# Methods of evaluation and reflection in design research

PrePrint of: McKenney, S. & Reeves, T. C. (2014). Methods of evaluation and reflection in design research. Zeitschrift für Berufs- und Wirtschaftspädagogiek, 27, 141-153.

# Methods of evaluation and reflection in design research+

*Susan McKenney[1,2] & Thomas C. Reeves[3]*
*[1]University of Twente, Enschede, the Netherlands*
*[2]Open University of the Netherlands, Heerlen, the Netherlands*
*[3]University of Georgia, Athens, GA, USA*

**Abstract**
This contribution is concerned with the role of evaluation in design research with respect to the twin goals of informing the external scientific community about any new knowledge attained as well as driving the development of an effective intervention. This kind of evaluation is conscious, systematic, and formalized. Here, the term, evaluation, is used in a broad sense to refer to any kind of empirical testing of interventions that have been mapped out (designs) or constructed (prototypes). Reflection pertains to retrospective consideration of findings and observations rather than to assessment of personal performance. In this contribution, evaluation is described as a series of 10 steps, starting with establishing a focus and concluding with reporting. Thereafter, the role of reflection and processes that support it are addressed.

*+Portions of this contribution are based on previously published work (McKenney & Reeves, 2012)*

## Positioning evaluation and reflection in a larger process

Educational researchers have long sought to answer the question "What works?" with respect to teaching and learning with little success (Schoenfeld, 2006). Design research fundamentally changes the focus of research from "what works?" questions to "how can we make this work and why?" intentions. The goals of design research are twofold, first to solve real world problems through the development of effective interventions and second to distill theoretical understanding that can serve the work of others.. Educational interventions developed through design research are created, tested, adopted, implemented, re-tested and refined through iterative cycles in authentic settings (Bannan-Ritland & Baek, 2008; Euler, this issue; Jonassen, Cernusca & Ionas, 2007). McKenney and Reeves (2012, 2013) portray evaluation and reflection in relation to the overall process as shown in Figure 1. Ideally, design research does not cease until the desired levels of problem resolution are attained.

This contribution discusses the Evaluation and Reflection phase of design research with respect to the twin goals of driving the development of an effective intervention and informing the external scientific community about any new knowledge attained through the initiative. The Evaluation and Reflection phase provides a better understanding of the intervention, the appropriateness of its intentions, what it looks like when implemented, and the effects it yields under certain circumstances; this provides the basis for recommendations concerning refinement of the intervention. At the same time, this phase contributes to developing theoretical understanding.

*Recommendations for practice*
For many researchers, part of the appeal of educational design research is that insights from one study can be applied in the subsequent phase of design, a process that can be particularly gratifying. Consideration of potential refinements to design ideas or constructed prototypes is given in light of

the original intentions, and should take into account potential risks of pulling internal aspects of the design out of alignment. As noted by Wang and Hannafin (2005, p.10), "In addition to improving the ongoing design, researchers also consider the influence of en route changes on the integrity of the design … any changes to one aspect of the design need to be compatible with other aspects of the design." This process is not always clear-cut or smooth. Design research requires rigorous attention to detail accompanied by high skill in coping with ambiguity.

*Theoretical understanding*

Asking and answering research questions gradually leads to new insights, and often quickly leads to new questions. The theoretical understanding that is constructed through evaluation and reflection may hold local, middle-range or high-level applicability; this depends on the focus and methods of the study. A single cycle of evaluation and reflection can contribute empirical findings, hypotheses or conjectures that constitute the building blocks of theory. This understanding can be used to shape descriptions (e.g. how boys and girls engage with a particular learning environment); explanations (e.g. why boys tend to exhibit certain behavior more than girls when using a particular learning environment); or predictions concerning certain phenomena (e.g. if given specific scaffolds in a particular learning environment, boys will tend to do X and girls will tend to do Y). When integrated, such understanding can serve prescriptive purposes. For example, initial design *propositions* (untested, theory-based conjectures) that were used to shape the intervention are tested through evaluation. Reflection on the findings helps generate explanations for the results. As initial design propositions are validated, refuted or refined, design *principles* emerge. Design principles are empirically and theoretically underpinned guidelines for creating a particular class of intervention within a particular kind of context. Van den Akker (1999) recommends the following formula for conceptualizing design principles: "If you want to design intervention X [for purpose/function Y in context Z], then you are best advised to give that intervention the characteristics $C^1$, $C^2$, …$C^m$ [substantive emphasis] and to do that via procedures $P^1$, $P^2$, $P^n$ [procedural emphasis] because of theoretical arguments $T^1$, $T^2$…$T^p$ and empirical arguments $E^1$, $E^2$, … $E^q$." Aptly portraying salient features of the context, including actors, is critical to ensuring both external validity and case-to-case generalization (Firestone, 1993) of the findings

## Evaluation processes

Formative evaluation is conducted in earlier stages of design research, with the goal of investigating how design ideas and/or prototypes can be improved. Summative evaluation may also be conducted in design research, typically after interventions have matured, with the goal of assessing intervention quality and thereby validating or refuting design principles. Scriven (1991) clarified distinctions between formative and summative evaluation. Synthesis of these descriptions shows that the differences pertain mainly to seven parameters: aim (of the evaluation); application (how the information will be used); timing (in relation to the development of the evaluand); audience (who will use this information); evaluator (in relation to that which is being evaluated); methods (of investigation); and instrumentation (used to collect data). The ways in which these approaches tend to be employed in design research are illustrated in Figure 1.

| Parameters | Formative | Summative |
|---|---|---|
| *Aim* | Intent to improve evaluand | Basis for decision-making |
| *Application* | Varied types of improvements e.g.:<br>▪ Optimizing the development process<br>▪ Fine-tuning the intervention<br>▪ Refine the underlying design propositions | Varied types of decisions concerning intervention, e.g.:<br>▪ Adoption/implementation<br>▪ Go/continue/stop support or efforts<br>Varied types of decisions concerning principles, e.g.:<br>▪ Generalizability<br>▪ Validate/refute |
| *Timing* | During development | After development is complete or stabilized |
| *Audience* | Mostly internal | Mostly external |
| *Evaluator* | Usually internal (could also be external or both) | Usually external (could also be internal or both) |
| *Methods* | Exploratory, flexible | Unobtrusive, pre-determined |
| *Instrumentation* | Varied, often tailor-made, sometimes standardized | Reliable, validated, often publicly accepted |

*Figure 1: Formative and summative tendencies in design research evaluation*

Evaluation in design research may serve formative or summative purposes, depending on the stage of the project. The evaluation process is described below in terms of 10 steps, starting with establishing a focus and concluding with reporting. The recommendations are based on evaluation literature as well as ideas stemming from our own experiences conducting evaluations within design research.

*1: Establish the focus*

Evaluation may be conducted with more formative goals (identifying ways to improve an intervention) and/or more summative goals (assessing overall value of an intervention). The first steps in evaluation are to establish the decisions that must be made and to identify the evaluation questions that must be answered to guide those decisions. Simply put, we step back from the research and development work thus far and ask, "What do we need to know now?" to make the best possible decisions concerning the design or implementation of the intervention. Usually the questions that arise have to do with either testing a specific feature of the intervention or perhaps the intervention as a whole; exploring some phenomena that the intervention is known to or might engender; and/or gathering new inputs for design. Often new questions come directly from findings in a previous micro-cycle of evaluation and reflection. Establishing the focus is like setting an agenda, and thus benefits from being conducted through dialogue with fellow researchers and practitioners, while also informed by the relevant literature.

An intervention's stage of development usually has implications for the kinds of questions that can or should be addressed. Building on previous work (McKenney, 2001; Reeves & Hedberg, 2003), we distinguish six foci important in educational design research evaluation. Not all design studies attend to each focus. However, these foci exhibit similarities with those mentioned elsewhere in literature (Burkhardt, 2006; McKenney, Nieveen & van den Akker, 2006; Schoenfeld, 2009).

The six foci can be clustered into three concerns: internal structures of an intervention; how it is used in practice; and effects. These three clusters are well aligned with three stages of testing

adapted from software engineering and often seen in design research: alpha, beta and gamma. In the description below, two foci are described for each stage of testing.

Alpha testing concerns early assessment of design ideas, to test the internal structures of an intervention to make sure they work logically. Alpha studies involve the collection of data to primarily ascertain s*oundness* (the ideas underpinning a design and/or how those ideas are instantiated) and f*easibility* (the potential temporal, financial, emotional and human resource costs associated with creating the intervention).

Beta testing is conducted with working system components within a more or less functional system, and focuses on use in context. In design research, beta studies mainly explore *local viability* (how an intervention survives in the immediate context and why) and *institutionalization* (how an intervention becomes absorbed within the broader educational organization).

Gamma testing in design research takes place with a nearly final, or at least highly stable, version of the intervention. Gamma studies are mostly used to determine *effectiveness* (the extent to which the intervention meets its goals when implemented under representative conditions) and *impact* (the extent to which the intervention engenders a measurable change in the target setting).

Design researchers often struggle more with establishing the focus for an evaluative activity, more so earlier in the process than later. This is because so many factors are new that it can be difficult to choose which ones to investigate deeply. But practical limitations (e.g. time, resources), generally force tradeoffs between a broad-shallow versus narrow-deep orientation. Both options have their affordances and risks, and identifying the most appropriate orientation constitutes a key challenge during this step.

*2: Frame guiding questions*

Shavelson, Phillips, Towne and Feuer (2003) distinguished three generic questions often raised in design research and recommended methods that could suit each: What is happening?; Is there a systematic effect?; and Why or how is it happening? From the practical perspective, the primary concern is with how and to what extent the problem is being addressed by the intervention. From a theoretical perspective, the main concern is understanding how an intervention does or does not work and especially, why. Studying if, how, and why interventions work, requires attending to their goals, the ways those goals are pursued, and if or to what extent, they are achieved. This varies depending on the stage (alpha, beta and gamma) in which testing takes place.

Questions about the intentions of an intervention are typically the main focus in alpha testing. Here, the feasibility and soundness of the design ideas are studied by seeking answers to questions like:

- How robust and complete is the theoretical and pragmatic justification for these design ideas?
- How well are the core design propositions embodied in the design?
- What changes must be made to the underlying ideas or the design itself to increase the plausibility and probability that it could meet the intended goals?

Many factors affect how interventions are implemented in practice, including how well they are aligned with core elements of the context, such as the curriculum; assessment frameworks, practitioner expertise, prevailing attitudes and beliefs, local culture, and the textbooks, programs and materials that have already been adopted. Questions about the implemented intervention especially focus on if and how it engenders intended processes when used in the intended setting, and how sustainable they are likely to be (viability and institutionalization). Sample guiding questions during beta testing include:

- How relevant and usable do practitioners perceive and experience the intervention?
- What intended and unintended processes are engendered by the intervention?
- What makes embodiments of certain mechanisms more resilient than others?

Questions about the attained intervention guide investigation of effectiveness and impact. Sample guiding questions during gamma testing include:
- How effectively does the intervention solve the problem? Under which conditions?
- What is the long-term impact of the intervention?
- In representative classrooms, how prevalent are the conditions that foster and inhibit success?

Identifying appropriate questions tends to present less of a challenge to design researchers than selecting which ones will be answered. Choosing which questions to answer relates of course to the focus (Step 1) but is also influenced by the stakeholders involved. Even if a shared focus has been agreed upon, different stakeholders may value answers to certain questions in different ways. Also, some questions may be more feasible to investigate. The (in)feasibility of answering certain questions is not always clear at the time they are posed. Rather, this often becomes clearer as subsequent steps are undertaken; emerging insights concerning strategies, methods and planning often prompt refinement of questions asked.

*3: Select basic strategies*
There are many ways to study interventions in various stages of development. Inspired by Nieveen's (1999) approaches to formative evaluation, four are discussed here: developer screening, expert appraisal, pilots and tryouts. Developer screening is especially useful for studying the internal structure of a design or constructed prototype. It can also provide insight about how it will likely work in the target setting. Expert appraisal refers to a process whereby external experts in a particular area are enlisted to review (elements of) the intervention. Different types of experts can provide guidance for the systematic improvement and/or validation of interventions. Pilots can help researchers and practitioners begin to get a sense of how the intervention will perform in various contexts and what kind of real-world realities need to be addressed for the design to have a chance of success under representative conditions. The term, pilot, refers to any field testing of the intervention in settings that approximate, but do not completely represent, the target context. Tryouts are used to study how interventions work, what participants think or feel about them, and the results they yield. Tryouts take place when (a prototype of) the intervention is field tested in a natural setting.

Often these strategies are used in combination with one another, and different strategies are generally used at different stages in the lifecycle of a project. Strategies for evaluation are selected based primarily on the research questions, and secondarily on the constraints of the study (time, personnel, costs, access to respondents, etc.). To decide which strategies to use, it is useful to visualize how each one will relate to the evaluation functions. A matrix like the one offered in Figure 2 may help. Here, recommendations are given for strategies that serve particular evaluation functions, related to the alpha, beta and gamma testing phases.

| | | Developer screening | Expert appraisal | Pilot | Tryout |
|---|---|---|---|---|---|
| Alpha | Feasibility | | | | |
| | Soundness | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Beta | Local viability | | | | |
| | Institutionalization | | | | |
| Gamma | Effectiveness | | | | |
| | Impact | | | | |

Legend: dark grey = frequently useful; light grey = maybe useful, depending on the circumstances
Figure 2. *Functions and strategies matrix.*

Especially for new design researchers, the interpersonal dynamics of educational design can present particular challenges. For decades, scholars of educational design have cautioned that merely creating valid and consistent interventions will yield little impact (Plomp, 1982, Verhagen, 2000). While high quality designs are scientifically valid and internally consistent, they are also attuned to the contexts in which they will be used and – most critically – to the concerns of key stakeholders. For example, Kessels and Plomp (1999) stress the need for designs to be externally consistent (aligned with the priorities, values and interests of stakeholders); and Tessmer and Harris (1990) argue for educational designers to view themselves as change agents and in so doing, develop a sound understanding of the instructional environment (e.g. learners, teachers, physical spaces, resources and their use) and the support environment (e.g. management, dissemination). While most design researchers have a background in research, fewer have formal training in educational design. This may explain why critically important attention to local viability and institutionalization (during both design and formative evaluation) frequently seem under-appreciated.

*4: Determine specific methods*
Once basic strategies have been selected, the methods that will be used for the investigation must be determined. Design researchers draw from both quantitative and qualitative methods, often using a combination of the two. The methods are selected based on the most accurate and productive way to answer the research questions. Accurate methods are able to collect the specific kind of information that is needed to answer the research question(s) well. Productive methods make sense within the constraints of the project.

Seven methods are among the most common used during the evaluation and reflection phase: interviews; focus groups; observations; questionnaires/checklists; (pre/post) tests; logs/journals; and document analysis. Figure 3 offers generic recommendations for methods suitable to the four strategies described above.

| Methods | Developer screening | Expert appraisal | Pilot | Tryout |
|---|---|---|---|---|
| Interviews | | | | |
| Focus groups | | | | |
| Observations | | | | |
| Questionnaires/checklists | | | | |
| (Pre/post)tests | | | | |
| Logs/journals | | | | |
| Document analysis | | | | |

Legend: dark grey = frequently useful; light grey = may be useful, depending on the intervention type
Figure 3. *Strategies and methods matrix.*

One controversial issue concerns the reliability and validity of the instruments used to collect information when different methods are used. Traditional educational research places a great deal of attention on issues such as reliability and validity whereas in design research studies prototyping and testing often takes place so quickly that there is time insufficient time to validate instruments to the same degree that other types of studies might demand. The degree of effort put into validating instruments is related to the uses of the information collected with those instruments (Wiggins, 1989). In design studies, instruments of undetermined validity may suffice to provide adequate guidance for making design decisions on the fly during a small-scale pilot with primarily formative goals whereas if the same instruments were to be utilized within the context of a large-scale tryout with both formative and summative goals, the instruments would need to be more carefully validated.

*5: Draft and revise a planning document*

As methods are selected and methodological ideas begin to solidify, it is important to document and check the emerging evaluation plans. A planning document provides an overview of the activities and timelines for a study, and is useful for assessing how well the study is likely to meet its goals. A planning document can be checked for its methodological soundness (e.g. triangulation of data sources and data collection times) and feasibility (e.g. levels of invasiveness, costs or time needed). It can also help to plan and keep track of the many different tasks involved (e.g. meeting with stakeholders, creating resources, hiring facilitators, obtaining IRB approval, creating instruments, coaching assistants).

Reeves and Hedberg (2003) stressed the importance of sharing planning documents and instruments with all of the relevant stakeholders within an evaluation. Open sharing of planning documents is especially important in design studies when educational practitioners are involved as they usually are. For example, school districts often have rigorous requirements for data collection, e.g. only being able to apply to conduct research in schools at the beginning of a school year or at best a semester. Design researchers are often unable to specify in advance exactly what data will be needed when and how the data will be collected. Therefore, publically sharing an evolving plan with all stakeholders is required to build the trust and tolerance for ambiguity that is often needed for design research.

*6: Create or seek instruments*

In order to collect data, it is necessary to create the instruments to be used in the study, or to review, obtain and fine-tune existing instruments that have already been designed and validated. Because design research is often conducted to create new solutions to problems, it can be difficult to find suitable instruments. But searching can be worthwhile, as the time and effort involved in creating new ones that are reliable and valid can be substantial. Lean data collection whereby only the minimum amount of information needed to make specific decisions is recommended over strategies that might be used to collect "just in case" information (Reeves & Hedberg, 2003).

*7: Select participants*

Different participant populations may be sampled for different purposes. Common participants in intervention evaluation include developers, practitioners, experts and learners. Developers can range from members of the design team who have been engaged with the design task from early on, to individuals constructing sub-components of the intervention with minimal involvement in its conception. Practitioner involvement in evaluation may take place through many roles, including that of developer, expert, facilitator or implementer. Teachers are most frequently involved in evaluation. Other practitioner groups include educational specialists (e.g. library and media

specialists, remedial instructors or school psychologists) and educational leaders (e.g. principals, instructional coaches or department heads). Learners can be involved in design research when they are tested and/or observed during implementation of an intervention. In interventions targeting the education of young children, it is not unusual to ask parents to serve as respondents.

Whereas it would be ideal if exact representatives of the target population for an intervention could be recruited to participate in pilots and tryouts, this is often not feasible and sometimes not desirable. For example, it may not be feasible to involve certain respondent groups before an intervention is mature. When testing effectiveness, it may not be desirable for respondent groups to have had any previous exposure (e.g. during pilot testing) to the intervention. In such situations, it is recommended that the participants sought are as much like the eventual target audience as possible.

*8: Collect the data*

General recommendations concerning data collection are available in standard qualitative and quantitative research manuals. This section discusses an issue that tends to crop up often during data collection in educational design research; specifically the conflicting roles of advocate and critic (cf. Design-Based Research Collective, 2003). In educational design research, the same individuals are often simultaneously researchers, developers, facilitators and/or evaluators of the intervention.

For the advocate role, first-hand, detailed understanding of research findings is very beneficial. When designers participate in the implementation and testing of interventions, they are afforded the sometimes humbling opportunity to gain deeper and often sharper insights into not only the aspects of the design that succeed or fail, but underlying assumptions (e.g. about what motivates learners or what details teachers will care about). This provides rich opportunity for critical reflection and generating new ideas, as inputs for redesign are immediate and, coming from live observation, often more powerful than second-hand findings. An open-minded designer is quite likely to 'see' implications for re-design during (or inspired by) pilot or tryout activities. For the critic role, the designer mindset also has benefits, as the intentions of the design are clearly understood. The critical researcher can see, for example, how well instruments are measuring the phenomenon they were intended to measure. The need for making methodological adjustments may be more apparent to someone who deeply understands both the design intentions and the research goals.

But this comes at a cost, for the advocate and especially the critic. The methodological concerns, particularly bias, are substantial. Despite efforts to stimulate criticism, the potential for socially desirable responses stands to increase when participants know that the researcher is also the developer. Participants may react differently due to the designer's presence, and the designers may be, unintentionally or not, less receptive to critique. And even if the researcher collecting the data is not a developer, this kind of research in context can be fraught with challenges, such as these threats described by Krathwohl (1993): the Hawthorne effect (involvement in the study influences participant behavior); hypothesis guessing (participants try to guess what the researcher seeks and react accordingly); and diffusion (knowledge of the treatment influences other participants). In addition to triangulation, using unobtrusive data collection methods can mitigate some of these threats (Gray 2009). Alternatively, design researchers may choose to embrace their role as a 'nuisance variable,' and compensate for this by clearly describing their presence in the research setting and discussing their real or potential influence on the data.

*9: Analyze the data*
Guidance on processing, analyzing and interpreting both quantitative and qualitative data are available in many sources (cf. Creswell, 2009; Reeves & Hedberg, 2003). In the evaluation of interventions, the data analysis is often framed, directly or indirectly, by design propositions, if they have been woven into the research questions. That is, when the intervention is being evaluated, it is common to look for evidence of a certain construct in the intervention, its enactment, or its results. Data analysis culminates in drawing conclusions – answers to the main research questions. Very often, the next step for researchers is to decide how to address concerns in re-design, but preferably not before making time for reflection.

Reflection is both a solitary and a social process. Design researchers must afford themselves time for solitary reflection on the meaning and application of the results. In so doing, they must resist the temptation to jump to conclusions, especially when seemingly ambiguous findings (which are not at all infrequent, particularly when assessing reactions to intervention features), could easily be used to justify already preferred decisions. Indeed it is the people, not the data, that make decisions. To facilitate this process, a balance of solitary reflection and social reflection (e.g. critical peers are invited to play 'devil's advocate' positions) are helpful; reflection is discussed further in the following section.

*10: Report the study*
Studies documenting the inception, process and findings from intervention testing are almost always reported internally, in the form of reports or memos. They may also be reported externally, in the form of: conference presentations; articles in refereed journals or other periodicals; books; or publically-accessible reports. Both internal and external reports attend to both practical and theoretical goals.

Researchers often have very deep-seated notions about the nature of scientific inquiry, and can balk at reports that seem to conflict with what they deem important. In reporting educational design research, it can be useful to demonstrate commonalities, rather than highlight differences between what Kelly (2006) refers to as the commissive space of educational design research and that of other approaches in the social sciences. Kelly's notion of commissive space has to do with commitments to certain assumptions about, in this case, the nature of scientific research. Simply put, it could be summarized as one's perception of the 'rules of the game.' For example, the commissive space of those who view randomized controlled trials using only quantitative methods as the gold standard of all research and those interested in design research is likely to have relatively little common ground. This allows very little room for productive debate, and the result can rather quickly be flat out rejection, as soon as one member violates the 'rules of the game' or commitments to assumptions, within the other's commissive space. While design researchers violate many of the assumptions of the randomized field trials commissive space (Kelly, 2006), most share the same basic assumptions of, e.g. the mixed methods commissive space. By reporting sound educational design research in high-quality journals, the commissive space of design research can be made more transparent, exposed for both skeptics and supporters alike.

## Reflection processes

*About reflection*
Donald Schön, a master of musical improvisation and conceptual structure who was also trained as a philosopher, called for professionals to better understand their actions by reflecting on them in his highly acclaimed work, *The Reflective Practitioner: How Professionals Think in Action* (1983). It can be useful for design researchers to view themselves as reflective practitioners. Schön (1983, p. 68) claimed that in so doing, each of us:

> … allows himself to experience surprise, puzzlement, or confusion in a situation which he finds uncertain or unique. He reflects on the phenomenon before him, and on the prior understandings which have been implicit in his [designing] behavior. He carries out an experiment which serves to generate both a new understanding of the phenomenon and a change in the situation.

In educational design research, reflection involves active and thoughtful consideration of what has come together in both research and development (including theoretical inputs, empirical findings and subjective reactions) with the aim of producing new (theoretical) understanding. This is part and parcel of all long term, thoughtful research. Indeed, if we look at the scientific journals and books in our fields, it is clear that only some kinds of new knowledge grow forth directly from empirical testing. New theories, for example, do not present themselves in the form of research findings. Rather, they grow out of reflection. In scholarly publications, we typically share the products of reflection, and indeed some of the process, under the heading of *Discussion*.

*Fostering reflection*

Reflection is usually driven by reasoning. Through this fairly transparent, rational process, connection between existing ideas can lead to new ones. Reasoning is used, for example, to hypothesize cause and effect, to explicate assumptions or to justify predictions. Reasoning is essential for both research and design, but so are creative thoughts and feelings. Hammer and Reyman (2002) point out the need for inspiration and emotion to play a role in engineering design reflection, alongside rationality. New, useful insights may also be are born out of less transparent, less planned processes, whereby insights and novel ideas seem to present themselves, sometimes quite suddenly. Creativity researchers do not yet fully understand the processes through which these, seemingly spontaneous, connections are made. But more is known about conditions under which they tend to occur. In the remainder of this section, recommendations are given on what design researchers can do to nurture the more spontaneous connections (referred to as organic reflection) and to encourage the reasoned connections (referred to as structured reflection).

*Organic reflection*

The term, organic reflection, refers to a kind of intended contemplation. For many people, this kind of reflection takes place under the shower, or during a commute to/from work. Sometimes it stems from the process of dialogue. It takes place during times when there is very little agenda, and the mind is relatively free to wander and make its own connections between ideas. While organic reflection is not typically associated with professional work, it can certainly serve the work of design research. Three techniques that may fertilize organic reflection include:

- *Well-timed breaks, with input*: Look for a moment in the day where the work flow is naturally paused so as not to be overly disruptive (e.g. between articles, if the task is literature review), that can also afford a break away from the work place (take a shower or go for a walk). Use that break time, in silence or in dialogue, for reflection.
- *Seek **un**like-minded sparring partners*: Find people with ideas that are not just new, but foreign. In print, in dialogue, or in silence, explore the ways of knowing and lenses for viewing that are concomitant with those ideas.
- *Engage in 'background' projects:* Johnson (2010) discussed the value of maintaining background projects that can springboard new ideas. For example, innovative design projects, intriguing inquiry or literature study in new areas can trigger ideas, often by presenting new ways of looking at the world.

*Structured reflection*

Reyman et al. (2006) assert that design reflection should focus on two areas. Applied to the context of educational design research, these are: the design challenge (e.g. difference between existing and desired situations and/or important factors in the design context); and aspects of the integrated research and development process. Based on the steps in a basic design cycle and mechanisms of reflective practice, they define reflection as a process that consists of three main phases: *preparation, image forming* and *conclusion*.

Preparation and image-forming mainly look into the past. Preparation consists of collecting the relevant facts or observations to be considered. Image forming involves selection and synthesis of those facts and observations. Conclusion drawing looks ahead, using the results from the first two steps to determine next activities. Reymen et al. (2006) point out the importance of setting aside certain moments for reflection. In the case of educational design research, an obvious moment for reflection is between one micro-cycle of evaluation and reflection and another of design and construction. But, especially if those cycles are long, interim moments may need to be identified for reflection. Interim moments should both start and end with reflection on the design challenge and the design process, and they should also give attention to when the next reflection will be, and what that means for structuring the work in the meantime.

Another important point Reyman et al. (2006) make is that the reflection should occur on a regular basis. While their work is focused on individual designers, they also emphasize that design teams (with all their variation in personality, intuition, creativity, and other skills) should reflect together regularly, and that support for this is desirable.

In design research, structured reflection can be used (among others) to probe the data or revisit the methods. Table 1 below gives examples of how reflective processes can be shaped to develop understanding related to each of these themes. The plain text describes what to do; italicized text gives examples.

Table 1. *Strategies for structured reflection on educational design research*

| | **Preparation** | **Image forming** | **Conclusion drawing** |
|---|---|---|---|
| **Data** | Identify one or more data points from which, unplanned insight may be gleaned and ask a question. | Consider/discuss not the potential lesson to be learned, but think about the experience. Ask not only why questions, but also how and what. | Use the results to formulate new hypotheses, questions for investigation or revised design ideas. |
| | *Were there unanticipated processes through which the learners were highly engaged?* | *Why did the learners seem so engaged? What were they doing? When? How did they interact?* | *How might this reflection be put to use? Do design requirements need to be revised? Or design propositions?* |

| | | | |
|---|---|---|---|
| **Methods** | Identify the methods that have been used.<br><br>*The study used interviews, video observations, pre/post test data and analysis of learner assignments to explore pupil and teacher experiences with the new learning environment.* | Describe issues, questions or problems that have been ignored or insufficiently addressed by those methods; which ones were addressed well? What made that method work?<br><br>*Observation of inter-learner interactions proved less useful than hoped to understand individual reasoning pathways; this also yielded insufficient insight into learner processes to fully understand implications from learner assignments.* | What can be done differently? What (more) do we need to investigate in order to make improvements? What can be learned from what did yield 'eye-opening' or powerful findings?<br><br>*Instead of studying inter-learner interactions, conduct the video observations of learners while they are completing the assignments to try and understand the reasoning pathways.* |

Both the theoretical and practical goals of educational design research are furthered through reflection. Through organic and structured reflection, theoretical understanding is developed through consideration of the findings (e.g. "What do these results tell us?") in light of how they were obtained (e.g. the soundness of the measurements, or how appropriate the choice of setting). In addition, consideration is given to what might have been happening, but was not measured; or what could have occurred if the research had been structured differently. These, and other reflective insights, often have immediate implications for refining the intervention. For example, reflecting on how teachers altered a learning environment may generate new ideas about features to include or needs to be met. Often, reflection also generates new questions to be explored.

## Discussion

Many interim products are generated during evaluation and reflection, ranging from written theoretical frameworks, to research instruments and plans, to reports on intervention implementation and testing. Reflective musings may even be shared in written form. But toward furthering the main goals of design research, the main outputs from this phase are answers to research questions (theoretical understanding) and implications for the related design (practical applications).

Anyone picking up the challenge of educational design research must enter into it with eyes wide open and sleeves rolled up, aware of its enormous complexity. This unique research genre requires long-term in-depth collaboration among researchers, practitioners, and others. It requires a strong commitment to work that will often take years rather than months. However, the potential impact makes it all worthwhile. Desforges (2001) wrote "The status of research deemed educational would have to be judged, first in terms of its disciplined quality and secondly in terms of its impact. Poor discipline is no discipline. And excellent research without impact is not educational" (p. 2). The impact of educational research must increase, design research provides a promising way forward, and well-conceived and executed evaluation is a crucial element in the overall process.

Despite the widely-recognized failure of initiatives such as the "What Works Clearinghouse" initiated by the Department of Education in the USA (Schoenfeld, 2006), there are many who still insist that education be viewed as a science that can yield "proofs" akin to those found in the hard sciences such as physics and chemistry. We question the extent to which it is possible to prove

immutable laws regarding human actitivities such as teaching and learning. We also note that, even if possible, this alone will not advance the field. In the wake of increasing emphasis on evidence-based practice, descriptive theories are insufficient to solve complex problems (Green, 2000). We contend that education is essentially a design field more akin to architecture and engineering, and that research-based design knowledge is critical to the advancing the enterprise. Research that demonstrates how, why, with whom and under what conditions certain phenomena are engendered, prevented or otherwise manipulated, is urgently needed to improve education.

Considering how research is shaped relates to more than epistemic views. Design researchers tend to be associated with academic institutions. The sharing of knowledge, predominantly through written publications, is a cornerstone of academia and central to the mission of knowledge creation and dissemination. Given this, it can be helpful to note that in addition to epistemic views ( and of course, research financing), researcher survival concerns can also influence the choice to conduct design research in the first place (or not); they can also influence how activities such as evaluation and reflection are carried out. Specifically, priorities might be established first by what is deemed publishable, and second by what addresses real scientific and/or practical concerns.

The current international publication culture in the field of education powerfully privileges descriptive knowledge, and this is by all means extremely useful. However, because explanatory, predictive and normative theories are also needed to enable productive change, it would seem that the current publication culture is insufficiently aligned to the knowledge needs in educational practice. Thus, another challenge facing (design) researchers is that of cultivating a publications culture that values and encourages high-quality, well-justified design knowledge. Because leadership by example is often considered good practice, this special issue constitutes an important step in demonstrating the nature and value design knowledge publication.

## References


Bannan-Ritland, B., & Baek, J. Y. (2008). Investigating the act of design in design research: The road taken. *Handbook of design research methods in education: Innovations in science, technology, mathematics and engineering* (pp. 299-319)*. Mahway, NJ: Taylor & Francis*.

Burkhardt, H. (2006). From design research to large-scale impact: Engineering research in education. In J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational design research* (pp. 121–150). London: Routledge.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Desforges, C. (2001). *Familiar challenges and new approaches: Necessary advances in theory and methods in research on teaching and learning*. The Desmond Nuttall/ Carfax Memorial Lecture, British Educational Research Association (BERA) Annual Conference, Cardiff. Retrieved from http://www.tlrp.org/acadpub/Desforges2000a.pdf

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher 32*(1), 5–8.

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher, 22*(4), 16-23.

Gorard, S., Roberts, K., & Taylor, C. (2004). What kind of creature is a design experiment? *British Educational Research Journal, 30*(4), 577–590.

Gray, D. E. (2009). *Doing research in the real world*. London: Sage.

Green, J. (2000). The role of theory in evidence-based health promotion practice. *Health Education Research, 15(*2), 125-129.

Hammer, D. K. & Reymen, I. M. M. J. (2002). *The role of emotion in design reflection.* Paper presented at the 3rd International Conference on Design and Emotion, Loughborough, UK. Retrieved from http://doc.utwente.nl/58076/1/role_of_emotion.pdf

Johnson, S. (2010). *Where good ideas come from: The natural history of innovation.* New York: Riverhead Books.

Jonassen, D., Cernusca, D., & Ionas, G. (2007). Constructivism and instructional design: The emergence of the learning sciences and design research. *Trends and Issues in Instructional Design and Technology*, *2*, 45-52.

Kessels, J., & Plomp, T. (1999). A systematic and relational approach to obtaining curriculum consistency in corporate education. *Journal of Curriculum Studies*, *31*(6), 679-709.

Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach*. New York: Longman.

McKenney, S., Nieveen, N., & van den Akker, J. (2006). Design research from the curriculum perspective. In J. van den Akker, K. Gravemeijer, S. E. McKenney & N. Nieveen (Eds.), *Educational design research* (pp. 67–90). London: Routledge.

McKenney, S. E., & Reeves, T. C. (2012). *Conducting educational design research*. New York: Routledge.

McKenney, S. & Reeves, T. C. (2013). Cite other chapter with pd in model here – was that jonassen or aect?

Nieveen, N. (1999). Prototyping to reach product quality. In J. van den Akker, R. Branch, K. Gustafson, N. Nieveen & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 125–136). Dordrecht, NL: Kluwer Academic Publishers.

Plomp, T. (1982). *Onderwijskundige technologie: Enige verkenningen* [Educational technology: Several explorations]. Enschede, the Netherlands, University of Twente, Faculty of Educational Science and Technology.

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

Reymen, I. M. M. J., Hammer, D. K., Kroes, P. A., Aken van, J. E., Dorst, C. H., Bax, M. F. T., & Basten, T. (2006). A domain-independent descriptive design model and its application to structured reflection on design processes. *Research in Engineering Design*, *16*(4), 147-173.

Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher*, *35*(2), 13-21.

Schoenfeld, A. H. (2009b). Instructional research and the improvement of practice. In J. D. Bransford, D. J. Stipek, N. J. Vye, L. M. Gomez & D. Lam (Eds.), *The role of research in educational improvement* (pp. 161–188). Cambridge, MA: Harvard Education Press.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (Vol. Part II, pp. 19-64). Chicago: The University of Chicago Press.

Shavelson, R., Phillips, D. C., Towne, L. & Feuer, M. (2003). On the science of education design studies. *Educational Researcher*, *32*(25), 25-28.

Tessmer, M., & Harris, D. (1990). Beyond instructional effectiveness: Key environmental decisions for instructional designers as change agents. *Educational Technology*, *30*(7), 16-20.

Verhagen, P. (2000). *Inaugurale rede: Over het opleiden van onderwijskundig ontwerpers [Inauguration speech: About educating educational designers]*. Enschede: Universiteit Twente.

Wang, F. & Hannafin, M. (2005). Design-based research and technology-enhances learning environments. *Educational Technology Research and Development*, 53(4), 5–23.

Wiggins, G. (1989). A true test. *Phi Delta Kappan*, *70*(9), 703-713.