All Faculty Publications

2018

Reconsidering Design and Evaluation

Andrew S. Gibbons III
Brigham Young University, andy_gibbons@byu.edu

David D. Williams
david_williams@byu.edu

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub

Part of the Educational Psychology Commons, and the Educational Technology Commons

Reconsidering Design and Evaluation

Andrew S. Gibbons
David D. Williams
Brigham Young University

Abstract

In this paper we have put design and evaluation as it were under a microscope with multiple lenses. We use four different snapshots of design, each at a different level of resolution, to reveal new perspectives on the design-evaluation relationship. We believe that the disparate views of designers and evaluators can be resolved by resorting to the middle ground described by Klir. We believe also that the disparate views of educational technologists and learning scientists can be similarly resolved by appealing to the similar principle of Edelson.

Everything we evaluate is designed. Every evaluation we conduct is designed. Every report, graph, or figure we present is designed. In our profession, design and evaluation are woven together to support the same purpose—making the world a better place.

> President John Gargani
> American Evaluation Association
> Announcing the Annual Conference theme for 2016

Design and Evaluation Emerge as Formalisms

Instructional design and educational evaluation have emerged within the last century as formalized systems pertinent to the creation of education and training. The practices of instructional design and educational evaluation should be symbiotically related. But they have become academic silos, which has contributed to a lack of progress of theory and practice in both areas. Design practice is the main loser in this separation of interests. We propose that a more detailed examination reveals the natural value of this relationship, which may lead to advances in our understanding of design. In turn this may lead to advances in the teaching of design and evaluation.

Instructional design as a formal practice began to emerge in the early 1900s. Interest in systematic processes originated in public and education (see Gibbons, 2015) and accelerated during both World Wars, joining with psychological, technological, and educational movements that coalesced beginning in the mid-1950s (Finn, 1953; Gibbons, 2014; Saettler, 1968). Following the Second World War, new ideas from systems development added new substance and detail to formal instructional design processes

(Gagne, 1965), and the field of instructional design began to stabilize as a profession.

Evaluation as a field of formal study and practice formed and grew much more rapidly, beginning in the mid-1950s. Stimulus for rapid growth came from a new emphasis on accountability occasioned by both a sense of crisis in public education and rapid increases in expenditures on large social programs in the 1960s (McLaughlin & Phillips, 1991; Jardini, 2000).

Evaluation has two embodiments in practice: (a) it is a process of examining the details of a product, process, or phenomenon in order to judge its value, and (b) it is a process carried out during and after instructional design for the purpose of quality control, administration, and product improvement. Evaluation is distinct from and may incorporate performance assessment. Evaluation as a design-related practice originated from three different starting points: (a) the introduction of systems design processes developed during the war (Hughes & Hughes, 2000), (b) an empirical approach to the design of programmed instruction that employed a cybernetic build-test-evaluate-revise strategy to improve program effectiveness (Markle, 1969), and (c) educational research into product and program effectiveness (Tyler, 1969; McLaughlin& Phillips, 1991).

The Need for Unifying Theory

Design and evaluation grew up in a climate of expediency, necessity, and intense competition (see, for example, McLaughlin & Phillips, 1991). In both fields the creation of theory trailed practice rather than leading it. In both fields intellectual concepts grew by hypothesizing, rather than by derivation from core principles. Validation of competing approaches, though often recommended in both design and evaluation literatures, was seldom accomplished. Traditional research methods were not suited to the task of verifying technical models for solving highly contexted and unique problems. Commercial interests of contractors and laboratories often found it more profitable in terms of funding to capitalize on the invention of new terms and approaches than to engage in argumentation from basic principles. Though leading practitioners and theorists in both fields spoke out for empirical verification, change was so rapid that it seldom occured.

In this climate of entrepreneurship, and without a firm grounding, terminology in both fields suffered. Concepts in both fields lost precision rather than gaining it, as differences in usage occurred in the literature. Some terms began to refer to different things, and sometimes different terms were used to refer to the same thing.

The Broken Symmetry of Evaluation and Design

Where once in the literature descriptions of design and evaluation referred to each other (Flanagan, 1969; Bloom, Hastings, & Madaus, 1971), the symmetry of the design-evaluation relationship was broken, and opportunities for joint innovation, theory building, and practical user guidance were lost. Today, though instructional designers speak a language of evaluation, it is in relatively shallow and formulaic terms. Few instructional design graduate programs offer courses dedicated to evaluation design, and formal evaluation literature tends to dismiss interest in evaluation as it is carried out by designers (Scriven, 1991).

Several factors can be identified as contributors to this separation of interests. Among these are: (a) a rapid escalation of evaluation economics compared to design in the 1970s, (b)

heavier emphasis on theory development and use of theory in the evaluation field, (c) the superficiality of descriptions of evaluation by instructional designers, and (d) a resulting stagnation of the design field's view of evaluation. These and other factors created an imbalance in the mutual contributions these fields should have been making to each other. They also allowed descriptions of both design and evaluation to become increasingly trivialized in the mind of the instructional designer, stultifying further inquiry into the nature of design, and blocking potential insights and contributions from evaluation.

The Rapid Escalation of the Evaluation Economy

The launching of Sputnik in 1957 added impetus to the re-examination of the American educational system. Concern for its status and other social programs reached critical levels in the 1960s (Jardini, 2000). As pressures for reform increased, evaluating the status and impact of programs became a priority, and well-funded evaluation projects were launched. In this period of relative prosperity, evaluation became an industry. Theories of evaluation multiplied as the complexities of carrying out evaluations became more evident. During the same period, though considerable demand for designed curricula also increased, funding tended to go to non-designers and brand-name psychological theorists, and so did not contribute many insights into design practice.

The focus of evaluators during this period was not on the details of product or program design but on the results they produced. There was a distinctly scientific character to the manner in which evaluation studies were carried out. The interest of instructional designers in evaluation stemmed mainly from the need to quality control products under design, and the value of evaluations performed by designers was either ignored or deprecated by the scientific-minded professional evaluator.

These factors set up a pattern of the isolation of design-related evaluation and its eventual trivialization to being a somewhat adjunctive and optional activity to instructional design. The attention of the evaluation theorist was on diagnosis and causation at much higher levels of concern than the details of product effectiveness. The literature of scientific evaluation slowly became less relevant to the concerns of the designer, so the designer was left to gather out just those concepts relevant to designing, which over time became fewer. Designers began to produce their own practical literature on evaluation and over time became less interested in formal evaluation literature. There was little money available for serious scholarship on evaluation and its deeper relationship to design.

Emphasis on Theory Development in Evaluation

The development of formal evaluation theory expanded, as the complexities associated with evaluations of diverse kinds became apparent. Theory development for large-scale educational assessment increased as well, and so standardized testing also grew into an industry.

A further distraction for evaluators away from design concerns was the migration of evaluation toward social programs for fighting poverty and promoting social change. These competed for evaluation services as well. Increasingly, evaluation became a major professional field, built on competing theories associated with prominent names, perspectives, and philosophies.

Superficiality of Designers' Descriptions of Evaluation

The attention of designers became focused on the development of instructional design models (Andrews & Goodson, 1980; Gustafson, 1981; Gustafson & Branch, 1997, 2002; Gustafson & Powell, 1991). Components of design models became a major competitive issue, and though evaluation was unfailingly included and design models, it was a secondary and less prominent concern. It was not examined as closely as processes for design, development, and media concerns.

Lacking attention from professional evaluators, designers began to create their own literature on evaluation. However, this literature also tended to separate itself from the literature of design and development, so evaluation continued to be considered as a topic held at arm's length from design development. Though some evaluators addressed instructional issues, their literature normally did not try to examine evaluation in relation to design (Worthen & Sanders, 1987; Patton, 2011).

The concerns of designers writing about evaluation were practical rather than theoretical: pragmatic rather than philosophical. At the same time the trend in design models became less exploratory and more prescriptive. The desire of the military to create large numbers of in-house designers led to the publication of the Interservice Procedures for Instructional Systems Development (Branson, 1975), which was intended to simplify and standardize design and development processes to the point where otherwise untrained instructional designers could apply a systematic approach to produce usable instructional material. Design and development procedures had to be simplified in order to address this audience. Evaluation procedures had to be likewise simplified and related directly to development processes for quality control purposes.

From the mid-1950s on, the field of educational technology was responding to a call from Finn for professionalization (Finn, 1953). This led to the publication of a series of attempts to define the professional field (Association for Educational Communications and Technology, 1972, 1977; Ely, 1963, 1973, 1983; Finn, 1953, 1960; Seels & Richie, 1994; see also Januszewski, 2001 for a review of these attempts). These publications laid emphasis on design and development activities and on technology, placing less emphasis on the details of evaluation. Richey, Klein, and Tracey (2011) focus on the instructional design knowledge base but devote attention mainly to the theory of instruction at the expense of evaluation practice or theory. There are some exceptions to the trend of evaluation away from design (Hamilton & Feldman, 2014), but they are few.

Stagnation of Evaluation in a New Design Literature

Today instructional design literature, evaluation receives little more than cursory attention. Design models themselves are questioned, but no clear new description of design has emerged to take their place. Instead, there is a search for a set of concepts out of which a new description of design can be created. Discussion is beginning to focus on design thinking (Boling & Smith, 2008; Dorst, 2015; Lawson, 1997; Lawson & Dorst, 2009), design imagination (Kelley & Littman, 2001), the social role of the designer (Campbell, Schwier & Kenny, 2005; Hokanson, 2009; Kelley & Littman, 2005; Schwier, Campbell & Kenny, 2004), design as narrative (Parrish, 2006, 2008), and design architecture (Gibbons, 2014; Gibbons & Rogers, 2009). In these discussions evaluation has continued to receive scant attention. One of the purposes of this paper is to renew interest in evaluation and its necessarily close

association with the activities of design and development.

Design: Four Definitions and Their Impact

Visualizing the close relationship between evaluation and design depends on how design is described (Gibbons & Rogers, 2009; Gibbons & Yanchar, 2010). We have identified four approaches to describing design that illustrate this:

- Design as a process
- Design as a cybernetic activity
- Design as a logical exercise
- Design beyond logic

Design as A Process

Instructional design is most often described as a process consisting of steps and stages. Some form of process model is employed in virtually every design field, regardless of the content of the field, so instructional design process models are an instance of a larger class of design descriptions. Design process models have many uses: (a) they provide waypoints for marking progress through a design and development project, (b) they provide a general order for design activities, showing which stages must take precedence, (c) they provide security for novice designers who have no other guide, helping them to avoid inadvertently omitting important steps, (d) they facilitate resource and personnel scheduling, and (e) they show convergence points were independent strands of design activity flow together and become integrated.

Design process models are a boon to managers, who must account for schedules, personnel allocation, resource use, and deliverable products. Design process model steps are usually associated imperatives, such as standards for process, quality, and documentation. They also relate to staffing patterns. The downside of process models is that because they are geared to generic standards, they give less guidance for projects that do not fit a standard mold. For this reason, design models are less prone to produce *innovative* design solutions.

Evaluation's role in design process models is to provide quality control at each process step and for each deliverable. Judgments usually comprise product accuracy, effectiveness, acceptance, and accomplishment of initial project goals (budget, time frame, etc.). Sub-assemblies of a product may be subjected to testing by themselves, in intermediate assemblies, and as a whole product. This may occur during design, during roll-out, during early implementation, and continuously throughout the lifetime of the product.

This description of design process and its relationship to evaluation is generic and could easily be applied to software design, hardware or device design, or the design (and development) of a bridge. The evaluand is the product, its components, and the integrated product's ability to function as planned within its context.

Industrial process models were originally adopted by instructional designers for good reasons. During and after World War II the design of complex engineered systems and the design of training for users, maintainers, and managers had to take place in parallel to ensure that the system and its operator were ready to function together as soon as possible.

As instructional designers witnessed the effectiveness and efficiency of process models for engineers, they began to adopt them for their own work. Evidence of this adoption into the instructional design world can be seen in the work of Gagne (1965) and Briggs (Briggs, 1967, 1970) and later in the Interservice Procedures for Instructional Systems Development (Branson, 1975). As in the industrial models, the close coupling of evaluation with instructional design models was an unquestioned necessity.

However, this relationship has become somewhat loose. Though design must still be carried out, under pressures of time and money evaluation processes for low-stakes projects can be, and often is, given lower priority.

Among designers in industry, evaluation is devalued except for high-stakes projects and military and government contracts that specify evaluation standards. Much of evaluation today in commercial and corporate practice is conducted using "smile sheets" that remove the threat of embarrassing failure for corporate project managers and sponsors.

Design as A Cybernetic Activity

Originally, a primary theme of instructional design process models was a cybernetic cycle between design acts and evaluation acts. The most visible application of the cybernetic principle in instructional design emerged in the 1960s, advocated by Markle (1964, 1967) for the proving of programmed instruction products. Each frame of a program under development might become an evaluand, as well as the program as a whole. As instructional programming was adopted by a wider population of designers, time and resource pressures became more confining, and the close cybernetic link between design and evaluation was in most cases lost: designers began to follow formulas rather than conducting frequent product tests. The quality of programs declined, and eventually designers and clients lost interest in a demanding design process that required too many resources.

As design models trended upward and programmed instruction trended downward, interest in the cybernetic principle, applied at a more gross level, became one of the selling points for the acceptance of design models. Box models showed arrows recycling back to the high-level boxes of an ID process model, and the evaluand began to have larger scope. As instructional goals were created, they were subjected to reviews and checks to ensure completeness, accuracy, and proper form. As lessons of instruction were created, they were tested and revised until they supported instructional goals, but there was little evaluation below these levels. This corrective cybernetic closure was written into most design models, usually for each product element and process stage of the design. As sub-assemblies of the product were completed, they were sometimes subjected to testing, followed by whole-product testing. This cybernetic principle can be easily detected in the earlier instructional design textbooks (Banathy, 1968; Dick & Carey, 1976; Gagné & Briggs, 1979; Romiszowski, 1981; Smith & Ragan, 1999). But over time, it became apparent that many customers were not prepared to pay for evaluations at a fine-grained level. In general, it also eschewed evaluations that might be embarrassing. Design theorists found it easier to isolate evaluation as a core topic in design, and there has been very little new research on the relationship with design.

The cybernetic principle is still used today, but mostly in design models from outside the instructional design field that including agile design, rapid-cycle design, and prototype-test methods (Kelley & Littman, 2001; Krippendorff, 2007; Schrage, 1999; Scott et al, 2007;

Tripp & Bichelmeyer, 1990). In design models of this type, which are popular in the software development field, the design and development processes are cycled rapidly to produce a testable product. After testing, a design is modified according to the result, then it is produced in modified form and is tested again. This cycle continues until an acceptable product architecture and featuring emerges. In design systems of this type evaluation is elevated in importance and given the role of informing the design and providing data for its improvement. In this kind of design environment, evaluation cannot be treated as a secondary process because the design cannot move forward without it. Adoption of these models from the outside is reintroducing the concept of a finer-grained evaluation in some ID models (Allen, 2012).

Design Thinking and Design as A Logical Exercise

Design process can be described at multiple levels. Over the past decade there has been heightened interest in describing design as a reasoning process. One approach has been to analyze it at the level of logic and reason. Rittel (1987/1988) addresses design at this level:

> *Many forms of mental activity take place in the course of design.... A very significant part of design happens under conscious intellectual control. Since design is intentional, purposive, [and] goal-seeking, it decisively relies on reasoning. (p. 2)*

Among instructional design theorists, there is evidence of a change in focus away from high-level process descriptions of design and toward more detailed descriptions of how designs are formed. Rittel proposes that in this perspective, "the designer's reasoning appears as a process of *argumentation*" (p. 3, emphasis in the original). He proposes that "the process appears as one of formation of judgment, alternating with the search for ideas" (p. 5). In this view, evaluative judgment becomes quite frequent and takes place at many levels of design decision-making: at high levels related to gross features, and at the smallest level of detail.

Schön (1987) describes an oscillation that takes place in the reasoning of the designer between ideation and evaluation prior to firming a design commitment. What is notable is the length of time that a decision can remain tentative, before being made firm. It is also notable that even once a firm decision is reached, it can be undone and re-made later. Schön describes the design thinking of an architectural tutor named Quist during a problem solving session with a student:

> *Quist spins out a web of moves, subjecting each cluster of moves to multiple evaluations drawn from his repertoire of design domains. As he does so, he shifts from embracing freedom of choice to accepting implications, from involvement in the local units to a distanced consideration of the resulting whole, and from a stance of tentative exploration to one of commitment. He discovers in the situation's back talk a whole new idea, which generates a system of implications for further moves (p. 64).*

For Schön, "designing takes the form of a reflective conversation with the situation" (p. 56):

> *Each move [tentative decision] has consequences described and* evaluated *in terms drawn from one or more design domains. Each has implications binding on later moves. Each creates new problems to be described and solved. Quist [the designer] designs by spinning out a web of moves, consequences, implications, appreciations, and further moves (p. 57, emphasis added).*

Schön's design "conversation" is involves back-and-forth movement between active *decision-making* and *evaluation* of the impact of new decisions on an emergent design that consists of nothing more than prior decisions.

At this detailed level of description of design, the relationship between design and evaluation becomes readily apparent, as in the description by Schön above. What is surprising, however, is the degree to which evaluation becomes prominent. For example Schön's account makes it clear that before any firm commitment is made, there is considerable back-and-forth movement between hypothesizing a decision and then stepping back from it to reconcile it with other decisions and possible influence on future ones. Even once a decision is made firm, it can be unmade, and the process of designing can proceed forward from there again.

Dorst (2015) gives an even more detailed account of design, which he calls "frame design". It shows clearly the major role of evaluation in decision-making. Although his description is illustrated with applications at a macro level, its parallels with Schön's description at the micro level are informative.

A frame, according to Dorst, is a pattern of relationships between design elements that a designer can select or create (much like Schön's "cluster of moves") and then be used to explore solution options. A frame is an abstraction with parametric positions that can be filled by substitution with different "hows". The point of Dorst's frame concept, and the origin of the name, is that the process of framing is one step that can lead to workable, practical, and sustainable design decisions. It is significant that the many of the steps in this framing process are exploratory or evaluative, rather than committal.

Dorst's process of frame creation includes:

- Archaeology—An in-depth investigation of the apparent problem, as well as previous attempts at solutions
- Paradox—Identification of the "core paradox": the combination of forces has kept a solution from being found
- Context—Exploration of the practices, and the values, of key stakeholders of a solution
- Field—Definition of the "playing field" or "force field" were cultural, economic, social, and symbolic assets are exchanged between players [stakeholders]
- Themes—Identification of "the deeper factors that underlie the needs, motivations, and experience of the players": the basis of the assets exchanged
- Frames—Finding the set of themes shared by stakeholders and using them to create frames
- Futures—Reshaping proposed themes and frames through a process of negotiation to fit the requirements of all stakeholders
- Transformation—A critical evaluation of the short-term and long-term prospects of the proposed themes and frames
- Integration—A final negotiation *to produce a decision* acceptable to all interests

At a societal or organizational level these steps are intuitively sound. What is not as apparent is the degree to which they can be applied also to decisions about individual

features or clusters of features for a specific design for a specific learning experience. What changes is not the sequence of considerations, but the fact that it can be an individual designer carrying out the evaluation and decision processes, sometimes alone.

The narrative below shows the plausibility of application at the detailed level of individual frame innovation decisions:

- Archaeology—Most instructional designs are created to solve a problem previously unsolved. Previous attempts to solve the problem may have missed the mark, the problem itself may have been mis-identified, or the problem may not yet have been identified as a problem. In any case, pausing to search the true nature of the problem has value. Why have previous designs failed, or why can the desired outcome not be achieved? Very minor differences in design impact the effectiveness of designs (McDonald, 2006 ; McDonald & Gibbons, 2009).
- Paradox—An unsolved problem, whether at the macroscopic or microscopic level, presents an undiscovered and yet discoverable paradox: a positioning of forces that either through tradition or active opposition resists a solution. These forces can be thought of in tectonic terms. At the level of societal problems there are clearly problems locked in by opposing agendas. Is there not an analogous situation within the learner that can be addressed by understanding the forces at work? Can a higher degree of flow be achieved by knowing and avoiding what is opposing it?
- Context—Finding the existence of a paradox, a designer will work to reveal the traditions, habits, and motives that exert opposing forces holding the problem in place. The context explicitly addresses the interests of all of the stakeholders of a decision. In a learning design, these interests include not only those of the learner and the designer, but the producer, distributor, and delivery system for produced artifacts. In addition, there are the concerns of those supporting the learner (parents, community), and administrators (parents, school administrators). Each of these is capable of imposing a constraint, as well as being part of a solution for an individual learner.
- Field—The "currencies" exchanged in the learning place are often psychological where the individual learner is concerned. Currencies—perhaps best called "micro-currencies"—that should be given extra consideration by instructional designers are described in detail by Kahneman (2011) and Immordino-Yang (2015).
- Themes—Transactional currencies of instruction lead to themes that provide the basis for building up frames that can lead to solutions. The work of instructional researchers is to seek for themes that provide clues to design solutions through traditional and design-based research. Identifying themes leads a designer the vast body of research results and instructional theory.
- Frames, Futures—From themes should flow a set of frames that can be reconciled among stakeholders, providing a tool for the negotiation of varied interests. For the individual designer, the identification of frames and the negotiation of futures takes place within the designer's personal reasoning space and represents the personal judgments of the designer. For design teams, the formation of frames and their reconciliation takes place in the group's discussions in a manner described by Bucciarelli (1994).
- Transformation—Finding workable frames and futures is, finally, the beginning of a *commitment* process. Frames and futures represent tentative commitments that can be negotiated in the manner described by Schön earlier. Transformation represents

a further step toward the commitment in which final "sanity checks" can be performed in the terms used by all of the stakeholders.

- Integration—Final adjustments lead to firm commitments: decisions made.

Two things are striking about Dorst's description of the framing process: (a) that it describes a process that is a plausible description of design decision making at different levels of scale, and (b) that it is made up mainly of *exploratory or evaluative* acts rather than *committal* ones. This speaks directly to the main theme of this paper, which now becomes clearly discernible at this level of design description: that design and evaluation are symbiotically, inextricably related and that evaluation decisions are equally if not more crucial to design success than final decisions about features and properties. At this level of detail the term evaluation achieves a deeper meaning for the designer. Seeing evaluation from this perspective should perhaps result in a tectonic shift in the amount of attention designers should pay to each decision within a frame.

Design Beyond Logic

The importance of acknowledging the key role of evaluation during design does not end with a recognition of its *logical* implications: there is an *illogical* or *emotional* aspect as well. Design is often described in terms that go beyond rational, intellectual, and objective (Boling et al, 2015; Kim & Ryu, 2014; Norman, 2005).

Miller, Galanter, and Pribram (1960, hereafter MGP), propose that decision-making is an agentive process and that the influences brought to bear on a decision can be rational, but that they may also include emotional and tacit, or unconscious, influences.

In their book, *Plans and the Structure of Behavior*, MGP argue in favor of a basic unit of human behavior capable of subsuming the reflex arc described in the literature of behaviorism. They maintain that this basic unit must feature an expression of will, or choice, and that human behavior is not merely a reflexive response to the environment.

Therefore, as MGP make their argument, they revive the concept of agency: "something to bridge the gap from knowledge to action" (p. 11). They propose as the basic unit of human performance a new unit called the TOTE (test-operate-test-exit). The TOTE for them represents a basic unit of human action.

The TOTE unit has to be scalable. That is, it has to be a self-similar structure at all levels:

> *A proper description of behavior must be made on all levels simultaneously. That is to say, we are trying to describe a process that is organized on several different levels, and the pattern of units at one level can be indicated only by giving the units at the next higher, or more molar, level of description (p. 13).*

The TOTE is a cybernetic element consisting of "test-operate-test-exit". According the MGP, "the problem is to describe how actions are controlled by an organism's internal representation of the universe" (p. 12). Figure 1 portrays the TOTE unit as envisioned by MGP.
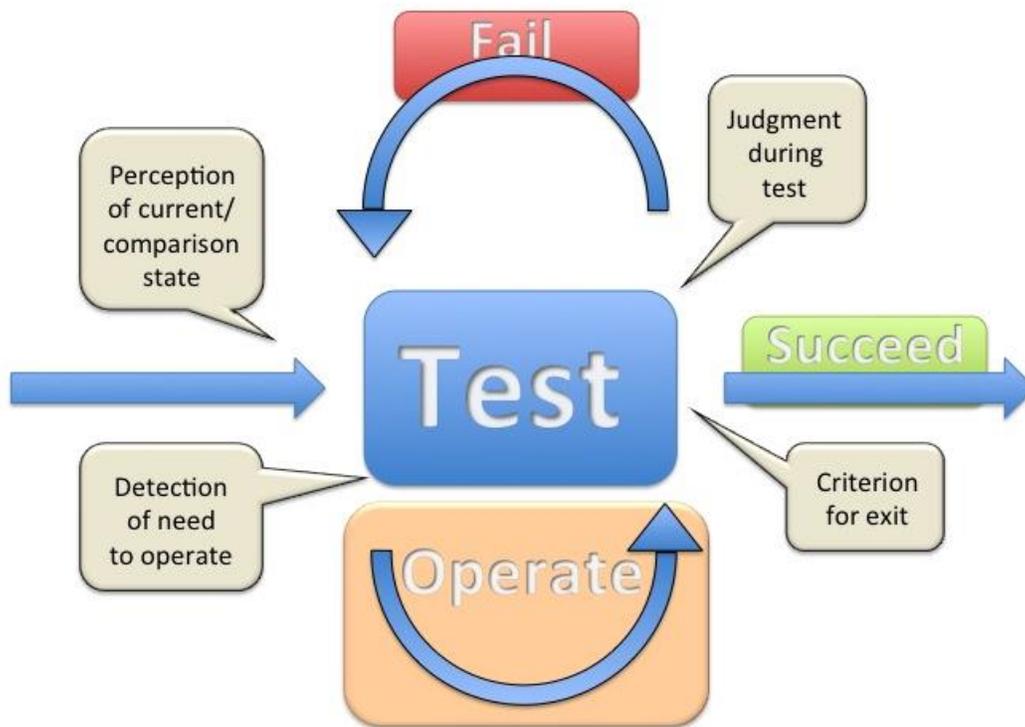
Figure 1. The TOTE unit depiction (after Miller, Galanter, and Pribram (1960)).

This portrayal indicates an ongoing test of "congruity" (in the internal or the external environment) that draws attention when an expected state is detected ("incongruity"). An incongruity leads to the selection and execution of a known operation. This is followed by a subsequent test to determine the success of the operation. Though the test and the operation are depicted as separate functions, it is clear that they must proceed in parallel in very close alternation. This brings evaluation and decision-making, alternating with action into consideration at a most basic level of performance.

MGP speculate on what is represented by the arrows in their diagram. They consider the answer at three levels of abstraction: energy, information, or control. Certainly the brain employs electrical energy in its operations. But at a level of higher abstraction, pulses of energy represent information. At the next level of abstraction, these pulses can be interpreted as control pulses that invoke or neutralize the execution of operations.

> *Here we are not concerned with a flow of energy or transmission of information from one [memory] page number to the next but merely with the order in which the "instructions" are executed. At this abstract level of description we are no longer required to think of the test is a simple [reflex arc] threshold that some stimulus energy must exceed. The test phase can be regarded as any process for determining that the operational phase is appropriate. (p. 29)*

In order to account for all human behavior, the test element of the TOTE unit must involve

both deliberate, rational, intellectual judgments (tests) and action choices (operations), and those below the level of conscious and deliberate control. This can mean that choices made by a designer can involve "hunches" or unexplainable biases in the designer's thinking. This gives validation to the idea that even though a designer cannot verbalize a rationale for a design decision, it may still have a reasonable basis based on the designer's previous experience.

> In its weakest form, the TOTE asserts simply that the operation [including decisions] an organism performs are constantly guided by the outcomes of various tests…. When the TOTE has been executed—the operations performed, the test satisfied, and the exit made—the creature may indeed appear to have attained a more desirable state (p. 29-30).

The concepts of agency and intentionality are a critical part of the MGP argument. MGP deconstruct the concept of motivation, describing it in terms of value and intention:

> The present authors take the position that a motive is comprised of two independent parts: value and intention. A value represents refers to an Image [of the world held in the individual's mind, capitalization in the original], whereas an intention refers to a Plan [capital in the original]. Presumably, a normal adult human being has constant access to a tremendous variety of plans that he might execute if he chose to do so…. As long as he is not using them, these various available plans form no part of his intentions. But as soon as the order to execute a particular plan is given, he begins to speak of its incomplete parts…. As things he intends to do. (p. ???)

The value of the TOTE to the present argument is that it opens for discussion a vision of human choice-making and design intimately associated through evaluation with all of the factors that influence human choices, including those that are emotional, unconscious, or tacit. Figure 2 illustrates a sampling of the factors that are brought to bear on judging (testing) and choosing operations (operating).
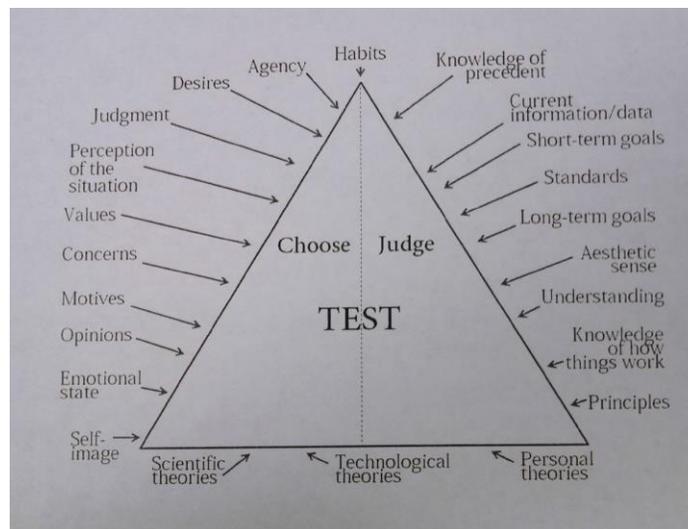

 Figure 2. A sampling of factors, deliberate and tacit, that influence judging (testing) and choosing.

At this level of the design and evaluation relationship, it becomes impossible to see how one of them can be considered in the absence of the other. Eisner's (1976) artistic appreciative approach to evaluation, dismissed by many designers as irrelevant, becomes more of a usable tool. At this level, evaluation may take as its evaluand not just one, but a succession of emotional states as well as the usual succession of intellectual states.

The view of the design-evaluation relationship at this level makes it hard to understand how design can adequately be taught without parallel instruction in evaluation and its close relation to decision-making. This becomes more important in light of the emerging recognition of the impact of emotions on learning (Immordino-Yang, 2015).

Conclusion

In this paper we have put design and evaluation as it were under a microscope with multiple lenses. We are not trying to explain design authoritatively, but rather to use four different snapshots of design, each at a different level of resolution, to reveal new perspectives on the design-evaluation relationship. We have observed that evaluative activity emerges larger and larger in the designer's field of view with each new level of detail. The process we usually perceive as *design* appears to be a fusion of *design and evaluation* processes. The more detailed the level of inspection, the more prominent judgmental and evaluative aspects appear to be. This presents a new aspect of design to us. Ironically, it runs counter to our observation that the formal worlds of design and evaluation have drifted apart.

What might the implications be of a more prominent role of evaluation in terms of current design and evaluation doctrine? Training for new instructional designers tends to emphasize decision-making aspects of design, describing evaluation as a follow-up activity, *after* decisions are made, rather than a process of applying judgment *before* deciding. This finding, should it be confirmed through further analysis, should lead to a curriculum for new designers that gives greater weight to the evaluative aspects of design—design thinking, design logic, and most importantly, design judgment. This will require not only giving more detailed descriptions of design activity, but in the process teasing out the role of value-seeking and value-assigning in the moment of design. It implies that we teach design thinking and design judgment rather than design process.

Relating Educational Technology and the Learning Sciences Through Design and Evaluation

If there can be a redefinition of these two essential activities, then it may also represent a point on the path toward *rapprochement* between disciplinary areas like the learning sciences and educational technology that now consider themselves in isolation. These areas have been addressing questions of learning and designing beginning with different assumptions about the nature of what they are studying and how it can be studied.

Edelson (2002) describes an alternative view of research by designing, in which:

> *…Design researchers proceed through iterative cycles of design and implementation using each implementation as an opportunity to collect data to inform subsequent design. Through a parallel and retrospective process of reflection upon the design and its outcomes the design researchers elaborate upon their initial hypotheses and principles,*

refining, adding, and discarding—gradually knitting together a coherent theory that reflects their understanding of the design experience. (p. 106)

This approach, which is widely recognized as design research, or design-based research, studies the processes of learning and designing simultaneously: teaching the researcher at the same time about designing, the elements of designs, and the elements of learning experiences. Edelson notes that: "an important characteristic of design research is that it eliminates the boundary between design and research" (p. 107).

It may be that "the boundary between design and research" in peoples' minds has been a major factor separating the interests and communities of the learning sciences and educational technology. The learning sciences, by adopting the term "science" in their identity, subscribe to the metaphors of science and all of the assumptions and associations they carry. Educational technologists carry on their backs the assumptions and associations of the term "technology". We submit that neither the learning sciences nor educational technology subscribe solely and exclusively to their nominal metaphors.

One way of understanding the divide that, unfortunately, many accept between the interests of these communities is to appeal to a more disinterested definition of the reasoning processes each school of thought relies upon.

George Klir (1969), a general systems theorist, describes the underlying continuity between scientific and design reasoning that may be a key not only to the design-evaluation relationship but to the learning sciences educational technology relationship as well. Klir proposes that the common goal of scientists and designers is the study of the properties of systems: "in our experimental [scientific] investigation of objects," says Klir, "we concentrate on…the search for a simple expression of the time-invariant relation[ships] between…quantities." He continues:

> *In the engineering [design] branches, the system [being studied] has the same traits as in the experimental sciences. As a rule, however, the problems involved are different. The relations between its quantities are usually prescribed [or constrained in some way], and we are to find a suitable manner for implementing them…or, conversely, a distinct realization [outcome] is given and we are to find the relations between certain quantities [to achieve it]. (p. 39-40)*

For Klir, the systems being studied are exactly the same, but scientists and engineers (designers) are solving different problems with respect to them. In general, scientists hold certain system quantities fixed while allowing others to vary so that unknown relationships within the system can be studied. Independent variables are controlled to determine their relationships to dependent variables. In design, what would be to the scientist a dependent variable—the outcome—becomes fixed, and the designer's job is to determine what relationships among the scientist's "independent" variables can achieve it.

Given Klir's description, the examples in the previous section of the reasoning behavior of Quist the designer and Einstein the scientist seem very similar. Both involve making a tentative decision regarding the resolution of forces represented by variables. This is followed by evaluation of the fitness of the decision within its context. The scientist's decision is about what variables to control using designed experimental materials and an experimental design. The engineer's decision is about what variables to involve and how to

balance their influence. It is interesting to note that in order to conduct an experiment, the scientist must prepare (design) the experimental materials, the tools of observation, and the measures of outcomes. It is likewise interesting that the designer has no information about the utility of a design without subjecting it to a quasi-experimental trial.

Making decisions based on judgment, followed by the evaluation of the decisions by trial, lies at the heart of both processes. This not only unifies our concepts of science and engineering, but it also shows the closeness of the design-evaluation relationship, even down to the finest level of detail.

Edelson concludes:

> *Taken together, these three arguments present a view of educational research as an applied science that differs from the view of research that is held by the broad community of educational researchers and has been taught in graduate schools of education in the past. Changing that view may be the key to enabling educational research to play a larger role in educational reform in the future. (p. 119)*

We believe that the disparate views of designers and evaluators can be resolved by resorting to the middle ground described by Edelson. We believe also that the disparate views of educational technologists and learning scientists can be similarly resolved by appealing to the same principle.

References

Allen, M. (2012). *Leaving ADDIE for SAM: An agile model for developing the best learning experiences*. New York: ASTD Press.

Andrews, D. H., & Goodson, L. A. (1980). A comparative analysis of models of instructional design. *Journal of Instructional Development*, *3*(4), 2-16.

Association for Educational Communications and Technology (1972). The field of educational technology: A statement of definition. *Audiovisual instruction*, *17*(8), 36-43.

Association for Educational Communications and Technology (1977). *The definition of educational technology*. Washington, DC: Association for Educational Communications and Technology.

Banathy, B. H. (1968). *Instructional systems*. Belmont, CA: Fearon.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (Eds.)(1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill Book Company.

Boling, E., & Smith, K. M. (2008). Artifacts as tools in the design process. In J. M. Spector, M. D. Merrill, J. Elen, and M. J. Bishop (Eds.), *Handbook of research on educational communications and technology.* New York: Springer.

Boling, E., Gray, C., Dagli, C., Demiral-Uzan, M., Ergulec, F., Tan, V., Altuwaijri, A., Gyabak, K., Hilligoss, M., Kizilboga, R., & Tomita, K. (2015) Judgment and instructional design: How ID practitioners work in practice. *Performance Improvement Quarterly*, *28*(3), 25-49.

Branson, R. K., Rayner, G. T., Cox, J. L., Furman, J.P., King, F.J., & Hannum, W. J. (1975). *Interservice procedures for instructional systems development (5 vols.)* (TRADOC Pamphlet 350-30). Ft. Monroe, Va.: U.S. Army Training and Doctrine Command, August, 1975. (NTIS Nos. AD-A019 4860-AD-A019 490).

Briggs, L. J. (1967). *Instructional media: A procedure for the design of multi-media instruction, a critical review of research, and suggestions for future research*. Washington, DC: American Institutes for Research.

Briggs, L. J. (1970). *Handbook of procedures for the design of instruction.* Washington, DC: American Institutes for Research.

Bucciarelli, L. L. (1994). *Designing engineers.* Cambridge, MA: The MIT Press.

Campbell, K. Schwier, R. A., & Kenny, R. F. (2005). Agency of the instructional designer: Moral coherence and transformative social practice. *Australasian journal of educational technology*, *20*(1), 69-100.

Dick, W. & Carey, L. M. (1976). *The systematic design of instruction (1st ed.)*. Instructional Design and Development Department, College of Education, Florida State University, Tallahassee, FL.

Dorst, K. (2015). *Frame innovation: Create new thinking by design. Cambridge, MA: The MIT Press.*

Edelson, D. C. (200). Design research: What we learn when we engage in design. *Journal of the learning sciences*, *11*(1), 105-121.

Eisner, E. W. (1976). Educational connoisseurship and educational criticism Their form and functions in educational evaluation. *The Journal of Aesthetic Education, 10*. 135-150.

Ely, D. P. (1963). The changing role of the audiovisual process: A definition and glossary of related terms. TCP Monograph No. 1, *Audiovisual communication review*, 11(1), supplement No. 6.

Ely D. P. (1973). Defining the field of educational technology. *Audiovisual instruction*, *8*(3), 52-53.

Ely, D. P. (1983). The definition of educational technology: An emerging stability. *Educational considerations*, *10*(2), 2-4.

Finn, J. D. (1953). Professionalizing the audio-visual field. *Audio-visual communication review*, *1*(1), 6-17.

Finn, J. D. (1960). A new theory for instructional technology. *Audiovisual communication review*, *8*(1), 84-94.

Flanagan, J. C. (1969). The uses of educational evaluation in the development of programs, courses, instructional materials and equipment, instructional and learning procedures, and administrative arrangements. In R. W. Tyler (Ed.) (1969). *Educational evaluation: New roles, new methods: Sixty-eighth yearbook of the National Society for the Study of Education, Part II* (pp. 221-241). Chicago, IL: University of Chicago Press.

Gagné, R. M. (Ed.)(1965). *Psychological principles in system development*. New York: Holt Rinehart & Winston.

Gagné, R. M., & Briggs, L. J. (1979). *Principles of instructional design* (2nd ed.). New York: Holt, Rinehart and Winston.

Gibbons, A. S. (2014). *An architectural approach to instructional design*. New York: Routledge.

Gibbons, A. S. (2015). Instructional design models. In J. M. Spector (Ed.), The SAGE *encyclopedia of educational technology* (pp. 386-389). Thousand Oaks, CA: Sage Publications.

Gibbons, A. S. & Yanchar, S. C. (2010). An alternative view of the instructional design process: A response to Smith and Boling. *Educational technology*, 50(4), 16-26.

Gibbons, A. S., Boling, E., Smith, K. M. (2014). Instructional design models. *Handbook of research on educational communications technology* (4th ed.). New York: Springer.

Gustafson, K. L. (1981). *Survey of instructional development models*. Syracuse, NY: ERIC Clearinghouse on Information & Technology.

Gustafson, K. L., & Powell, G. C. (1991). *Survey of instructional development models with an annotated ERIC bibliography* (2nd ed.). Syracuse, NY: ERIC Clearinghouse on Information & Technology,

Gustafson, K. L., & Branch, R. M. (1997). *Survey of instructional development models* (3rd ed.). Syracuse, NY: ERIC Clearinghouse on Information & Technology.

Gustafson, K. L., & Branch, R. M. (2002). *Survey of instructional development models* (4th ed.). Syracuse, NY: ERIC Clearinghouse on Information & Technology.

Hokanson, B., & Miller, C. (2009). Role-based design: A contemporary framework for innovation and creativity in instructional design. *Educational Technology*, *49*(2), 21–28.

Hughes, A. C., & Hughes, T. P. (2000). *Systems, experts, and computers: The systems approach in management and engineering, World War II and after.* Cambridge, MA: The MIT Press.

Immordino-Yang, M. H. (2015). *Emotions, learning, and the brain*. New York: Norton.

Januszewski, (2001). *Educational technology: The development of a concept.* Englewood, CO: Libraries Unlimited.

Jardini, D. (2000). Out of the blue yonder: The transfer of systems thinking from the Pentagon to the Great Society, 1961-1965. In A. Hughes and T. Hughes (Eds.), *Systems, experts, and computers: The systems approach in management and engineering, World War II and after*. Cambridge, MA: The MIT Press.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kelley, T., & Littman, J. (2001). *The art of innovation.* New York: Crown Business.

Kelley, T., & Littman, J. (2005). *The ten faces of innovation.* New York: Currency Doubleday.

Kim, J. & Ryu, H. (2014). A Design Thinking Rationality Framework: Framing and Solving Design Problems in Early Concept Generation. *Human–computer interaction*, *29*(5-6), 516-553. DOI: 10.1080/07370024.2014.896706.

Klir, G. J. (1972). *An approach to general systems theory*. New York: Van Nostrand Reinhold.

Krippendorff, K. (2007). Cybernetics of design and the design of cybernetics. *Kybernetes*, *36*(9/10), 1362-1380.

Lawson, B. (1997). *How designers think: The design process demystified (3rd ed.)*. Oxford, UK: Architectural Press.

Lawson, B., & Dorst, K. (2009). *Design expertise*. Oxford, UK: Elsevier.

Markle, S. M. M. (1964). Teaching machines and programmed instruction: The Harvard teaching machine project: The first hundred days. *Audiovisual communication review, 12*(3), 344-351.

Markle, S. M. M. (1967). Empirical testing of programs. In P. C. Lange (Ed.), *Programmed instruction: Sixty–sixth yearbook of the National Society for the Study of Education: Part II* (pp. 104–138). Chicago: University of Chicago Press.

Markle, S. M.M. (1969). *Good frames and bad: A grammar of frame writing (2nd ed.)*. New York: John Wiley & Sons.

McDonald, J. K. (2006). Technology I, II, and III: Criteria for understanding and improving the practice of instructional technology. Unpublished Thesis (Ph.D.), Brigham Young University, Department of Instructional Psychology and Technology.

McDonald, J., & Gibbons, A. (2009). Technology I, II, and III: Criteria for understanding and improving the practice of instructional technology. *Educational technology research and development, 57*, 377-392.

McLaughlin, M. W., & Phillips, D. C. (Eds.) (1991). *Evaluation and education: At quarter century: Ninetieth yearbook of the National Society for the Study of Education, Part II*.

Chicago, IL: University of Chicago Press.

Miller, G., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Henry Holt & Company.

Norman, D. (2005). *Emotional design: Why we love (or hate) everyday things.* New York: Basic Books.

Parrish, P. (2006). Design as storytelling. *TechTrends*, *50*(4), 72-82.

Parrish, P. (2008). Plotting a learning experience. In L. Botturi and T. Stubbs (Eds.), *Handbook of visual languages in instructional design* (pp. 91-111). Hershey, PA: IDEA Group.

Patton M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use.* New York: Guilford Press.

Richey, R. C., Klein, J. D., & Tracey, M. W. (2011). *The instructional design knowledge base: Theory, research and practice.* New York, NY: Routledge

Rittel, H. W. J. (1987/1988). The reasoning of designers. Paper presented at the Congress on Planning and Design Theory, Boston, August, 1987, and at the *Schrieftenreihe des Instuts fuer Grundlagen der Planung*, *Universitaet Stuttgart*, 1988.

Romiszowski, A. J. (1981). *Designing instructional systems: Decision making in course planning and curriculum design.* New York: Kogan Page/Nichols Publishing.

Saettler, P. (1968). *A history of instructional technology*. New York: McGraw-Hill.

Schrage, M. (1999). *Serious games: How the best companies simulate to innovate.* Cambridge, MA: Harvard Business School Press.

Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco, CA: Jossey-Bass Publishers.

Scott, B., Shurville, S., Maclean, P. & Cong, C. (2007). Cybernetic principles for learning design. *Kybernetes, 36*(9/10), 1497-1514.

Scriven, M. (1991). Beyond formative and summative evaluation. In M.W. McLaughlin and D. C. Phillips (Eds.), *Evaluation and education: At quarter century: Ninetieth yearbook of the National Society for the Study of Education, Part II*. Chicago, IL: University of Chicago Press.

Seels, B. B., & Richey, R. C. (1994). *Instructional technology: The definition and domains of the field.* Bloomington, IN:: Association for Educational Communications and Technology.

Smith, P. L., & Ragan, T. J. (1999). *Instructional design (2nd ed.).* Upper Saddle River, NJ: Merrill.

Tripp, S., & Bichelmeyer, B. (1990). Rapid Prototyping: An Alternative Instructional Design Strategy. *Educational technology research and development*, *38*(1), 31-44.

Tyler, R. W. (1969). *Educational evaluation: New roles, new methods: Sixty-eighth yearbook of the National Society for the Study of Education, Part II*. Chicago, IL: University of Chicago Press.

Worthen, B. B., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines.* White Plains, NY: Longman.