Evaluation Methods for Learning Experience Design

Schmidt, M. , Earnshaw, Y. , Tawfik, A. A. , & Jahnke, I.

This chapter addresses the methodological vacuum in evaluating LXD practices. It elucidates common evaluation methods for LXD, providing a structured approach amidst the existing challenges in terminology, theoretical foundation, and method-application from user experience design (UXD) in learning design contexts. This chapter aims to bridge the gap by offering methodological guidance, thus fostering a more robust framework for evaluating LXD initiatives.

Author's Note

This chapter is a companion to the chapter entitled *Learning Experience Design*, also available in this volume.

Learning experience design (LXD) is being practiced as a modern manifestation of learning design at an increasing rate (Schmidt & Huang, 2022; Schmidt & Tawfik, 2022). However, given the recency of LXD, a range of challenges present themselves when learning designers desire to apply LXD in their own design practice. Of these, Schmidt and colleagues (2020) identified three major, troublesome issues: (a) there is little agreement in terminology (i.e., what is it?), (b) no substantial efforts have been made to connect LXD practice with the theoretical foundations of learning design and technology (i.e., how does it work?), and (c) there are no guidelines for applying methods and processes derived from user experience design (UXD) in learning design contexts (i.e., how do you do it?). In response to the lack of methodological guidance in this area, the current chapter seeks to introduce evaluation methods that are commonly used for LXD.

Learning experience design has been characterized as encompassing two broad forms of interaction: (a) interaction with the learning space and (b) interaction with the learning environment, which Tawfik and colleagues (2022) describe as follows:

> [Interaction with the learning environment is] focused on UX elements and includes learner's utility of the technology in terms of customization, expectation of content placement, functionality of component parts, interface terms aligned with existing mental models, and navigation. Interaction with the learning space describes how the student perceives the interface elements, including engagement with the modality of content, dynamic interaction, perceived value of technology features to support learning, and scaffolding. Rather than see these as mutually exclusive, [they represent a] confluence of these design constructs. (p. 331)

This characterization of LXD highlights the importance of human-computer interaction (HCI) to technology-mediated learning. It is therefore unsurprising that many of the evaluation methods of LXD center around learner-computer interaction, that is, how learners actually use a digital learning technology product or service. However, learning designers cannot know a priori how learners will actually interact with a product (Gregg et al., 2020). Therefore, evaluation methods are critical to explore not only learners' perceptions of prototypes and fully developed products, but also to gain insight into learners' needs, preferences, and values related to envisioned products. In the following sections, we present various evaluation methods that are commonly used in UXD and usability research, as well as recommendations for when these evaluation methods are most appropriate.

Learn More About LXD and UXD Research in LIDT

To learn more about learner and user experience research in the field of LIDT, we recommend the open access edited volume *Learner and user experience research: An introduction for the field of learning design & technology*, provided here in EdTech Books!

Schmidt, M., Tawfik, A. A., Jahnke, I., & Earnshaw, Y. (2020). *Learner and user experience research: An introduction for the field of learning design & technology*. EdTech Books. https://edtechbooks.org/ux

Evaluation Methods

In LXD, knowing when and under what conditions to apply evaluation methodologies is a challenge. In the following sections, several evaluation methodologies commonly used in LXD are described, with descriptions of how these evaluation methodologies can be used in a learning design context. These can be applied during various phases across the learning design and development process (i.e., front-end analysis, low fidelity to high-fidelity prototyping). While a case can be made that any of the approaches can be applied to a given design phase, some evaluation methodologies are more appropriate to the overall learning experience, while others focus more on usability. Table 1 provides an overview of methods, the associated design phases in which they can most optimally be implemented, and their associated data sources.

Method	Design phase				Data source
		Prototyping			
	Front-end analysis	Paper (low fidelity)	Wireframe (medium fidelity)	Functional (high fidelity)	
Ethnography	x				Single user or users
Focus groups	x	x			Group of users
Card sorting	x	x			Single user, multiple users, or group fo users
Cognitive walkthrough		x	x	x	Expert
Heuristic evaluation		x	x	x	Experts
A/B testing		x	x	x	Multiple users
Think-aloud		x	x	x	Multiple users
EEG/Eye Tracking				x	Multiple users
Analytics				x	Multiple users

Table 1. Evaluation Methodologies, Design Phases, and Data Sources

Ethnography

A method that is used early in the front-end analysis phase, especially for requirements gathering, is ethnography. Ethnography is a qualitative research method in which a researcher studies people in their native setting (not in a lab or controlled setting). During data collection, the researcher observes the group, gathers artifacts, records notes, and performs interviews. In this phase, the researcher is focused on unobtrusive observations to fully understand the phenomenon in situ. For example, in an ethnographic interview, the researcher might ask open-ended questions but would ensure that the questions were not leading. The researcher would note the difference between what the user is doing versus what the user is saying and take care not to introduce their own bias. Although this method has its roots in the field of cultural anthropology, UCD-focused ethnography can support thinking about design from activity theory and distributed cognition perspectives (Nardi, 1996). This allows the researcher to gather information about the users, their work environment, their culture, and how they interact with the device or website in context (Nardi, 1997). This information is particularly valuable when writing user personas and scenarios. Ethnography is also useful if the researcher cannot conduct user testing on systems or larger equipment due to size or security restrictions.

A specific example of how ethnography can be applied in learning design is in the development of learner personas. Representative learners can be recruited for key informant interviews with the purpose of gathering specific data on what a learner says, thinks, does, and feels, as well as what difficulties or notable accomplishments they describe. The number of participants needed depends on the particular design context but does not need to be large. Indeed, learning designers can glean critical insights from just a few participants, and there is little question that even small numbers of participants is better than none. For example, to develop online learning resources for parents of children with traumatic brain injuries, a learning designer might interview two or three parents and ask them to relay what their typical day looks like, tell a story about a particular challenge they have encountered with parenting their child, or describe how they use online resources to find information about traumatic brain injuries. The interviews could then be transcribed, and the learning designer could use a variety of analysis techniques to categorize the interview data thematically. This information from thematic categories could then be generalized into the development of learner personas that are illustrative of themes derived from the key informant interviews.

Learn More About Conducting Thematic Analysis

For information on how to conduct a thematic analysis on interviews, refer to Mortensen (2020).

Learning Check

(True/False) Ethnography can be used to gather information about users' work environment and culture.

True

False

Focus Groups

Focus groups are often used during the front-end analysis phase. Rather than the researcher going into the field to study a larger group as is done in ethnography, a small group of participants (n = 5–10) are recruited based on shared characteristics. Focus group sessions are led by a skilled moderator who uses a semi-structured set of questions. For instance, a moderator might ask what challenges a user faces at work (i.e., actuals vs. optimals gap), suggestions for how to resolve them, and provide feedback on present technologies. The participants are then asked to discuss their thoughts on products or concepts that the moderator/group of learning designers propose. The moderator may also present a low-

fidelity prototype to the prospective user and ask for feedback. The role of the researcher in a focus group is to ensure that no single person dominates the conversation in order to hear everyone's opinions, preferences, and reactions. This helps to determine what users want and keeps the conversation on track. It is preferred to have multiple focus group sessions to ensure various perspectives are heard in case a conversation gets side-tracked. Analyzing data from a focus group can be as simple as providing a short summary with a few illustrative quotes for each session. The length of the sessions (typically 1–2 hours) may include some extraneous information, so it is best to keep the report simple.

For example, a learning designer developing an undergraduate-level introduction to nuclear engineering course invited a group of nuclear engineers, radiation protection technicians, and undergraduate-level nuclear engineering students to participate in a focus group. Before meeting with the focus group, the learning designer created a semi-structured set of questions to guide the session. These questions focused on issues such as the following: the upcoming challenge of an aging workforce on the brink of retirement and with no immediate replacements, the stigma of nuclear power, and the perceived difficulty of pursuing a career in nuclear engineering that the designer had gleaned from discussions with SMEs and from a document analysis. These issues were then explored with the focus group participants during a focus group session, with the designer acting as a facilitator. Sticky notes were used to document key ideas and posted around the room. Participants were asked to use sticky notes to provide brief responses to facilitator questions. The facilitator then asked the participants to find the sticky notes posted on the walls that best aligned with the responses they had provided and post their sticky notes near those sticky notes. These groups of notes were then reviewed by the participant groups, refined, and then named. The entire process took two hours. These categorized groups of sticky notes served as the foundation for the content units in the online course, covering topics like the application of nuclear medicine in cancer diagnosis and treatment, as well as the use of irradiation to extend the shelf life of food.

Learning Check

(True/False) Analyzing data from a focus group should involve providing a detailed report with extensive quotes for each session.

True

False

Card Sorting

Aligning designs with users' mental models is important for effective UX design. A method used to achieve this is card sorting. Card sorting is used during front-end analysis and paper prototyping. Card sorting is commonly used in psychology to identify how people organize and categorize information (Hudson, 2012). In the early 1980s, card sorting was applied to organizing menuing systems (Tullis, 1985) and information spaces (Nielsen & Sano, 1995).

Card sorting can be conducted physically using tools like index cards and sticky notes or it can be conducted electronically. It can involve a single participant or a group of participants. With a single participant, they group content (individual index cards) into categories, allowing the researcher to evaluate the information architecture or navigation structure of a website. For example, a participant might organize "Phone Number" and "Address" cards together. When a set of cards is placed together by multiple participants, this suggests to the designer distinct pages that can be created (e.g., a "Contact Us" page). When card sorting with a group of participants instead of just one person, the same method is employed, but the group negotiates how they will sort content into categories. How participants arrange cards provides insight into mental models and how they group content.

No-cost tools like [Lloyd Rieber's (2017) Q Sort](http://lrieber.coe.uga.edu/qsort/index.html) (http://lrieber.coe.uga.edu/qsort/index.html) can be used for card sorting.

There are two types of card sorting methods: open and closed. In an open card sort, a participant or group of participants will first group content (menu labels on separate notecards) into piles and then name the category. Participants can also place the notecards in an "I don't know" pile if the menu label is not clear or may not belong to a designated pile of cards. In a closed card sort, the categories will be pre-defined by the researcher. It is recommended to start with an open card sort and then follow-up with a closed card sort (Wood & Wood, 2008). As the arrangement of participants are compared, the designer designs a new prototype where the menu information and other features align with how the participants organize the information within their mind.

Learn More About Card Sorting Best Practices

For card sorting best practices, refer to "Card sort analysis best practices" ([Righi et al., 2013](#)).

Card sorting is particularly useful for learning designers who are creating courses in learning management systems. After identifying the various units, content categories, content sections, and so on, the learning designer can write what they identified on cards (or use other methods discussed above), present them to a SME, course instructor, or student, and ask them to arrange the cards into what they perceive to be the most logical sequence or organization. This approach can be particularly informative when comparing how instructors feel a course should be organized with how a learner feels a course should be organized, which can sometimes be quite different. Findings can then be used to inform the organization of the online course and potential navigational structures that are important to LXD.

Learning Check

What is the main difference between open card sorting and closed card sorting?

Open card sorting allows participants to create their own categories, while closed card sorting provides predefined categories.

Open card sorting involves using digital tools, while closed card sorting uses physical index cards.

Open card sorting is conducted with a group of participants, while closed card sorting is done individually.

Open card sorting requires paid technology tools, while closed card sorting can be performed using low-cost or no-cost tools.

Cognitive Walkthroughs

Cognitive walkthroughs (CW) can be used during all prototyping phases. CW is a hands-on inspection method in which an evaluator (not a user) evaluates the interface by walking through a series of realistic tasks (Lewis & Wharton, 1997). CW is not a user test based on data from users, but instead is based on the evaluator's judgments.

During a CW, a UX or LXD expert evaluates specific tasks and considers the user's mental processes while completing those tasks. For example, an evaluator might be given the following task: Recently you have been experiencing a technical problem with software on your laptop and you been unable to find a solution to your problem online. Locate the place where you would go to send a request for assistance to the Customer Service Center. The evaluator then identifies the correct paths to complete the task but does not make a prediction as to what a user will actually do. In order to assist designers, the evaluator also provides reasons for making errors (Wharton et al., 1994). The feedback received during the course of the CW provides insight into various aspects of the user experience including

- first impressions of the interface,
- how easy it is for the user to determine the correct course of action,
- whether the organization of the tools or functions matches the ways that users think of their work,
- how well the application flow matches user expectations,
- whether the terminology used in the application is familiar to users, and
- whether all data needed for a task is present on screen.

LIDT in the World

CW is particularly valuable when working in teams that consist of senior and junior learning experience designers. Junior learning experience designers can develop prototype learning designs (e.g., learning modules, screencasts, infographics), which can then be presented to the senior designer to perform a cognitive walkthrough. For example, a junior designer creates a series of five videos and sequences them in the LMS logically so as to provide sufficient information for a learner to correctly answer a set of corresponding informal assessment questions (e.g., a knowledge check). The junior designer then presents this to the senior designer with the following scenario: "You don't know the answer to the third question in the knowledge check, so you decide to review what you learned to find the answer." The senior designer then maps out the most efficient path to complete this task but finds that videos cannot be easily scrubbed by moving the playhead rapidly across the timeline. Instead, the playhead resets to the beginning of the video when it is moved. The senior designer explains to the junior designer that learners would have to completely rewatch each video to find the correct answer. The junior designer then has specific feedback that can be used to improve the learning experience for this learning module.

Heuristic Evaluation

Heuristic evaluation is an inspection method that does not involve working directly with the user. In a heuristic evaluation, it is recommended that at least two evaluators work independently to review the design of an interface against a predetermined set of usability principles (heuristics) before communicating their findings. Ideally, each evaluator will work through the interface at least twice: once for an overview of the interface and the second time to focus on specific interface elements (Nielsen, 1994). The evaluators then meet and reconcile their findings. This method can be used during any phase of the prototyping cycle.

Many heuristic lists exist that are commonly used in heuristic evaluations. The most well-known heuristic checklist was developed over 25 years ago by Jakob Nielsen and Rolf Molich (1990). This list was later simplified and reduced to 10 heuristics which were derived from 249 identified usability problems (Nielsen, 1994). In the field of LIDT, researchers have embraced and extended Nielsen's 10 heuristics to make them more applicable to the evaluation of eLearning systems (Mehlenbacher et al., 2005; Reeves et al., 2002). Not all heuristics are applicable in all evaluation scenarios, so UX designers tend to pull from existing lists to create customized heuristic lists that are most applicable and appropriate to their local context, as do LX designers.

Nielsen's 10 Heuristics

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

Schmidt provides this easy-to-use learning design heuristics worksheet (MS Excel format) at no cost, based on Mehlenbacher et al. (2005) task-oriented usability heuristics for web -based instruction design and evaluation.

https://docs.google.com/spreadsheets/d/1cmqopfqgc-ejhCCBtvEkVLkWn86UOcKi/

An approach that bears similarities with a heuristic evaluation is the expert review. In an expert review, the expert is knowledgeable about usability principles and has worked directly with users in the past. Expert reviewers do not always use a set of heuristics, but instead they may produce a document that details the overall issues, ranks them in order of severity, and then provides recommendations on how to mitigate the issues. This more informal approach allows for more flexibility than using a heuristic list. As is the case with the heuristic evaluation, multiple experts should be involved and data from all experts should be aggregated. This is because expert review is particularly vulnerable to the expert's implicit biases. Different experts will have different perspectives and therefore will uncover different issues. Involving multiple experts helps ensure that implicit bias is minimized and that problems are not overlooked.

For learning designers developing online courses, established quality metrics such as Quality Matters (QM) can be used for guiding heuristic evaluations (Zimmerman et al., 2020). QM provides evaluation rubrics for certified evaluators to assess the degree to which an online course meets QM standards. The aggregate QM score can then be used as a quality benchmark for that course. However, when applied in the context of a heuristic evaluation, the QM materials should only be used to evaluate prototypes for making improvements—not for establishing a quality benchmark for a finalized course. A QM-guided heuristic evaluation performed by a skilled evaluator can provide tremendously valuable insights along the dimensions of learning experience that are outlined above. These can serve as the basis for subsequent design refinements to an online course. These insights, in turn, promote a more positive learning experience.

Learn More About Heuristics

For details on heuristics, we recommend reading Jahnke et al. (2021)'s article titled "Advancing sociotechnical-pedagogical (STP) heuristics for the usability evaluation of online courses for adult learners," https://olj.onlinelearningconsortium.org/index.php/olj/article/view/2439

Learning Check

Select the most appropriate response to complete the following statement: Usability testing and Nielsen's heuristics are for . . .

Testing the user's ability to effectively and efficiently complete a task

Evaluating the user's interaction with the digital technology, product, or service

A/B Testing

A/B testing or split-testing compares two versions of a user interface; because of the nature of this method, all three prototyping phases can be employed at the same time. The different interface versions might utilize different screen elements (such as the color or size of a button), typefaces, textbox placements, or overall general layouts. During A/B testing, it is important that the two versions are tested at the same time by the same user. For instance, Version A can be a control and Version B should only have one variable that is different (e.g.,

navigation structure). A randomized assignment, in which some participants receive Version A first and then Version B (versus receiving Version B and then Version A), should be used.

LIDT in the World

Learning experience designers do not frequently have access to large numbers of learners for A/B testing, and therefore need to consider how to adapt this approach to specific design contexts. For example, a design team building a case library for a case-based learning environment is struggling with the design of the cases themselves. One learning experience designer has created a set of cases that highlight the central theme of the different cases (i.e., constant responsibilities, preparatory activities, recruitment, training, and the selection process); however, the chosen texts are fairly text heavy. Another learning experience designer has taken a different design approach and created a comic book layout for the cases, which has visual appeal. However, the central theme of the cases is not emphasized. The design team asks six students to review the designs. Three students review the more thematically focused cases and three review the comic book cases. The students are then asked to create a concept map that shows the central themes of the cases and how those themes are connected. The design team learns that students who used the thematically focused cases spent much less time reviewing the cases, and their concept maps show a very shallow understanding of the topic—although, they did appropriately identify thematic areas of the cases (i.e., constant responsibilities, recruitment, etc.). The students who used the comic book cases spent more time reviewing the cases. Their concept maps are richer and show a more nuanced understanding of the topic but are missing the specific names of the thematic areas (although they describe the areas in their own words). With this information, the team decides to continue to iterate prototypes of the comic book design while focusing on better emphasizing the central themes within those cases. On this basis, a potentially more effective learning experience was uncovered.

Learn More About A/B Testing

To learn more about A/B Testing, we recommend reading Kimmons (2021).

Think-Aloud User Study

Unlike A/B testing, a think-aloud study is only used during the functional prototyping phase. According to Jakob Nielsen (1993), "thinking aloud may be the single most valuable usability engineering method" (p. 195). In a think-aloud user study, a single participant is tested at any given time. The participant narrates what he or she is doing, feeling, and thinking while looking at a prototype (or fully functional system) or completing a task. This method can seem unnatural for participants, so it is important for the researcher to encourage the participant to continue verbalizing throughout a study session.

Learn More About Think-Aloud Usability Studies

To view an example of a think-aloud usability study, we recommend the video (24 minutes) from Peachpit TV (2010) on *Rocket Surgery Made Easy by Steve Krug: Usability Demo*.

Krug (n.d.) also provides useful scripts that are freely available for you to download and adjust.

A great deal of valuable data can come from a think-aloud user study (Krug, 2009). Sometimes participants will mention things they like or dislike about a user interface. This is important to capture because their opinions may not be discovered in other methods. However, the researcher needs to also be cautious about changing an interface based on a single comment.

Users do not necessarily have to think aloud while they are using the system. The retrospective think aloud is an alternative approach that allows a participant to review the recorded testing session and talk to the researcher about what he or she was thinking during the process. This approach can provide additional helpful information, although it may be difficult for some participants to remember what they were thinking after some time. Hence, it is important to conduct retrospective think aloud user testing as soon after a recorded testing session as possible.

Learn More About Conducting Think-Aloud User Testing

For a primer on how to conduct think-aloud user testing, refer to the U.S. government's online resources for usability at https://www.usability.gov (U.S. Dept. of Health and Human Services, n.d.)

Think-aloud testing does not test the user but the interaction of the user with the technology, product, or service. It is the most widely used method of usability evaluation in practice, including in the field of LIDT. Indeed, usability testing has long been recognized as a useful evaluation method in the design of interactive learning systems (cf. Reeves & Hedberg, 2003). Increasingly, usability testing is gaining acceptance in LIDT as a viable and valuable evaluation method for informing research related to advanced or novel learning technologies, for which existing research is neither substantial nor sufficient, such as 360-video based virtual reality (Schmidt et al., 2019) or digital badging (Stefaniak & Carey, 2019). Given the limited resources provided to learning designers, think aloud user testing is particularly attractive because it can be conducted with relatively small numbers of participants (5–12 users depending on the complexity of the system) and with open source or free-to-use tools.

LIDT in the World

Learn how learning designers apply think-aloud techniques in the AECT Design & Development Webinar (58 minutes) on "Think-Aloud Methods: Just-in-Time & Systematic Methods to Improve Course Design" by Gregg et al. (2022). https://edtechbooks.org/dd_chronicles/lxd_tao

Further details can be found in the chapter "Think-Aloud Observations to Improve Online Course Design: A Case Example and "How-to" Guide" by Gregg et al. (2020).

Learning Check

When is the best time for a think-aloud user study to be conducted in the design and development process?

During the initial brainstorming phase.

Only during A/B testing.

Primarily during the functional prototyping phase.

At any phase of the design process.

Eye-Tracking

Similar to the think-aloud user study, eye-tracking is an evaluation method that involves the user during the functional prototype phase. Eye-tracking is a psychophysiological method used to measure a participant's physical gaze behavior in responses to stimuli. Instead of relying on self-reported information from a user, these types of methods look at direct, objective measurements in the form of gaze behavior. Eye-tracking measures saccades (eye movements from one point to another) and fixations (areas where the participant stops to gaze at something). These saccades and fixations can be used to create heat maps and gaze plots, as shown in Figures 1–3, or for more sophisticated statistical analysis.
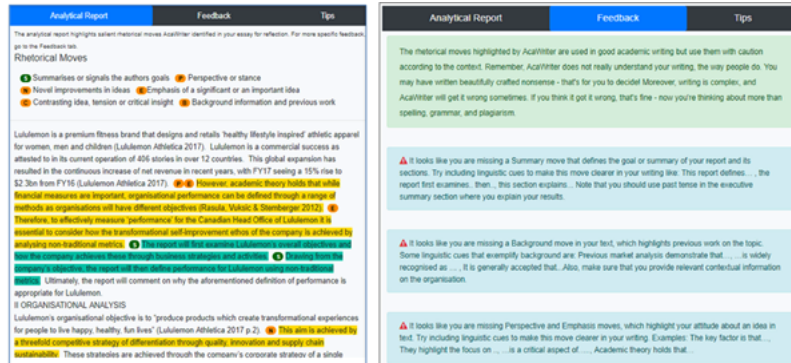

Figure 1. Heat map of a functional prototype's interface designed to help learners with Type 1 Diabetes learn to better manage their insulin adherence; here, eye fixations are shown with red indicating longer dwell time and green indicating shorter dwell time. Photo courtesy of the Advanced Learning Technologies Studio at the University of Florida. Used with permission.
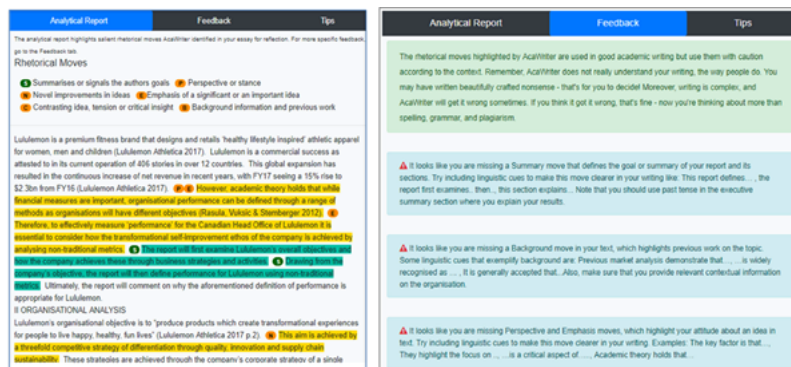

Figure 2. Heat map of a three-dimensional interface showing eye fixations and saccades in real-time, with yellow indicating longer dwell time and red indicating shorter dwell time. Adapted from "The best way to predict the future is to create it: Introducing the holodeck mixed-reality teaching and learning environment," by M. Schmidt, J., Kevan, P. McKimmy, and S. Fabel, 2013, *Proceedings of the 2013 International Convention of the Association for Educational Communications and Technology*, Anaheim, CA. Reprinted with permission.
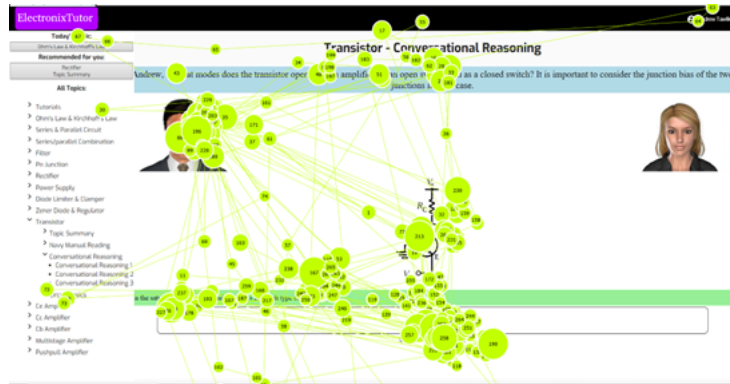
Figure 3. Gaze plot of a learner engaged with the ElectronixTutor learning environment adapted from Tawfik et al. (2022). Photo courtesy of the Instructional Design Studio at the University of Memphis. Used with permission.

LIDT in the World

Conley et al. (2020) used eye-tracking to examine two different layouts (functional and chronological) in Blackboard in their article "Examining course layouts in Blackboard: Using eye-tracking to evaluate usability in a learning management system," https://doi.org/10.1080/10447318.2019.1644841

Electroencephalogy

Another psychophysiological method used to directly observe participant behavior is electroencephalogy (EEG). EEG measures participant responses to stimuli in the form of electrical activity in the brain. An EEG records changes in the brain's electrical signals in real-time. A participant wears a skull cap (Figure 4) with tiny electrodes attached to it. While viewing a prototype, EEG data such as illustrated in Figure 5 can show when a participant is frustrated or confused with the user interface (Romano Bergstrom et al., 2014).

From the perspective of learning design, eye tracking and EEG-based user testing are typically reserved for very large training programs (i.e., for large corporations like Apple or Facebook) or for learning designs that are more focused on research than on practical application. It is not very common for small learning design teams to have access to EEG and eye tracking resources. Nonetheless, these approaches can serve as a way to understand when learners find something important, distracting, disturbing, etc., thereby informing learning designers of factors that can impact extraneous cognitive load, arousal, stress, and other factors relevant to learning and cognition. A disadvantage of this type of data, for example, is that it might not be clear why a learner was fixated on a search box, why a learner showed evidence of stress when viewing a flower, or if a fixation on a 3D model of an isotope suggests learner interest or confusion. In these situations, a retrospective think-aloud can be beneficial. After the eye-tracking data have been collected, the learning designer can sit down with a participant and review the eye-tracking data while asking about eye movements and particular focus areas.

Figure 4. A research study participant wears an EEG while viewing an interface. Photo courtesy of the Neuroscience Applications for Learning (NeurAL Lab) at the University of Florida's Institute for Advanced Learning Technologies (IALT). Used with permission.
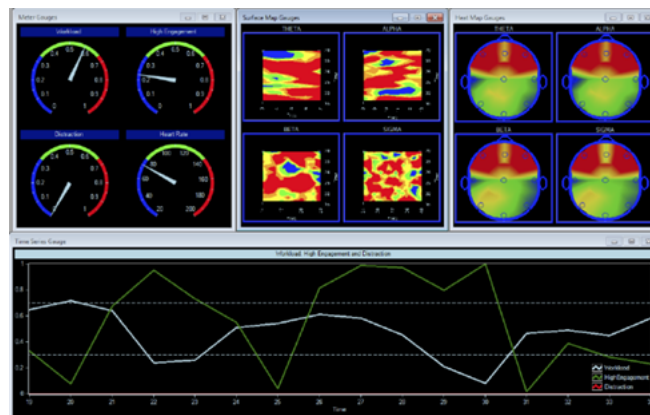

Figure 5. Output from an EEG device in a data dashboard displaying a variety of psychophysiological measures (e.g., workload, engagement, distraction, heart rate). Photo courtesy of the Neuroscience Applications for Learning (NeurAL Lab) at the University of Florida's Institute for Advanced Learning Technologies (IALT). Used with permission.

Analytics

A type of evaluation method that is gaining significant traction in the field of learning design due to advances in machine learning and data science is analytics (e.g., learning analytics). Analytics are typically collected automatically in the background while a user is interfacing with a system and sometimes without the user even being aware the data is being collected. An example of analytics data is a clickstream analysis in which the participants' clicks are captured while browsing the web or using a software application (see Figure 6). This information can be beneficial because it can show the researcher the path the participant was taking while navigating a system. Typically, these data need to be triangulated with other data sources to paint a broader picture.

Where can we find the data?

Student Information System	Learning Environment	Research Instruments	Multimodal	Social Media
• demographics • timetables • enrolment • attendance • previous assessment results	• trace data • simulation data • assessment data • social interaction • content interaction • intelligent tutoring systems • educational context data	• surveys • questionnaires • interviews • focus groups • observations	• video • audio • gesture • gaze • psychophysiological data • EEGs • fMRI	• Twitter • Instagram • Facebook • blogs

Figure 6. An example of a clickstream, showing users' paths through a system. Adapted from "Transforming a problem-based case library through learning analytics and gaming principles: An educational design research project," by M. Schmidt and A. Tawfik, 2017, *Interdisciplinary Journal of Problem-Based Learning*. Reprinted with permission.

Increasingly, learning analytics and data dashboards such as LMSs, video conferencing suites, video hosting providers, and more are being incorporated into the tools of the learning design trade. Indeed, the massive collection of learners' personal usage data has become so ubiquitous that it is taken for granted. However, analytics and data dashboards remain novel tools that learning designers do not necessarily have the training to use for making data-based decisions for improving learning designs. That said, data dashboards are maturing quickly. Less than a decade ago, only the most elite learning designers could incorporate learning analytics and data dashboards into their designs, whereas today these tools are built-in to most tools. Clearly, these tools have enormous potential in the field of LIDT; for example, these tools could be beneficial for creating personalized learning environments, providing individualized feedback, improving motivation, and so-on. With advances in machine learning and artificial intelligence (AI), learning analytics hold great promise. However, privacy concerns, questions of who owns and controls learner data, and other issues remain. Learning designers are encouraged to carefully review the data usage agreements of the software used for developing and deploying digital environments for learning. LXD considers the entire experience of the learner when using a technology, which includes their experiences with the collection of personal data. Carefully safeguarding this data and using it judiciously is paramount for a positive learning experience.

Learning Check

What is one of the primary benefits of analytics data, such as clickstream analysis, in the context of learning design?

Analytics data can replace the need for user testing.

Analytics data are collected manually by users during their interactions.

Analytics data provide insights into the path participants take while navigating a system.

Analytics data are typically used as the sole source of information for decision-making.

Evaluating the Educational Impact of Digital Learning Experiences

A range of evaluation techniques can be used to evaluate the educational impact of digital learning experiences, including pre/posttests and concept maps. Pre/posttests help answer the question of whether the learning design is effective. Pre/posttests are performed with an identical set of measurement items before and after the learning design. In the pretest,

the learner's knowledge is captured as baseline, and in the posttest, the difference between the pre- and posttest scores indicates the level of learning growth. This technique is quick and easy to apply; however, it is limited in that it typically is only able to measure lower-order learning outcomes such as memorization/recall. For more intricate higher-order learning objectives, such as synthesis and problem-solving, alternative methods prove more suitable. These may include collaborative design, simulation tasks, or more advanced pre/posttest designs that extend beyond mere information recall. For example, concept mapping allows learners to represent their understanding of concepts using line-and-node visualizations (Borrego et al., 2009). A concept mapping task might ask learners to map out all of the things they know about a particular content area (i.e., different kinds of poems, how the animal kingdom is classified, etc.). This is done both before and after the learning experience, after which researchers can compare the differences.

Learn More About Evaluating the Educational Impact of Digital Learning Experiences

For further details, we suggest two case studies of learning experience design that give practical insights into iterative development and testing of the LXD including effectiveness, efficiency and appealing:

Lee et al. (2021), "Mobile microlearning design and effects on learning efficacy and learner experience," https://doi.org/10.1007/s11423-020-09931-w

Li et al. (2021), "Digital learning experience design and research of a self-paced online course for risk-based inspection of food imports," https://doi.org/10.1016/j.foodcont.2021.108698

Conclusion

In this chapter, we have provided examples of commonly used evaluation methodologies that can be employed to advance usable and pleasing learning designs, along with illustrative examples of how these methods can be used in practice. A design approach that connects the evaluation methods of UX and HCI with LXD can help ensure that digital environments for learning are constructed to support learners' achievement of their learning goals in ways that are effective, efficient, and satisfying.

Think About It!

1. In your role as a learning experience designer, reflect on a project you are currently involved in. How can you strategically combine different evaluation methods discussed in this chapter to maximize the effectiveness of your evaluation process for this project? Explain your rationale for selecting specific evaluation methods and their potential synergies.
2. In the context of the evaluation methods discussed in this chapter, how might you adapt and combine different evaluation approaches to gain a deeper understanding of the impact of a learning design? Consider the potential challenges and benefits of combining quantitative and qualitative data, user testing, and analytics in your evaluation process.
3. Evaluating learning experiences often involves collecting and analyzing user data. What ethical considerations should learning designers and evaluators keep in mind when working with user data for evaluation purposes? How can you ensure that the collection and use of data respect the privacy and rights of learners while still providing valuable insights?

References

Borrego, M., Newswander, C. B., McNair, L. D., McGinnis, S., & Paretti, M. C. (2009). Using concept maps to assess interdisciplinary integration of green engineering knowledge. *Advances in Engineering Education*, *1*(3), 1–26.

Conley, Q., Earnshaw, Y., & McWatters, G. (2020). Examining course layouts in Blackboard: Using eye-tracking to evaluate usability in a learning management system. *International Journal of Human-Computer Interaction*, *36*(4), 373–385. https://doi.org/10.1080/10447318.2019.1644841

Gregg, A., Reid, R., Aldemir, T., Garbrick, A., & Gray, J. (2022). LXD webinar series - Think-aloud methods: Just-in-time & systematic methods to improve course design. In T. R. Huang & M. Schmidt, Design and Development Chronicles. EdTech Books. https://edtechbooks.org/dd_chronicles/lxd_tao

Gregg, A., Reid, R., Aldemir, T., Gray, J., Frederick, M., & Garbrick, A. (2020). Think-aloud observations to improve online course design: A case example and "how-to" guide. In M. Schmidt, A. A. Tawfik, I. Jahnke, & Y. Earnshaw (Eds.), *Learner and user experience research: An introduction to the field of learning design & technology*. EdTech Books. https://edtechbooks.org/ux/15_think_aloud_obser

Hudson, W. (2012). Card sorting. In M. Soegaard & R. F. Dam (Eds.), *The encyclopedia of human-computer interaction* (2nd ed.). Interaction Design Foundation. https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/card-sorting

Jahnke, I., Riedel, N., Singh, K., & Moore, J. (2021). Advancing sociotechnical-pedagogical heuristics for the usability evaluation of online courses for adult learners. *Online Learning*, *25*(4). https://doi.org/10.24059/olj.v25i4.2439

Krug, S. (n.d.). Downloads. Steve Krug. https://sensible.com/download-files/

Krug, S. (2009). *Rocket surgery made easy: The do-it-yourself guide to finding and fixing usability problems*. New Riders.

Lee, Y.-M., Jahnke, I., & Austin, L. (2021). Mobile microlearning design and effects on learning efficacy and learner experience. *Educational Technology Research and Development*, *69*, 885–915. https://doi.org/10.1007/s11423-020-09931-w

Lewis, C., & Wharton, C. (1997). Cognitive walkthroughs. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 717–732). Elsevier.

Li, S., Singh, K., Riedel, N., Yu, F., & Jahnke, I. (2021). Digital learning experience design and research of a self-paced online course for risk-based inspection of food imports. *Food Control*, *135*. https://doi.org/10.1016/j.foodcont.2021.108698

Mehlenbacher, B., Bennett, L., Bird, T., Ivey, I., Lucas, J., Morton, J., & Whitman, L. (2005). Usable e-learning: A conceptual model for evaluation and design. *Proceedings of HCI International 2005: 11th International Conference on Human-Computer Interaction, Volume 4 — Theories, Models, and Processes in HCI*. Las Vegas, NV.

Mortensen, D. H. (2020). How to do a thematic analysis of user interviews. https://www.interaction-design.org/literature/article/how-to-do-a-thematic-analysis-of-user-interviews

Nardi, B. A. (1996). Studying context: A comparison of activity theory, situated action models, and distributed cognition. In B. A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 69–102). The MIT Press.

Nardi, B. A. (1997). The use of ethnographic methods in design and evaluation. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 361–366). Elsevier.

Nielsen, J. (1993). *Usability engineering*. Morgan Kaufmann.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25–62). John Wiley & Sons.

Nielsen, J., & Molich, R. (1990). *Heuristics evaluation of user interfaces. Proceedings of ACM CHI'90 Conference*. Seattle, WA.

Nielsen, J., & Sano, D. (1995). SunWeb: user interface design for Sun Microsystem's internal Web. *Computer Networks and ISDN Systems*, *28*(1), 179–188.

Norman, D. A. (2013). *The design of everyday things*. Basic Books.

Peachpit TV. (2010, February 3). Rocket surgery made easy by Steve Krug: Usability demo [video]. YouTube. https://www.youtube.com/watch?v=QckIzHC99Xc

Reeves, T. C., Benson, L., Elliott, D., Grant, M., Holschuh, D., Kim, B., Kim, H., Lauber, E., & Loh, S. (2002). Usability and instructional design heuristics for e-learning evaluation. *Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications*. Denver, CO.

Reeves, T. C., & Hedberg, J. G. (2003). *Interactive learning systems evaluation*. Educational Technology Publications.

Rieber, L. (2017). Lloyd's Q Sort Tool for Teaching [Computer Software]. http://lrieber.coe.uga.edu/qsort/index.html

Righi, C., James, J., Beasley, M., Day, D. L., Fox, J. E., Gieber, J., Howe, C., & Ruby, L. (2013). Card sort analysis best practices. *Journal of Usability Studies*, *8*(3), 69–89. http://uxpajournal.org/card-sort-analysis-best-practices-2/

Romano Bergstrom, J. C., Duda, S., Hawkins, D., & McGill, M. (2014). Physiological response measurements. In J. Romano Bergstrom & A. Schall (Eds.), *Eye tracking in user experience design* (pp. 81–110). Morgan Kaufmann.

Schmidt, M., & Huang, R. (2022). Defining learning experience design: Voices from the field of learning design and technology. *TechTrends*, *66*(2), 141–158. https://doi.org/10.1007/s11528-021-00656-y

Schmidt, M., Kevan, J., McKimmy, P., & Fabel, S. (2013). The best way to predict the future is to create it: Introducing the Holodeck mixed-reality teaching and learning environment. *Proceedings of the 2013 International Convention of the Association for Educational Communications and Technology*, Anaheim, CA.

Schmidt, M., Schmidt, C., Glaser, N., Beck, D., Lim, M., & Palmer, H. (2019). Evaluation of a spherical video-based virtual reality intervention designed to teach adaptive skills for adults with autism: A preliminary report. *Interactive Learning Environments*, 1–20. https://doi.org/10.1080/10494820.2019.1579236

Schmidt, M., & Tawfik, A. (2017). Transforming a problem-based case library through learning analytics and gaming principles: An educational design research approach. Interdisciplinary *Journal of Problem-Based Learning*, *12*(1). https://doi.org/10.7771/1541-5015.1635

Schmidt, M., & Tawfik, A. (2022). Activity theory as a lens for developing and applying personas and scenarios in learning experience design. *The Journal of Applied Instructional Design*, *11*(1). https://edtechbooks.org/jaid_11_1/activity_theory_as_a

Schmidt, M., Tawfik, A. A., Jahnke, I., & Earnshaw, Y. (2020). *Learner and user experience research: An introduction for the field of learning design & technology*. EdTech Books. https://edtechbooks.org/ux

Stefaniak, J., & Carey, K. (2019). Instilling purpose and value in the implementation of digital badges in higher education. *International Journal of Educational Technology in Higher Education*, *16*(44). https://doi.org/10.1186/s41239-019-0175-9

Tawfik, A. A., Gatewood, J., Gish-Lieberman, J. J., & Hampton, A. J. (2022). Toward a definition of learning experience design. *Technology, Knowledge and Learning*, *27*(1), 309–334. https://doi.org/10.1007/s10758-020-09482-2

Tullis, T. S. (1985). Designing a menu-based interface to an operating system. In L. Borman and B. Curtis (Eds.). *Proceedings of the ACM CHI 85 Human Factors in Computing Systems Conference*. San Francisco, CA.

U.S. Dept. of Health and Human Services. (n.d.). *usability.gov: Improving the user experience*. https://usability.gov

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 105–140). John Wiley & Sons.

Wood, J. R., & Wood, L. E. (2008). Card sorting: Current practice and beyond. *Journal of Usability Studies*, *4*(1), 1–6.

Zimmerman, W., Altman, B., Simunich, B., Shattuck, K., & Burch, B. (2020). Evaluating online course quality: A study on implementation of course quality standards. *Online Learning*, *24*(4), 147–163. https://doi.org/10.24059/olj.v24i4.2325