlin and Lee 2010).

Nonetheless, the high hope held for the PLE to be a key enabler for lifelong learning is yet to be shown because the research and practice on PLE is still evolving. Specifically, substantive claims about the power of PLE should be grounded in relevant case studies, which, however, are limited in number and scope (Johnson et al. 2011). The paucity of case studies and missing evidence on the success or usefulness of PLEs can be attributed to the lack of a comprehensive evaluation framework for PLEs. The difficulties of evaluating PLEs have been documented (e.g., Gillet et al. 2011; Giovannella 2011). While technical implementations have demonstrated some significant progress (see Chaps. 5 and 8 in this volume), the empirical evaluation of PLEs lacks behind. Indeed, the development of an evaluation framework for PLEs poses several major challenges:

– PLEs are not a stable technology that can be prepared and used in a controlled environment. In fact, PLEs do change over time and can be highly dynamic.
– PLEs integrate other technological artifacts that are designed independently from each other and can stem from different providers. This leads to possible (unintended) interdependencies, usability issues, and update state problems.
– PLEs are used to combine formal and non-formal learning contexts. Therefore the purpose of using a PLE can be highly heterogeneous, rendering systematic comparisons across different learners very difficult.
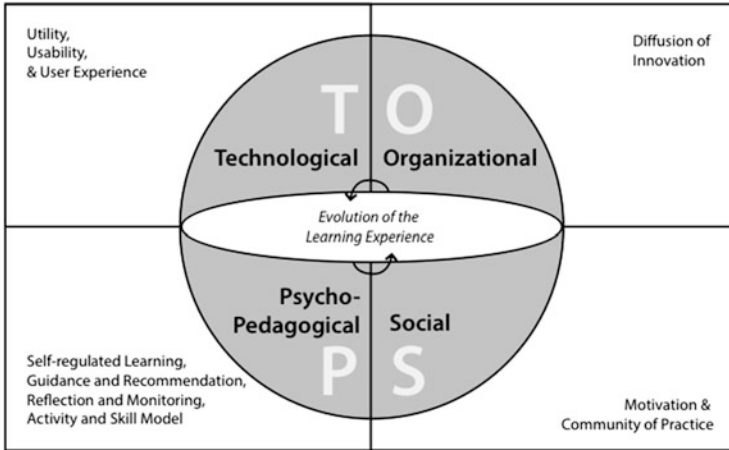
Fig. 1 Four perspectives (the TOPS model) for PLE evaluation

To tackle these challenges, mixed-method and multi-perspective evaluation approaches are deemed relevant to address the complexity of PLE usage and its effects on learning behaviors and learning outcomes.

Four main perspectives can be identified: *technological*, *organizational*, *psycho-pedagogical*, and *social* (short: "TOPS"), with each being informed by specific concepts and theories and subsuming certain methods and tools (see Fig. 1). They are elaborated in the following with reference to related work of the ROLE project (http://www.role-project.eu/).

The "TOPS" Model for Evaluating PLEs

In this section we delineate the individual perspectives of the TOPS model—with specific emphasis on their respective underlying conceptual and theoretical frameworks.

Technological Perspective

The technological perspective comprises two main aspects: *utility* and *usability* and *user experience*. It is to emphasize that the user-centered design (UCD) approaches underpin the work of PLEs, so not only end-users' but also developers' perspectives should be taken into account.

Utility

Two major elements of utility to be evaluated are *software* and *documentation*, which are discussed in detail in the following.

Software evaluation pertains to the functionality of different software components constituting PLEs, including widgets, widget containers, the widget store, libraries, services, tools, and the overall interoperability framework. It is essential to evaluate how useful these components are to enable end-users to accomplish specific tasks and goals.

As already indicated above, the strict separation of end-users from developers can be seen as artificial (at least under the UCD approach), thus requiring an evaluation approach to look also at developers and "power" users who engage in customization, configuration, or even end-user-driven development (keyword: mash-ups). Typically, such developers and power users use configuration options, authoring tools, and APIs allowing for the mash-up of components to customize or even create new software artifacts.

Specifically, we highlight a list of factors critical for the technical evaluation of software, which are adopted from constructs of the Information System Success Model (ISSM) (DeLone and McLean 2003) and the Technology Acceptance Model (TAM) (Davis 1989; Venkatesh et al. 2003; Venkatesh and Bala 2008) (see Table 1).

Table 1 Constructs relevant to the utility evaluation of PLEs

Construct	Factor
System quality	Integration
	Portability
	Availability
	Performance
	Reliability
	Usefulness of features
	Completeness of features
	Security
Information quality	Usefulness of information
	Completeness
	Correctness
	Appropriateness of presentation format
Service quality	Responsiveness
	Support
	Feedback mechanism
	Trustworthiness
Use	Users, communities
	Functionality accessed
	Information items accessed
	Duration of use
	Frequency of use

An increasingly popular data collection method is *automated monitoring*: monitoring and data logging for capturing how frequently a service or a feature has been used and how often different significant events have occurred (Helms et al. 2000). More specifically, raw data of system use are recorded and then aggregated for computing measures for individual factors. For example, "mean time to failure" is a measure for the factor "reliability" under the construct of "system quality" that can be derived from monitoring data, including information about what and when errors occur.

Contextual information gathering (i.e., information about the current situation where a learner deploys specific software) is also important. Noteworthy is that context has a technical and social aspect: which software/browser is used for accessing a PLE, which types of data are accessed, and which people interact with each other in a certain community. In principle, physical context information can also be recorded if the sensor data required would be available (e.g., GPS coordinates for spatial location). An example of including contextual data in evaluation is a factor "Browser Compatibility," where the number of errors occurred related to a particular browser can be measured. Similarly, a factor "Widget Container Interoperability" can be measured by associating errors with a particular widget container where they occur.

Where it is possible and does not infringe privacy and security regulations, it may often be safer to capture a broad standard range of monitoring data, especially as capturing technologies typically require no further human effort beyond initial setup and the setup is often already integrated into the related software. Subsequently, such data can be selected, refined, and processed based on the actual needs and goals of an evaluation project.

Croll and Power (2009) provide an elaborate list of metrics that can be used for monitoring usage of web-based technologies. Some of the key metrics relevant to PLEs are: user-generated content, content popularity, loyalty, search effectiveness, reflection, enrolment, conversion, and abandonment.

There are a number of ways for collecting data for these metrics. *Google Analytics* are a free service to generate a comprehensive range of usage statistics for any web-based application. Following the insertion of a small JavaScript code snippet into a given web application, Google starts to record usage statistics (including simple demographic features and events). Some of the key aspects that Google can currently track are

– *Visitor Tracking*: Demographics, conversion, uniqueness, loyalty, etc.
– *User Profile*: browser, OS, screen resolution, Java availability, flash availability, connection speed, etc.
– *Events*: frequency of use of specific event categories, events per visit, total number of events.

One of the drawbacks of using analytics is the limited capability to provide data describing how users interact with content and tools (known as attention metadata) within their environments. Collecting *contextualized attention metadata* (CAM) will enable us to infer the ways learners use technologies and tools for specific

purposes. The CAM approach proposed by Wolpers et al. (2007) supports such tracking of attention metadata. This approach helps observe the user at the application level, enabling association of tool usage with content-specific behavior *in context*. The challenge of collecting observation data of user attention unobtrusively can be resolved by the CAM approach through integrating the data-capturing process into a user's daily working environment. This approach allows integrating data from web applications (e.g., by mapping the Apache open log file format to CAM) as well as from desktop applications. CAM helps track learning content usage, analyze behavioral patterns, provide similarity measures between users, and allow inferences about user goals. CAM data can be utilized to measure the effectiveness of PLE technologies in providing the learner with a highly responsive and personalized learning environment. CAM data can also be used to track and infer self-regulatory activities for measuring the effectiveness of the psycho-pedagogical model (Scheffel et al. 2010a, b).

All measures that cannot be derived with automatic monitoring need to be obtained from users explicitly. The challenge is to identify appropriate techniques for survey data acquisition with the possible lowest obtrusiveness and highest intuitiveness for users.

For instance, a lightweight "Requirements Bazaar" approach is integrated into the ROLE Widget Store (http://role-widgetstore.eu/) similar to other well-accepted systems such as Google's Android Market or the Chrome Extensions marketplace. This is a valuable source of data since their users provide feedback on the quality of tools, services, and widgets using means such as rating scales, and—where appropriate—free text comment boxes.

Documentation evaluation looks into the availability and quality of technical documentation—a prerequisite for software to be accepted by end-users as well as developers. To encourage developers to contribute new learning technologies by mashing up existing software components, it is necessary to ensure that documentation is correct, complete, and tailored to developers' needs.

With regard to the development of web-based software components, developer documentation of the infrastructure usually includes the following items:

– The set of initial documents (e.g., an overview of the underlying principles and overarching architecture).
– The reference documentation with complete information on all supported features, usually in the form of API documentation.
– The set of tutorials demonstrating how to use the technology for developments on simple and useful examples.

Specifically, technical documentation should be tested by inviting developers to practical sessions, where they are asked to use the infrastructure and accompanying documentation to realize a small but motivating use case beyond basic tutorial contents. In such sessions, the developers who authored the documentation can serve as tutors to be consulted to discuss any problems arising. Such discussions can be used as individual interviews or focus groups to collect feedback on the quality of the software as well as documentation. This approach, however, does not scale to

large groups of developers. This is where the required alternative means such as online tools are preferred over presence workshops.

Documentation of web-based software is usually supplemented by different technical means for communicating with the core developers of the original technology, authors of the documentation (who often are also its developers), and developers deploying these software artifacts. For instance, developers use online forums to get in contact with other developers to report problems and ask for help. Besides bug reports, such comments often contain practical questions about how to accomplish certain tasks, thus indicating where the existing documentation could be unclear or incomplete.

Further means to assess the utility of documentation is to directly integrate ratings, for instance, in the form of 5-star scales, like/dislike buttons or commenting functions, into the online documentation. In this manner, different factors from the dimension Information Quality can be surveyed. These (and additional) features are often already provided by software project management systems such as SourceForge, GitHub, and the like.

Usability and User Experience

First of all, it is deemed imperative to demarcate usability from user experience (UX)—two key concepts in the field of human–computer interaction (HCI). One main distinction is that usability targets *instrumental quality*, emphasizing the effectiveness and efficiency of task and goal attainment with interactive technologies, whereas user experience targets *non-instrumental quality* (e.g., aesthetics), going beyond the traditional task-oriented focuses to address users' affective and emotional responses (e.g., fun, pleasure, surprise, sad, happy) to interactive technologies (e.g., Hassenzahl 2013). Hassenzahl's (2005) oft-cited model on the pragmatic and hedonic quality illustrates similar arguments. Despite its decade-long history, some basic conceptual issues in UX are yet to be resolved (Law et al. 2009; Law, van Schaik & Roto, 2014). While a deeper exploration of such issues is beyond the scope of this chapter, here we highlight metrics and approaches relevant to the evaluation of PLEs.

Noteworthy is that usability and user experience evaluations focus on the interaction design of technological components underpinning PLEs, which nonetheless contribute to the holistic educative experience with PLEs (see also section "Psycho-pedagogical Aspect").

Usability

The usability of different technological components of PLEs (section "Utility") is to be evaluated based on a combination of metrics identified from the literature

(e.g., Nielsen 1994) and standards (ISO/IEC 25010:2011[1]; ISO/IEC 9241-110[2]: 2006 and ISO/IEC 9241-210: 2010[3]). The metrics are listed as follows:

– *Learnability*: The ability of the technology to enable users to learn with great ease how to assemble a PLE themselves. If users find it difficult to assemble a PLE, then the acceptance and uptake may be drastically hindered. Hence, the assembly process for such an open learning environment should be relatively straightforward for end-users. Some factors that enable us to ascertain learnability are consistency of user interface design and predictable system behavior. Learnability of PLEs is equally important for developers as for end-users. If developers find it difficult to use PLE software, they may not be able to create new widgets.
– *Efficiency*: The ability of the technology to support users to be highly productive. Features such as consistent look and feel, consistent navigation, frequent feedback, and availability of templates to help them quickly assemble their environments can contribute to the overall efficiency of the PLE software.
– *Memorabilit*y: The ability of the technology not to require users to reinvest time in remembering how to use it after a period of nonuse. Closely related with learnability, memorability can influence the uptake and usage of PLE. The key success factor for PLE is to make the assembly process of the environment highly intuitive, using relevant standardized visual cues.
– *Error Tolerance*: The ability of the technology to avoid catastrophic errors by making users reconfirm critical actions (e.g., deleting a software component) and to recover from errors by providing the "un-do" feature that allows users to reverse their actions.
– *Effectiveness*: The ability of the technology to help users achieve their goals. Using PLEs, if learners are able to assemble and personalize their environments with ease, while at the same time they find the recommendations and rated/ranked content useful for fulfilling their goal, then we can infer that the technology is effective and that learners are likely to feel satisfied. More explicit methods are mentioned above in section "Utility."
– *Flexibility*: The ability of the technology to offer a range of services so as to be able to adapt to task changes. The ability of learners to seamlessly integrate and use a range of web-based tools and services for assembling their learning environments and for exporting/importing data as well as settings to other similar technologies.
– *Operability*: The ability of the platform to allow users to operate and control it.
– *Satisfaction*: The ability of the platform to be deployed by users without discomfort. It is highly subjective as compared with the other qualities listed above, which when realized to a sufficiently large extent, can contribute to overall user

---

[1] Systems and software engineering: Systems and software Quality Requirements and Evaluation.

[2] Ergonomics of human-system interaction: Part 110: Dialogue principles.

[3] Ergonomics of human-system interaction: Part 210: Human-centered design for interactive systems.

satisfaction. Note that in addition to the system and service qualities, information quality can play a key part in user satisfaction, according to the ISSM (DeLone and McLean 2003).

Usability evaluation methods comprise a range of *usability inspection methods*, *user-based tests*, and *user surveys*, which can be used to evaluate PLEs using the metrics described above. Inspection methods rely on experts, whereas user-based tests and user surveys, as the names suggest, involve end-users (an overview, see Holzinger 2005).

Two commonly used inspection methods are *heuristic evaluation* and *cognitive walkthrough*. For heuristic evaluations, experts examine a system based on ten usability heuristics or principles that were originally derived from a large database of common problems. Violating any of such principles is identified as usability problem of which the severity is estimated so as to inform the urgency and necessity of its being fixed (Nielsen 1994). The major advantages of this method are that it can be applied throughout the whole development lifecycle and is, relatively, less time-consuming. In a cognitive walkthrough, experts analyze a system's functionality with a set of four questions (e.g., "Will the user notice that the correct action is available?") to estimate how the user would interact with the system (Lewis and Wharton 1997). A negative response to any of the questions suggests the identification of a usability problem.

All inspection methods, as prediction methods, are prone to false alarms and results thereof are typically to be verified with user-based tests, such as *think aloud* or *field design methods* and *observation methods* (e.g., video observation, screen sharing, mouse tracking, eye tracking). Usability evaluation feedback is deployed for further development of the system under scrutiny, as they can provide insights into where and why usability requirements are not met.

Think aloud is a method that requires end-users to constantly think aloud as they are using a system individually or collaboratively in order to understand how they perceive the features of the user interface, identify preferences, and discover any potential misconceptions at early design stages (Dumas and Fox 2007). The drawback of this method is that it can be tiring for end-users who have to focus and behave in a rather unnatural manner by giving a running commentary on their own actions.

Field methods are a collection of tools and techniques for conducting user studies in context. Among others, Contextual Inquiry (Beyer and Holtzblatt 1998) is commonly used field method in research as well as in practice. The main advantage of such methods is that they provide a development team with data about what and how (and why) people carry out their tasks in a given environment, thereby enabling the production of useful and usable systems that meet people's needs and goals. The main disadvantage is that they are time-consuming. Nonetheless, such methods can be streamlined with respect to the budget available for evaluation in a project (Wixon et al. 2002).

Furthermore, while the importance of automated monitoring techniques was already highlighted above, methods such as CAM and Google Analytics may not

provide sufficient granularity of data to determine the usability of the PLE software. The ability of CAM to provide granular and contextual data may be useful, but its appropriateness may not be established unless or until a sufficient amount of data has been collected. Apart from traditional methods mentioned above, there are two additional methods that can be useful for small-scale (eye tracking) and large-scale (mouse tracking) usability evaluations:

– *Eye tracking* measures visual attention as people navigate through websites. It is useful in quantifying which sections of an interface are read, glanced at, or skipped/ignored. Eye tracking is generally carried out in laboratories and at a small scale. It can provide useful information for evaluating the effectiveness of the learning design (Schwonke et al. 2009; van Gog and Scheiter 2010) and it can be used to gather data after every redesign phase before large-scale rollout.
– *Mouse tracking* is a technique for monitoring and visualizing mouse movements on any web interface. Mouse movements provide key data about usability issues on a large scale, as users can be observed in their natural habitat in an unobtrusive and continuous manner. In most cases, a JavaScript code snippet is inserted to track mouse movements. Privacy issues must be considered while adopting this method. Tools like Crazyegg,[4] Userfly,[5] and Simple Mouse Tracking[6] can be used for this purpose. It should be mentioned that even more so than eye tracking, data captured with this method represent only part of the story and, hence, must be triangulated with other qualitative data to ensure completeness and correct interpretation.

For summative usability evaluation, *user surveys* are deployed. They are normally administered in the final phase of a project after end-users interact with an executable prototype. Among others, the System Usability Scale (SUS) is widely used in research and practice, as it is simple with only ten items and standardized with psychometric properties (Brooke 1996).

To study the usage of PLEs, it is crucial to evaluate whether the associated services and features can help achieve learning objectives. This can be derived from *evaluation metadata* such as ratings, bookmarks, tags, and comments provided by users (Vuorikari and Berendt 2009): One important aspect here is to investigate how the PLE usage facilitates social interactions, triggers discussions, and improves the understanding of learning content (Mason and Rennie 2007; Farrell et al. 2007; Rollett et al. 2007). Moreover, when it comes to learning material recommended by the system, ratings and like/dislike evaluation metadata can help assess unobtrusively to what extent learners deem them useful.

---

[4] http://www.crazyegg.com/

[5] http://userfly.com/

[6] http://smt.speedzinemedia.com/smt/

User Experience

The literature on UX published since the turn of the millennium indicates that there are two disparate stances on how UX should be studied (i.e., qualitative versus quantitative) and that they are not necessarily compatible or can even be antagonistic. A major argument between the two positions is the legitimacy of breaking down experiential qualities into components, rendering them to be measurable. A rather comprehensive review on the recent UX publications (Bargas-Avila and Hornbæk 2011) identifies the following observations: UX research studies have hitherto relied primarily on qualitative methods; among others, emotions, enjoyment, and aesthetics are the most frequently measured dimensions; the products and use contexts studied are shifted from work to leisure and from controlled tasks to consumer products and art; the progress on UX measures has thus been slow.

Given that UX has at least to some extent developed from usability, it is not surprising that UX methods and measures are largely drawn from usability (Tullis and Albert 2008). However, the notion of UX is much more complex, given a mesh of psychological, social, and physiological concepts it can be associated with. Among others, a major concept is emotion or felt experience (McCarthy and Wright 2004). As emotion arises from our conscious cognitive interpretations of perceptual-sensory responses, UX can thus be seen as a cognitive process that can be modeled and measured (Hartmann et al. 2008).

Larsen and Fredrickson (1999) discussed measurement issues in emotion research with reference to the influential work of Ekman, Russell, Scherer, and other scholars in this area. More recent work along this direction has been conducted (cited in Bargas-Avila et al. 2011). These publications point to a common observation that measuring emotion is plausible, useful, and necessary. However, like most, if not all, psychological measurements, they are only approximations (Hand 2004) and should be considered critically. Employing quantitative measures to the exclusion of qualitative accounts of user experiences, or vice versa, is too restrictive and may even lead to wrong implications (Law et al. 2014).

There exist a range of UX evaluation methods (e.g., Vermeeren et al. 2010). For qualitative data, narrative or storytelling methods (e.g., Riessman 2008) are commonly employed. For instance, users' short descriptions about their positive and negative interaction experiences can be analyzed with the use of machine learning as well as manual coding approach (e.g., Tuch et al. 2013). For quantitative data, validated scales with good psychometric properties such as AttrakDiff2 (Hassenzahl and Monk 2010) and PANAS (Positive Affect and Negative Affect Scale; Watson et al. 1988) are increasingly used.

Especially challenging is to operationalize a diversity of emotions, be they positive and negative, because teasing out their nuances proves difficult. Common methods here are self-assessment manikins and Emocards (for a summary, see Stickel et al. 2011). It is even more demanding to measure the social aspect of UX, which has hitherto been defined as highly individual and contextualized (Law et al. 2009).

Organizational Aspect

With their capability for personalization and plasticity, PLEs help create a rich and diverse learning technology ecosystem promising perpetual change and innovation. The uptake and effects of PLEs at an organizational level can be understood in the light of theory of Diffusion of Innovation, which is advanced by Rogers (1995): "An innovation is an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (p.11).

Furthermore, Rogers (1995) states that the "innovation diffusion process" progresses over time through five stages: *knowledge* (when adopters learn about the innovation), *persuasion* (when they are persuaded of the value of the innovation), *decision* (when they decide to adopt it), *implementation* (when the innovation is put into operation), and *confirmation* (when the decision is reaffirmed or rejected).

The ROLE project conducted a study to identify factors that can have an effect on the adoption and diffusion of PLE-related technologies in organizations (Chatterjee et al. 2013). Table 2 presents an overview of the factors identified.

Among the main organizational factors, the outlook of the top management on introducing technological change matters, as this particularly influences persuasion strategies for facilitating positive decision-making in terms of PLE adoption. It is equally important to look at how coherent or unified the views on PLEs of the key stakeholders within the organization are. With the increasing popularity of social media within commercial organizations, extensive use of such platforms can have positive impacts on informing the stakeholders about key concepts and issues around PLEs.

The top management, as per the findings of the study, is particularly interested in the cost-effectiveness PLEs offer as compared to existing solutions in place—the perceived cost-effectiveness thus plays a key role here for evaluation. Compatibility with the existing technical infrastructure and high learnability are other key success factors of introducing innovation. These persuasive factors tend to act in a push–pull mechanism (Shih 2006) before embarking on the decision-making stage.

Table 2 Potential factors influencing organizational uptake

Categories	Factors
Organizational	Leaderships attitude towards change
	Strategic alignment
	Learning culture
	IT support
Innovation (PLE)	Perceived cost-effectiveness
	Compatibility with existing system
	Perceived effort expectancy
External factors	Perceived factuality
Communication channels and influence	Line manager
	Social networks

Once the key stakeholders within an organization are informed and persuaded about the usefulness and utility of PLEs within their organization, the top management may then take the two key factors into account when deciding upon the adoption of the new learning technologies.

PLEs enable the learners to take control of their own learning depending on their contextual needs and goals. It is therefore crucial to check whether a framework exists that allows relating personal goals directly to organizational goals. Similarly, the learning culture should not be dominated by didactic and trainer-facilitated approaches, as a healthy sign of PLE adoption is that learners take control of their own learning and managing the related technologies. It is necessary to look at the provision of IT support (particularly in the introduction phase), when stakeholders start using PLEs within their day-to-day activities. Another important factor that determines the PLE adoption is its use by line managers. If line managers and senior team do not lead by example, then the likelihood of PLE adoption can be adversely affected.

Psycho-pedagogical Aspect

From the psychological and pedagogical perspective, the key aspects to look at are the ability to foster self-regulated learning, the guidance and recommendation strategy, and the facilities for reflection and monitoring. Moreover, the availability and documentation of an activity and skill model play an important role—and how far this is put into practice.

Self-regulated Learning

From the psycho-pedagogical perspective, effective exploitation of PLEs, which support lifelong learning, hinges crucially upon the learner's self-regulated learning competence. The quality of learning outcomes varies with the extent to which learners are capable of regulating their own learning (Steffens 2006). Self-regulated learning approaches have been evolving since the 1970s in educational research and practice (Efklides 2009).

Successful deployment of PLEs relies on a self-regulated learning process model such as the following one (derived from Zimmerman 2002), where it is seen as a learner-centric cyclic model consisting of four recurring learning phases: learner profile information is defined or revised; learner finds and selects learning resources; learner works on selected learning resources; and learner reflects and reacts on strategies, achievements, and usefulness.

Note that while cognitive learning activities are rather related to actual learning (i.e., information receiving, debating, and experimenting), meta-cognitive learning activities are related to controlling and reflecting on one's own learning.

With respect to the evaluation of the success and extent of self-regulated learning, gathering data about the accuracy and usefulness of the learning process model is crucial. It is particularly relevant to find out, whether learners can actually follow the process model and whether they comprehend it and its implications. Another key question is, whether the process model supports the development of self-regulatory skills.

It should be taken into account that the process model can be applied in different contexts and situations. For example, learners might be in a collaborative learning situation, where they may learn together with peers. Or they may learn on their own. In addition, the actual learning technology mix may make a difference, since learners might use tools and widgets explicitly built to support self-regulated learning, whereas in other cases, performance of meta-cognitive learning activities may happen just in an implicit way (i.e., being aware of them).

One particularly useful instrument to help in the evaluation of self-directed learning is the questionnaire. While it certainly is supportive of all other aspects mentioned above and following below, this widely used instrument can help here in providing structured, often numerical data. Questionnaires can be administered without the presence of the researcher, and are often comparatively straightforward to analyze (Wilson and McLean 1994). According to Cohen et al (2000), "Though there is a large range of questionnaires that one can use, but there is a simple rule of thumb to follow: the larger the size of the sample, the more structured, closed and numerical the questionnaire may have to be, and the smaller the size of the sample, the less structured, more open and word based the questionnaire may be" (p. 247). Questionnaires are particularly useful when comparison across groups is required (Oppenheim 1992).

Guidance and Recommendation Strategies

Guidance for learning in the context of PLEs depends on the situation and on who is providing the guidance. Learners can learn in a blended learning situation with teachers structuring the learning process. Peers can be involved in the learning process, if learners collaborate in some way. Learners can also learn on their own without human interaction. In the first case, teachers can provide guidance. In the second case, peers can provide guidance either directly or indirectly (e.g., with peers attempting to master a problem together). In all cases, guidance can also be provided by the system through personalized recommendations.

Moreover, the scope of guidance can focus on a variety of things, including the search for learning resources (e.g., widgets, content, or peers), the composition of a PLE, the control over the learning process, and the improvement of self-regulation ability. Evaluating the effectiveness and appropriateness of such guidance strategies requires looking into its preconditions: the given abilities of learners are relevant, since it depends largely on concrete skills of learners, what they can do on their own and where they need help.

Furthermore, goals and preferences need to be investigated because the scope of guidance depends on these factors. It should be noted that it depends on who is delivering guidance, whether certain preconditions can be taken into account, and to which extent. If the system provides guidance, then this is done usually in terms of recommendations. Personalized recommendations are based on a learner model (e.g., goals, skills, learning history, learning progress, background of a learner, and the learner's preferred instructional technique), which models the preconditions for guidance.

The scope of recommendations can include concrete widgets, content resources, peers, learning activities, and complete learning environments (i.e., sets of learning resources). By recommending certain meta-cognitive learning activities, guidance for self-regulated learning can be provided. In case of teacher guidance, learning environments can be pre-configured. Especially in a blended learning situation, teachers can support the use of the learning environment and help improve self-regulated learning, providing further scaffolds to system guidance.

Regarding evaluation, it is important to assess the appropriateness and quality of guidance strategies. This includes evaluating, whether the respective guidance strategy helps learning effectively and whether the guidance provided helps overcome difficulties. Different guidance strategies have different purposes: it requires an evaluation of whether all purposes are actually achieved.

While of course the questionnaire (see above) can be utilized to evaluate the success of particular guidance and recommendation facilities in their context, other qualitative methods are suitable as well—such as focus groups, the nominal group technique, and a Delphi study. Quasi-experiments using test collections and statistical measurements are the dominant quantitative methods.

A focus group is a small group of people who get together to discuss a certain issue given to them normally by a researcher. It usually consists of 6–10 members and meets regularly during the lifetime of a project or in an ad hoc manner when a need arises (Vaughan et al 1996). The technique relies on interactions among group members. Focus groups are used to capture qualitative feedback to triangulate findings from some other data sources.

Two other techniques, namely Nominal Group technique and Delphi technique may be used to collect group opinion. The Nominal Group Technique was developed by Delbecq and Van de Van (1971, 1975) in the 1970s. It has been found to be useful in improving educational programs (Jones and Hunter 1995). There is further evidence in the literature that it was successfully used for evaluation purposes in higher education (Nisbet and Watt 1984). Grant et al. (2003) used the technique to determine the impact of student journals in postgraduate education.

The Delphi technique (Turoff 1970) is, like the Nominal Group technique, a structured process, but it does not require physical proximity among participants. The participants may be geographically dispersed and are not required to meet face to face. Either technique may be instantiated after validation trials to gather group data, augmenting and triangulating the monitoring or survey data.

Following the tradition of search engine evaluation, the relevance of recommendations can be evaluated in the so-called quasi-experiment with the help of a

specially prepared test collection. In such a case, the learning resources (e.g., content, peers) are evaluated by experts or representative users; this allows comparing how well the recommender system performs in bringing up the most relevant and most complete recommendation. Evaluation measures depend on the guidance strategy: for example, recommendations fostering serendipity have much more relaxed requirements on accuracy as compared to identifying potential peers who are currently in a similar learning situation. An overview on possible evaluation measures (and their application contexts) can be found in Herlocker et al. (2004).

Reflection and Monitoring

Learner information is important for guidance strategies; this can be the assessment of a teacher, peers, or the learner herself. A teacher and peers might form an opinion by observing, the learner can do this by self-monitoring or self-reflection, and the system can do that by tracking the learner's behavior and building a learner profile (or recommending profile information). Most importantly, a mixed procedure can be used if profile information is proposed by the system and the learner has to modify and update it. In this case the learner is made aware of certain assessment outcomes, which also stimulates self-reflection. As already mentioned above, learner profile may contain information about goals, skills, learning progress, etc. Evaluation should focus on the accuracy of this information.

While an interview can be used for the evaluation of many of the other aspects listed above and below, it is particularly useful for the evaluation of reflection and monitoring. An interview is a purposeful discussion between two or more people (Kahn and Cannel 1957). One of the most distinct advantages of interview over, for instance, questionnaires is that the researcher has personal contact with the respondent and hence more control over the questions and its context. The researcher is available to clarify confusing questions (Cohen et al 2000), which is difficult to do with questionnaires. This same advantage, however, can also turn into a disadvantage, when the researcher knowingly or unknowingly diverts the discussion and when allowing personal bias to directly impact on outcomes. Interviews consist of a more direct method that helps easily spot user preferences, satisfaction, and encountered problems.

Apart from qualitative approaches, quantitative evaluation techniques utilizing content analysis over learners' writings are emerging, some of which using automation techniques from text mining and statistical processing. Ullmann et al. (2013) provide an overview and a framework for the study of reflection by hand and with the help of automation techniques; from natural language processing as well as using crowd-sourcing of human coding on platforms such as CrowdFlower or Amazon's Mechanical Turk.

Activity and Skill Model

For successful deployment of PLEs, the underlying skill model is typically complex, since in addition to the developed domain knowledge, self-regulated learning and the handling of PLE services and tools have to be considered. Any PLE skill model encompasses at least these three different kinds of skills: domain, tool, and self-regulation skills:

– Domain skills are skills that a learner possesses, if he or she has a certain level of expertise in a knowledge domain. For instance, the learner can explain what percentages she estimates to have attained and, if she prefers, justifies with some qualitative comments.
– Tool skills are defined as skills which a learner possesses, if she is able to perform a learning activity with a learning tool in a domain context: for example, the learner can use a tool for setting goals or can use a tool in order to retrieve domain knowledge in a certain topic. Different learning activities with the same tool can require different skills.
– Self-regulated learning skills imply the ability of a learner to regulate her learning activities by herself: the learner can realistically set own goals, monitor own progress, apply effective time management, and self-evaluate. Self-regulated learning skills are skills on a meta-level and domain independent.

For the evaluation, focus should be set on documenting and subsequently assessing accuracy and usefulness of these skill models. Methods for the assessment of accuracy and usefulness are essentially the same as those valid for evaluating the utility of PLE utility (particularly automated monitoring and CAM).

Social Aspect

A Community of Practice approach is an effective way of sharing knowledge. They are usually characterized by anonymity and an addictive, but voluntary behavior, with a strong sense of belonging (Hampton and Wellman 2001). Trust, loyalty, and social usefulness are pertinent motivational features identified in the virtual community context.

Over the last century, a number of motivational theories were proposed (e.g., Maslow 1954; Herzberg 1987; Vroom 1964). At the foundation of these theories, it is claimed, lies the suggestion that each school of thought focuses on certain factors to the exclusion of all others—for example, reward, social needs, or psychological growth.

A few key inferences in the context of PLEs from the motivational models are mentioned below:

– Recognition of a range of individual needs: Learners have varying levels of motivation depending on their needs.

– Goal alignment in the provision of materials: If a given task does not align with the learner's goal, then the motivation to complete the task will obviously decrease.
– Varying incentives: Incentives can help instill a sense of achievement and motivation to keep going. Learners will require varying levels of incentives of different natures to keep themselves motivated (grades, peer recognition, altruism, to mention just a few).
– Connectedness to community performance: Link of these incentives to performance at an organizational or community level.

To assess the social aspect of PLEs, Kim's (2000) application of Maslow's Hierarchy of Needs to online communities can be further adapted: Table 3

Table 3 Community building and motivation (extended from Kim 2000)

Need	Offline (Maslow)	Online communities (Kim 2000)	Personal Learning Environment
Physiological	Food	System access	Access to PLE technology, widget store, user profile
	Clothing	Ability to maintain own identity while participating in the community	Use of templates for assembly of environment
	Shelter		
	health		
Security and safety	Protection from crimes and war	Protection from hacking and personal attacks	Data security (automated monitoring data) and encryption
	Sense of living in a fair society	level playing field	multi-level privacy framework
		maintain varying level of privacy	
Social	Give and receive love	Belonging to the community as whole and within subgroups	Share and consume tools, content, and resources
	Feeling of belongingness		Belongingness
			Ability to collaborate across several social networks
Self-esteem	Self-respect	Contribute to community and get recognized	Sharing modified PLE templates
	Ability to earn respect from others		Altruism
			Mentoring
			Giving and receiving feedback
			Rating and ranking
Self-actualization	Develop skills and fulfill one's potential	Take on community role that develops new skills and opens new opportunities	Acquiring expert status within the community
			Assembly and regulation of own learning

illustrates which constructs are relevant to the PLE evaluation from a motivational perspective.

Clustering techniques and social network analysis (SNA) can be used to trace whether the infrastructure supports the emergence and evolution of self-directed communities of interest and practice (Wenger 1998). Both rely on either implicit factors (looking at interaction and usage patterns) or explicit ones (utilizing evaluation metadata).

SNA originates from sociology and network analysis that is widely applied in physics, electrical science, civil engineering, and others. In SNA, entities and relations among them are mathematically modeled as graphs, (i.e., sets of nodes and edges connecting them). Nodes and edges can have different semantics: for instance, nodes can be people and edges between nodes can be based on communication between people, for example, through e-mails or chats. Edges can also be used to denote citations of resources that peers own or create. For instance, a peer is connected with the other one whose work he has cited. According to the Actor Network Theory (Latour 1991), we can consider every node as an arbitrary actor, which is not necessarily human. In this sense, it is also possible to analyze networks consisting of users and tools, both modeled as nodes.

SNA is a basis for assessing social learning and the interaction with tools used in learning (Klamma 2010). It helps discover information about social relationships. Based on this, it allows inspecting social presence of learners within their communities: for example, it helps in evaluating which roles learners adopt or how their positions evolve over time, positively as well as negatively.

Since 1967 with the discovery of the small world network phenomenon (Milgram 1967), the heterogeneity of networks has been examined intensively. Newman (2003) showed that in scale-free networks, connections between nodes are distributed unequally with a certain probability. While most of the nodes have few connections, there exist a few nodes exhibiting a large number of connections. The connectivity of a graph representing a network informs about robustness and cohesiveness of the network (Brandes and Erlebach 2005). Freeman (1979) also pays attention to centrality measures that help us to reveal special roles of network nodes. Moreover, brokerage phenomena can hardly be defined without the application of SNA (Barabási 2007). Considering the irregularity of peer connections of networks, Newman and Girvan (2004) developed one of the clustering algorithms, which find groups of network nodes that are densely connected to each other but sparsely connected with other nodes.

Networks typically consist of several groups of learners communicating with each other and with other groups. SNA techniques and clustering allow unveiling the structure underlying such a network. For example, networks can include groups of learners that have connections only to leaders of groups, but don't have communications with other groups.

SNA techniques allow following behaviors of learners within a time frame by examining network centrality measures, which reveal expertise or presence of a learner within a network. This method of evaluation may show us how learners evolve in their communities over time: do they become experts or brokers of information from one to the other community or do they lose their position and lock themselves in a community closed from communication?

In practice, SNA requires the availability of data containing information on the nodes, i.e., people, groups of people or even tools, and on the edges, i.e., relations between nodes. One possible source of input for SNA can be the raw monitoring data. Here, different kinds of interaction between users are captured.

The Unified PLE Evaluation Framework

Based on the TOPS model and the background literature reviewed above, we propose an integrated evaluation framework for PLEs. Specifically, the framework incorporates major dimensions with a gradual progression from the individual to community focus. Figure 2 lists the key dimensions (and its aspects) of this evaluation framework and shows how they relate to each other: the framework is organized in three circles from the inner Technological one, which lays the cornerstone of PLEs, through the middle Psycho-pedagogical circle, which addresses individual user's needs and goals, to the outer Organizational and Social circle, which brings in the social and organizational factors relevant to the exploitation of PLEs.

The constructs highlighted within the three circles are high-level concepts, which should be translated into low-level variables, selected from the review brought forward in the previous sections. Operationalizing and estimating such variables with particular techniques and tools leads to results, which can somehow and somewhat account for the extent to which PLEs successfully enable users to attain their learning goals. For instance, the construct *usability* is translated into two

Fig. 2 The "TOPS" integrated evaluation framework for PLEs

metrics—*effectiveness* and *efficiency*— which can be measured in terms of number and type of errors and in the time to complete a specific task with a PLE.

Nonetheless, not every construct can be operationalized in a straightforward manner. Indeed, it is a challenging task to develop structural and measurement models, where factors and measures are orthogonal in the ideal case, but at least exhibit a lowest degree of collinearity. Statistical analysis techniques such as *correlation*, *regression*, and *factor analysis* deem useful to sample, validate, and tune the underlying model in early evaluation runs in order to maximize validity throughout the overall process.

Table 4 relates these three sets of dimensions (with their main criteria) to the methods reviewed in the previous sections. Each of the dimensions (technological, psycho-pedagogical, and organizational/social) can be broken down into its main groups of constructs, as listed in the first column. The second column provides the selection of methods that have been used in the past and that we deem most appropriate for their study.

Table 4 Evaluation dimensions and recommended methods

Dimension	Group of constructs	Key methods
Technological	Openness	Questionnaires
	Responsiveness	Interviews (incl. storytelling)
	Security	Desk research (documentation)
	Scalability	Nominal group, Delphi
	Documentation	Inspection methods
	Interoperability	User tests
	Accessibility	Monitoring data (incl. web analytics, CAM)
	Availability, reliability	Observation
	Quality (content and UI)	Unit tests
	Effectiveness	user-based evaluation: behavioral measures, observations, and questionnaires
	Efficiency, satisfaction	
	Enjoyment	
Organizational/ Social	Trust	Questionnaires
	Social usefulness	Interviews
	New connections	Desk research
	Sharing	Monitoring data
	Privacy	Social network analysis
		Clustering
Psycho-pedagogical	Meta-cognitive	Questionnaires
	Motivation	Interviews
	Behavioral	Nominal group, Delphi, focus group
	Recommendations	Monitoring data
		Observation
		Quasi-experiments (relevance accuracy)

The PLE evaluation is ideally conducted in cycles of planning, actual evaluation, and reflection on results. A useful vehicle for this can be found in form of case studies and—concluding the final cycle—a cross-case analysis. Case study is a generic term for the investigation of an individual group or a phenomenon (Bogdan and Biklen 2006). Case studies are often used for exploratory research, but the technique can be varied and adapted to include the multi-method mix proposed above for the unified PLE evaluation framework.

While the techniques used may vary, the distinguishing feature of case study is the assumption that human systems develop a characteristic wholeness or integrity and are not simply a loose collection of traits. This approach enables researchers to investigate a given phenomenon to a much greater depth, bringing out the interdependencies of parts and emerging patterns. Besides, case study has the potential to accommodate the value context of the enquiry, is flexible to accommodate unanticipated events, does not attempt to generalize, and admit the problems of researcher bias in various ways (Nisbet and Watt 1984). Nonetheless, the inability to accommodate re-observation is a major cause of concern.

The final cycle of the cyclic evaluation process depicted above in Fig. 3 can then be concluded with the cross-case analysis. A cross-case analysis is "a qualitative, inductive, multi-case study that seeks to build abstractions across cases" (Merriam 1998, p.195). It is used to identify and compare patterns of similarities and differences across individual cases resulting in meaningful connections. Most importantly it empowers all stakeholders to access new knowledge from a rich holistic point of view (Khan and van Wynsberghe 2008).

There are two well-known techniques to carry out cross-case analysis, namely, variable- and case-oriented approaches (Ragin 2004). There are other techniques as
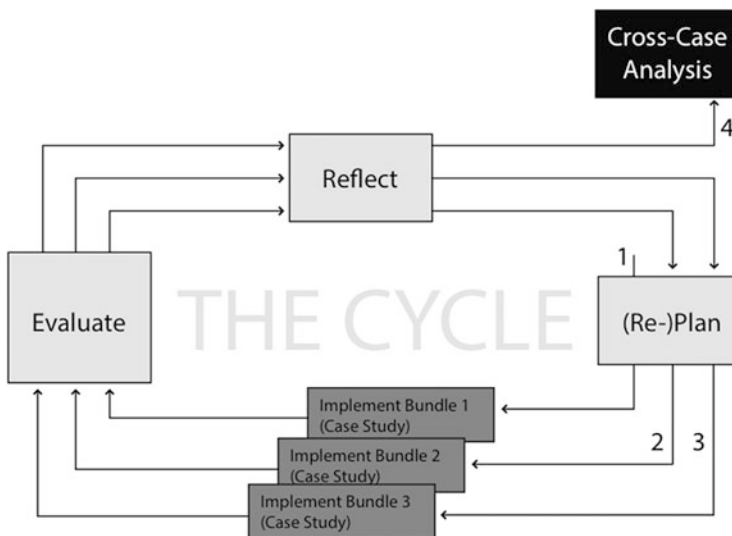

Fig. 3 Evaluation cycle for PLEs

well but are generally derived from the aforementioned ones. The variable-oriented technique focuses on comparison of identified variables across cases in order to delineate causal relationships. The case-oriented approach enables researchers to make sense of causal similarities between different cases by comparing them using visualization techniques such as stacking cases (Miles and Huberman 1994), thereby enabling the identification of new social phenomenon.

There are a number of ways in which case-oriented cross-case analysis could be carried out, namely, most different design (Przeworski and Teune 1982), typologies, multi-case methods (Smith 2004), and process tracing (George and Bennett 2005). The first two are of particular interest for PLE. The aim for adopting cross-case analysis for studying the implementation of PLEs across settings is to identify similarities in a diverse set of cases, which is what most different design offers. Additionally clustering of cases might also be relevant to identify and compare patterns and process pathways to seek typological regularity. We recommend the adoption of an iterative case study design with multi-method data collection to triangulate empirical findings. Cross-case analysis should be performed towards the end of a series of evaluations to obtain a holistic view on the outcomes of deploying PLEs (cf. Fig. 3).

General Discussion: Qualitative Versus Quantitative

In the foregoing sections we present an array of quantitative and qualitative methods for data collection and analysis. The selection of a particular type of method depends on individual researchers' assumptions, values, and expertise.

Some researchers defy the value of quantitative data with the argument that numbers cannot tell us anything, insisting on capturing solely qualitative data. Any method fundamentalism is wrong, not least in the light of a postulate for a wide repertoire of research skills among researchers. Still such standpoint is often found in practice, particularly by those critics instigating methodological discussions with the aim to dismantle or even discredit a particular piece of quantitative work they do not agree with.

It is in our opinion, however, not that simple: Methods cannot be differentiated into good and bad, and if a particular method fails to provide results (or even more often: results beyond tautologies), then this probably says more about their competent handling, rather than their validity or reliability. Exceptions prove the rule, of course.

In our view, there are two aspects to consider that influence methodological choices. First, it all depends on *why* the evaluation is needed, *what* the goal of the evaluation is, and *who* the recipient of the evaluation data is. For example, if the target is to feed back into psycho-pedagogical or technological development, qualitative means can provide deeper insights on what has gone wrong, what works, and what leaves room for improvement. Moreover, qualitative methods bear the potential to discover, *why* this is the case.

Furthermore, which approach to adopt depends on the *phase* of a research study. Qualitative approaches are particularly useful for exploring a topic and its phenomena in their context. They help in forming hypotheses and build understanding. Once such understanding is reached, however, more targeted questions can be posed. Also, if a phenomenon or an application is potentially relevant to a larger number of people, then it is well justified to conduct a quantitative follow-up to see if the qualitative findings, suspected dependencies, effects, and other observations hold when scaling out. Qualitative methods do not scale very well, which can pose a problem when the target is to, for instance, to assess the effects of an intervention on a full university, an entire company, or the general population.

This chapter aims to support researchers in determining which method they need, depending on purpose ("TOPS") and phase (from case-to-case to cross-case). It provides a rich repertoire of different methods for the multi-method, multi-perspective mix, and it helps in combining the strength of different approaches into a unified evaluation.

As can be seen from the review of the methodological state of the art, the frontiers in technology-enhanced learning are much more complex than the mere differentiation of quantitative and qualitative suggests: "mediated" observation using monitoring data, pictogram-based methods for affect measurement, quasi-experiments for relevance evaluation, and the like start blurring these boundaries and start claiming their own place in the standard canon of methods.

It is worth mentioning one class of methods listed in the chapter in particular, as it stands out through the paucity of research in the area of PLEs: While emotions and affects can play a critical role in influencing a learner's motivation to engage in technology-enhanced learning activities, this experiential aspect tends to be not only overlooked, but also under-researched.

At the turn of millennium, the psychological research on emotions has been rekindled, thanks to the work of psychologists such as Klaus Scherer (2005; "emotion wheel") and James A. Russell (2003 "core affect"). Coincidentally, this resurgence of interest in emotions and affects has resonated with the shift of emphasis in HCI around the same time, moving from cognitivist-behavioral performance-based usability to phenomenological-reflective experience-oriented user experience (UX) (Law et al. 2009) .

Alongside with this change of emphasis is the revived tension about the relative importance of qualitative and quantitative methods. This issue is actually an age-old debate in the realm of measurement theory. In brevity, some UX researchers argue that experience is holistic and cannot be reduced into components to be measured; any attempt to put down a number to infer the type or intensity of an emotion is methodologically flawed and inherently meaningless. In contrast, some other UX researchers believe that the process of experiencing/experienced emotions can be modeled like cognitive processes and thus they are measurable. These arguments have significant implications to the selection of evaluation methods for assessing the impact of interacting with technologies (Law et al. 2014).

Above all, putting aside the issue about the quantifiability of user experience, the main point we want to stress is the high relevance of emotions and affects to the

design and evaluation of learning environments. Both positive (e.g., fun, pleasure, engaged, liberating) and negative (e.g., anxious, defeated, frustrated, fear) emotions can substantially shape the effectiveness of any type of learning situations, including PLEs. Consequently, due attention should be heeded to this overlooked experiential aspect.

Conclusion and Future Work

Developing an evaluation framework for PLEs is challenging, since technological, organizational, psycho-pedagogical and social aspects need to be considered in an integrated manner and with a diverse set of stakeholder perspectives being taken into account.

Our attempt was to propose a unified framework encompassing the main valid constructs (derived from relevant theoretical models), yet at the same time providing a flexible and adaptive methodology that is capable of accommodating the changes that are inevitable in an emerging field.

In order to achieve this, we have elaborated an integrated framework that is by nature case study based and follows a multi-method approach. Furthermore, we recommended concluding the cyclic evaluation with a cross-case analysis in order to consolidate data from different contexts so as to establish a holistic view.

A number of metrics and possible methods have been identified and located in the proposed unified framework. The metrics, criteria, methods, techniques, and tools proposed are subjected to further refinement and improvement. A process model ensures the possibility to do so in a well-defined manner.

Obviously, more research efforts are called for to investigate the complex phenomenon of PLE—and this contribution provides the methodological basis on which such future endeavors can be built.

References

Attwell G. Personal learning environments—the future of eLearning? eLearning papers. 2007:2. http://www.elearningpapers.eu/index.php

Barabási A-L. From network structure to human dynamics. IEEE Contr Syst Mag. 2007;27(4):33–42.

Bargas-Avila JA, Kasper Hornbæk K. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In: Proc. CHI 2011; 2011. p. 2689–98.

Beyer H, Holtzblatt K. Contextual design: Defining customer-centered systems. San Francisco: Morgan Kaufmann; 1998.

Bogdan RC, Biklen SK. Qualitative research for education: an introduction to theory and methods. 5th ed. Needham Heights: Allyn & Bacon; 2006.

Brandes U, Erlebach T. Network analysis: methodological foundations. LNCS 3418. Berlin: Springer; 2005.

Brooke J. SUS: A quick and dirty usability scale. In: Jordan PW et al., editors. Usability evaluation in industry. Boca Raton: CRC Press; 1996. p. 189–94.

Chatterjee A, Law EL-C, Mikroyannidis A, Owen G, Valesco K. Personal learning environments in the workplace: an exploratory study into the key business decision factors. International Journal of Virtual and Personal Learning Environments (IJVPLE), 2013;4(4):44–58.

Cohen L, Manion L, Morrison K. Research methods in education. 4th ed. London: Routledge; 2000.

Croll A, Power S. Complete web monitoring. Sebastopol: Oreilly; 2009.

Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart. 1989;13(3):319–40.

Delbecq AL, van de Ven AH. Nominal versus interacting group processes for committee decision-making effectiveness. Acad Manage J. 1971;14:203–12.

Delbecq AL, van de Ven AH, Gustafson DH. Group techniques for program planners. Glenview: Scott Foresman; 1975.

Delone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manage Inform Syst. 2003;19(4):9–30.

Dumas JS, Fox JE. Usability testing: Current practice and future directions. In: Sears A, Jacko JA, editors. The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. 2nd ed. Boca Raton: CRC Press; 2007. p. 1129–50.

Efklides A. The role of metacognitive experiences in the learning process. Psicothema. 2009;21(1):76–82.

Farrell S, Lau T, Nusser S, Wilcox E, Muller M. Socially augmenting employee profiles with people-tagging. In: Proceedings of the 20th annual ACM symposium on user interface software and technology (UIST '07); 2007. p. 91–100.

Freeman LC. Centrality in social networks: conceptual clarification. Soc Netw. 1979;1(3):215–39.

George AL, Bennett A. Case studies and theory development in the social sciences. Cambridge: MIT Press; 2005.

Gillet D, El Helou S, McCluskey A. Evaluating agile PLE enablers. In: Proceedings of the PLE conference 2011, 11–13 July 2011, Southampton; 2011. http://journal.webscience.org/655/

Giovannella C. Virtual learning places: a perspective on future learning environments and experiences. In: Law EL-C, Mödritscher F, Wolpers M, Gillet D, editors. Proceedings of the 1st workshop on exploring fitness and evolvability of personal learning environments; 2011. p. 10–17. http://ceur-ws.org/Vol-773/

Grant A, Berlin A, Freeman GK. The impact of a student-learning journal: a two-stage evaluation using the nominal group technique. Med Teach. 2003;25:659–61.

Hampton K, Wellman B. Long distance community in the network society contact and support beyond Neville. Am Behav Scient. 2001;45(3):476–95.

Hand DJ. Measurement theory and practice. Chichester: Wiley-Blackwell; 2004.

Hartmann J, De Angeli A, Sutcliffe A. Framing the user experience: information biases on website quality judgment. In: Proc. CHI'08; 2008. p. 855–64.

Hassenzahl M. The thing and I: understanding the relationship between user and product. In: Blythe M, Overbeeke C, Monk AF, Wright PC, editors. Funology: from usability to enjoyment. Dordrecht: Kluwer Academic; 2005. p. 31–42.

Hassenzahl M. User experience and experience design. In: Soegaard M, Dam RF, editors. The encyclopedia of human-computer interaction. 2nd ed. Aarhus: The Interaction Design

Foundation; 2013. http://www.interaction-design.org/encyclopedia/user_experience_and_experience_design.html

Hassenzahl M, Monk A. The inference of perceived usability from beauty. Hum Comput Interact. 2010;25(3):235–60.

Helms J, Neale DC, Isenhour PL, Carroll JM. Data logging: higher-level capturing and multi-level abstracting of user activities. In: Proceedings of the 40th annual meeting of the Human Factors and Ergonomics Society. Santa Monica: Human Factors and Ergonomics Society; 2000. p. 303–6.

Herlocker J, Konstan J, Terveen L, Riedl J. Evaluating collaborative filtering recommender systems. ACM Trans Inform Syst. 2004;22(1):5–53.

Herzberg F. One more time—how do you motivate employees? Harv Bus Rev. 1987;65(5):109–20.

Holzinger A. Usability engineering methods (UEMS) for software developers. Commun ACM. 2005;48(1):71–4.

Johnson M, Liber O. The personal learning environment and the human condition: from theory to teaching practice. Interact Learn Environ. 2008;16(1):3–15.

Johnson M, Griffiths D, Wang M. Positioning theory, roles and the design and implementation of learning technology. J Univ Comput Sci. 2011;17(9):1329–46.

Jones J, Hunter D. Consensus methods for medical and health services research. BMJ. 1995;311.

Kahn RL, Cannel C. The dynamics of interviewing. New York: Wiley; 1957.

Khan S, van Wynsberghe R. Cultivating the under-mined: cross-case analysis as knowledge mobilization. Forum Qual Soc Res. 2008;9(1). http://www.qualitative-research.net/index.php/fqs/index

Kim AJ. Community building on the web: secret strategies for successful online communities. London: Addison Wesley; 2000.

Klamma R. Emerging research topics in social learning. In: Dirckinck-Holmfeld L, Hodgson V, Jones C, de Laat M, McConnell D, Ryberg T, editors. Proceedings of the 7th International conference on networked learning 2010, Aalborg, Denmark; 2010. p. 224–31.

Larsen RJ, Fredrickson BL. Measurement issues in emotion research. In: Kahneman D, Diener E, Schwarz N, editors. Well-being. New York: Sage; 1999.

Latour B. Technology is society made durable. In: Law J, editor. A sociology of monsters: essays on power, technology and domination. London: Routledge; 1991. p. 103–31.

Law EL-C, Roto V, Hassenzahl M, Vermeeren A, Kort J. Understanding, scoping and defining user experience: a survey approach. In: Proceedings of CHI 2009; 2009. p. 719–28.

Law EL-C, Van Schaik P, Roto V. Attitudes towards the measurability of user experience. Int J Hum Comput Studies 2014;72(6):526–541.

Lewis C, Wharton C. Cognitive walkthroughs. In: Helander M, Landauer TK, Prabhu P, editors. Handbook of human-computer interaction. 2nd ed. Amsterdam: Elsevier; 1997. p. 717–32.

Maslow AH. Motivation and personality. New York: Harper & Row; 1954.

Mason R, Rennie F. Using Web 2.0 for learning in the community. Internet Higher Educ. 2007;10(3):196–203.

McCarthy J, Wright P. Technology as experience. Cambridge: MIT Press; 2004.

McLoughlin C, Lee MJW. Personalised and self regulated learning in the Web 2.0 era: international exemplars of innovative pedagogy using social software. Australas J Educ Technol. 2010;26(1):28–43.

Merriam SB. Qualitative research and case study applications in education. San Francisco: Jossey-Bass; 1998.

Miles MB, Huberman M. Qualitative data analysis: an expanded sourcebook. Thousand Oaks: Sage; 1994.

Milgram S. The small world problem. Psychol Today. 1967;1:60–7.

Newman MEJ. The structure and function of complex networks. Siam Rev. 2003;45(2):167–256.

Newman ME, Girvan M. Finding and evaluating community structure in networks. Physical Rev E. 2004;69(2):026113.

Nielsen J, editor. Usability engineering. Morgan: Kaufmann; 1994.

Nisbet J, Watt J. Case study. In: Bell J, Bush T, Fox A, Goodey J, Goulding S, editors. Conducting small-scale investigations in educational management. London: Harper & Row; 1984. p. 79–92.

Oppenheim AN. Questionnaire design, interviewing and attitude measurement. London: Pinter; 1992.

Przeworski A, Teune H. The logic of comparative social inquiry. Malabar: Robert E. Krieger; 1982.

Ragin CC. Turning the tables: how case-oriented research challenges. In: Brandy HE, Collier D, editors. Rethinking social inquiry: diverse tools, shared standards. Lanham: Rowman & Littlefield; 2004. p. 123.

Riessman CK. Narrative methods for the human sciences. Los Angeles: Sage; 2008.

Rogers EM. Diffusion of innovations. In: Stacks DW, Salwen MB, editors. An integrated approach to communication theory and research, vol. 65. New York: Free Press; 1995. p. 519.

Rollett H, Lux M, Strohmaier M, Dosinger G, Tochtermann K. The Web 2.0 way of learning with technologies. Int J Learn Technol. 2007;3(1):87–107.

Russell JA. Core affect and the psychological construction of emotion. Psychol Rev. 2003;110 (1):145.

Schaffert S, Kalz M. Persönliche Lernumgebungen: Grundlagen, Möglichkeiten und Herausforderungen eines neuen Konzepts. In: Wilbers K, Hohenstein A (Hrsg.), Handbuch E-Learning (Gruppe 5, Nr. 5.16). Köln: Deutscher Wirtschaftsdienst (Wolters Kluwer Deutschland), 27, Erg.-Lfg; 2009. p. 1–24.

Scheffel M, Beer F, Wolpers M. Analysing contextualized attention metadata for self-regulated learning. In: Proc. 2nd Int. Conference on computer supported education, Valencia, Spain, April; 2010a.

Scheffel M, Beer F, Wolpers M. Analyzing contextualized attention metadata with rough set methodologies to support self-regulated learning. In: Proc. 10th IEEE International conference on advanced learning technologies (ICALT 2010), July, Sousse, Tunesia; 2010b.

Scherer KR. What are emotions? And how can they be measured? Soc Sci Inform. 2005;44 (4):695–729.

Schwonke R, Berthold K, Renkl A. How multiple external representations are used and how they can be made more useful. Appl Cogn Psychol. 2009;23(2009):1227–43.

Shih H. Technology-push and communication-pull forces driving message-based coordination performance. J Strat Inform Syst. 2006;15(2):105–23.

Smith AD. Knowledge management strategies: a multi-case study. J Knowl Manage. 2004;8(3):6–16.

Steffens K. Self-regulated learning in technology-enhanced learning environments: lessons of a European peer review. Eur J Educ. 2006;41(3/4):353–79.

Stickel C, Holzinger A, Felfernig A. Measuring emotions: towards rapid and low cost methodologies. In: Proc. RecSys'11 workshop on human decision making in recommender systems (Decisions@RecSys'11), Chicago; 2011.

Tuch AN, Trusell R, Hornbæk K. Analyzing users' narratives to understand experience with interactive products. In: Proc. CHI 2013; 2013. p. 2079–88.

Tullis T, Albert W. Measuring the user experience. San Francisco: Morgan Kaufman; 2008.

Turoff M. The policy Delphi. J Technol Forecast Soc Change. 1970;2(2).

Ullmann T, Wild F, Scott P. Reflection—quantifying a rare good. In: Proceedings of the 3rd workshop on awareness and reflection in technology-enhanced learning; 2013. http://ceur-ws.org/Vol-1103/

van Gog T, Scheiter K. Eye tracking as a tool to study and enhance multimedia learning. Learn Instruct. 2010;20(2):95–9.

van Harmelen M. Personal learning environments. In: Proceedings of the sixth IEEE international conference on advanced learning technologies; 2006. p. 815–6.

Vaughan S, Schumm JS, Sinagub J. Focus group interviews in education and psychology. Thousand Oaks: Sage; 1996.

Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci. 2008;39:273–315.

Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quart. 2003;27(3):425–78.

Vermeeren APOS, Law EL-C, Roto V, Obrist M, Hoonhout J, Väänänen-Vainio-Mattila K. User experience evaluation methods: current state and development needs. In: Proc NordiCHI; 2010. p. 521–30.

Vroom VH. Work and motivation. New York: Wiley; 1964.

Vuorikari R, Berendt B. Study on contexts in tracking usage and attention metadata in multilingual technology enhanced learning, GI Jahrestagung; 2009. p. 1654–63.

Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol. 1988;54(6).

Wenger E. Communities of practice: learning, meaning, identity. Cambridge: Cambridge University Press; 1998.

Wilson N, McLean S. Questionnaire design: a practical introduction. Newtown Abbey: University of Ulster Press; 1994.

Wixon DR, Ramey J, Holtzblatt K, Beyer H, Hackos J, Rosenbaum S, Page C, Laakso SA, Laakso K. Usability in practice: field methods evolution and revolution. In: Extended abstracts of CHI'2002; 2002. p. 880–4.

Wolpers M, Najjar J, Verbert K, Duval E. Tracking actual usage: the attention metadata approach. Educ Technol Soc. 2007;10(1):106–21.

Zimmerman BJ. Becoming a self-regulated learner: an overview. Theory into Practice. 2002;41(2):64–70.