Two views of scale: Design Principles for Scaling Reach and Empowerment

Chinmay Kulkarni

Carnegie Mellon University

Pittsburgh, PA, USA

chinmayk@cs.cmu.edu

ABSTRACT

How do we design technology for learning at scale? Based on an examination of a large number of influential systems for learning at scale, I argue that designing for scale is not an amorphous design undertaking. Instead, it builds on two distinct perspectives on scale. Systems built with a *scaling through efficiency* perspective make learning more efficient and allow the same number of instructors to help a much larger set of learners. Systems with a *scaling through empowerment* perspective empower a larger number of people to assist learners effectively. I outline how these simple differences in design perspective lead to large differences in design concerns, techniques, and evaluation criteria. Articulating prevalent design perspectives should make overlooked design opportunities more salient, help systems designers design for scale more deliberately and understand their trade-offs, and open up new opportunities to designers who shift their perspectives.

CCS CONCEPTS

• **Human-centered computing** → HCI theory, concepts and models.

KEYWORDS

design perspectives, systems for scale, empowerment, efficiency

1 INTRODUCTION

Systems for learning at scale have enormous diversity in how they work. For example, intelligent tutoring systems scale learning by creating a computational model of human learning and allow learners individualized tutoring at low cost [39]. MOOCs have no explicit computational learning model, but instructors scale their reach by recording their lessons over video for thousands to watch later, and by incorporating many active learning techniques [40]. Amortizing the large effort of creating high-quality instruction over thousands of learners allows instructors efficiently scale teaching. Other systems help many learners watch a video lecture together [49] or to form better teams [73]. And of course, some systems combine the benefits from approaches such as MOOCs and intelligent tutoring systems [2].

These individual systems have introduced powerful techniques to scale learning. But less attention has been devoted to understanding underlying patterns across individual systems, and to constructing general principles for designing systems that scale learning.

In this paper, I examine influential systems in the Learning@Scale and adjacent communities to unearth some of these patterns. The goal is not to create a taxonomy of systems for learning at scale (others have created a partial one already [43]) but rather to offer system designers guidance as they design new systems. I synthesize this analysis into two design perspectives that seem to influence, often implicitly, the design of systems for learning at scale.

A design perspective consists of beliefs, assumptions, and preferred visions of the future held by a designer. Design perspectives highlight designers' intentions and suggest particular design techniques and decisions, and frame how these decisions are evaluated [15]. As *perspectives*, they are not right or wrong, nor are specific perspectives always better than others. Over time, designers may also shift their perspective.

By discussing design perspectives explicitly, I hope to shift the discussion from whether we can design systems for learning at scale to what perspective should we bring to designing

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
L@S '19, June 24–25, 2019, Chicago, IL, USA

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6804-9/19/06...\$15.00

<https://doi.org/10.1145/3330430.3333620>

“at scale”, and how best to achieve it. Most papers about learning at scale systems so far talk about scale and its benefits broadly. For instance, creators of PeerStudio suggest that “scale of massive online classes enables systems that drastically and reliably reduce the time to obtain feedback” [46]. Similarly, Divide and Conquer, a system for clustering student submissions has the goal of “maintaining the teacher’s involvement while still allowing them to work with short answer responses at scale” [10]. Despite this vagueness about what scaling means, my analysis suggests that system designers actually think about scale from two distinct design perspectives.

The first perspective, which I call *scaling through efficiency*, scales learning by improving how *efficiently* instructors can reach and teach learners. The total number of instructional resources (e.g. teachers) remains roughly the same, but each instructor can efficiently teach many more learners. For example, systems in this vein allow instructors to teach more students by recording their classes, or to be more effective in noticing common misconceptions across thousands of learners by aggregating data.

The second perspective, which I call *scaling through empowerment* scales learning by allowing more *people* (not necessarily instructors) to help in a learner’s education. Systems embodying this perspective result in scaffolding that significantly increases the size and the diversity of the pool of their effective instructional resources (e.g. by including peers, parents, learners themselves etc.)

Intelligent tutoring systems (ITSs) offer an example of *scaling through efficiency*. One of the initial visions for these systems was to create a computational tutor as thoughtful and as caring as an expert human tutor [63, 65]. Broadly speaking, ITSs did so by computationally capturing and replicating the same pedagogical and tutorial strategies used by expert human tutors [54]. Expert human tutors would thus become dramatically more efficient when their tutoring models were replicated and distributed at scale (i.e. *scaling through efficiency*). This perspective has been generalized to a variety of skills (see [1, 39]) and has had sustained success over many decades [59].

Scaling through empowerment systems empower people other than traditional experts to help in a learner’s education. For instance, Mindset Tools is a system that empowers teachers, instead of only trained psychologists, to foster growth mindsets in their classroom by subtly changing their teaching practices through lesson plans, teacher stories, and discussions [70]. These systems are not limited to a “train-the-trainer” model either: for example, many peer feedback systems empower peer reviewers to write better feedback without training them to do so (e.g. with in-context computational aids [55].)

These design perspectives, and the differences between them may seem obvious, even trivial. However, these perspectives pervade all aspects of design: what systems are concerned with, the techniques used to address those concerns, and how systems are evaluated.

Articulating distinct design perspectives in detail has three benefits. First, it gives system designers a guide while designing new systems. When the benefits, drawbacks, and the emergent effects of distinct design perspectives is well articulated, designers can choose a perspective more deliberately.

Second, systems that employ the same design perspective could borrow design techniques from each other, creating a systematic way to guide designers. For example, MOOCs, which like ITSs also employ the *scaling through efficiency* perspective, are being enriched by techniques introduced in ITSs, such as building computational models of learning [2], and personalizing instruction through analysis of data.

Finally, system designers can deliberately change their design perspective (which is currently often implicit) and discover new ways to look at persistent problems. For example, early work on MOOCs including our own took a *scaling through efficiency* perspective to assessment. In this perspective, open-ended assessments were seen as a challenge as they could not be efficiently graded by expert teachers nor automated yet [6]. Peer assessment was seen as a practical alternative: teachers couldn’t scale to a MOOC classroom of thousands, so peers could be recruited to grade work instead (e.g. see [45]). Our later work and that of others recognized that peer feedback provided benefits that instructor feedback did not, and looked at *scaling through empowerment*: empowering peers with better matching [46], emphasizing interdependence [44], and creating better “recipe-books” [55].

The main contribution of this paper is to point out that successful systems for learning at scale are not amorphous design efforts, but systematically build on specific design perspectives. Further, by systematically documenting these design perspectives, this paper creates guide-posts that should help future systems designers. Finally, system designers might of course bring perspectives other than *scaling through empowerment* and *scaling through efficiency* to their work. Articulating these two perspectives, their benefits, and their challenges might also encourage other systems designers to similarly discuss their perspectives explicitly.

2 RELATED WORK

This paper builds on three areas of related work: 1. Research on design perspectives frames my analysis, 2. existing arguments for specific design choices that suggest possible implicit perspectives and classifications of systems, and 3. a large number of existing systems provide the raw material from which I distill perspectives on scale.

Design perspectives: The idea of design perspectives guiding design is pervasive in design practice [15]. Perspectives comprise beliefs, assumptions, and methods that frame design activity, and determine what it focuses on. Perspectives are not always articulated, but are often developed through reflection [62] or through an examination of possibilities through prototyping [50]. This paper develops design perspectives primarily by reflecting on existing systems for learning at scale, and on my own involvement in designing them.

Because design perspectives frame design activity, shifts in perspectives have enormous ramifications. For example, *service design* is a perspective holds that “a framework of value creation is built on service provision, rather than goods production” [51]. This perspective reframes a transaction (such as selling a computer) from being a result of manufacturing a unit of output to conceptualizing it as a process. In this changed perspective, dynamic resources such as knowledge are as important as static ones like manufacturing plants, and the generation of value as the result of a collaborative process between providers and consumers, rather than mere production and delivery of a good.

Specifically, changes in perspective change which problems are solved e.g. a service perspective changes which products are developed [11], what design techniques are used e.g. service design leads to techniques for managing value flow [9], and how results are evaluated e.g. how value is measured [72].

In this paper, I articulate the problems different perspectives on scale emphasize, techniques used, and how systems are evaluated.

Classifying systems for learning and scale and current provocations: Baker implicitly advocates for designing computational tutors that are designed intelligently to *leverage human intelligence* [4]. Baker argues that leveraging human intelligence is useful because more automated interventions are time-consuming to program, unable to predict when they have failed, and do not adapt to changing learner behavior. I develop Baker’s thesis further, not only arguing whether systems should leverage humans, but *how* to do so. In addition, I explore the outcomes of such empowerment, and emergent effects.

Kross and Guo catalog learning at scale systems based on how many participants are involved in their interactions: one-on-one (many peer-interaction based systems), one-to-many (many grading aids), and many-to-many/automated (all other systems) [43]. This catalog offered a starting point for my analysis, and I augment it with other systems from adjacent communities. Unlike this prior work, my analysis focuses on what the interactions between participants *do* (e.g. make teaching more efficient or empower more individuals to scaffold learning).

Existing systems for learning at scale See Table 1 lists the systems I studied to discover implicit design perspectives. This includes all Learning@Scale conference papers identified by [43] as “Systems” contributions, but excludes systems that have no user interface (such as knowledge tracing algorithms), or those that do not concern learning directly (such as website analytics frameworks.) I also studied systems in adjacent conferences published after 2012 that contribute to learning at scale (found using keywords such as learning at scale, MOOC, etc.). This latter list is approximate, and is intended to sample how researchers in adjacent domains think about learning at scale.

3 INFERRING DESIGN PERSPECTIVES

My process of distilling design perspectives started with talking to fellow researchers about how they operationalized scale in their own work at the Learning@Scale 2017 Program Committee meeting, and the conference itself. Based on these initial observations and reflections on my own research, I identified a few candidate perspectives of scale that might describe how systems were designed (such as “scaling is increasing fanout”, and “scaling is ensuring transferability”). I refined these perspectives by listing the design techniques used in papers at Learning@Scale (Table 1), and clustering them by possible intent, until the perspectives so found could describe a majority of systems at the conference and systems in adjacent conferences.

The perspectives I articulate are necessarily subjective. They reflect both the systems studied and my own biases and reflections on the area. This subjectivity is essential to discovering interpretable insights, even if other researchers may synthesize systems into an different set of perspectives [80]. In other words, this contribution is inherently subjective, and this subjectivity is what allows it to fulfill its purpose of guiding future research.

4 DESIGN TECHNIQUES FOR SCALING THROUGH EFFICIENCY

The *scaling through efficiency* perspective seeks to allow experts to reach and help a large number of learners across contexts. The total number of experts involved remains largely the same as previously required at smaller scale. Instead, each expert is much more efficient in scaffolding learners.

Systems that scale through efficiency also seek to yield benefits for a large number of learners despite differences in learner contexts (i.e. to have a large “main-effect” for their intervention). Design techniques for *scaling through efficiency* reflect this focus.

Technique 1: Generalize small-scale interventions

Many systems for learning at scale start as small-scale interventions. Removing the details of the local context in which

AT LEARNING AT SCALE	
Axis: Generating explanations at scale with learnersourcing and machine learning [75]	Efficiency
Divide and correct: using clusters to grade short answers at scale [10]	Efficiency
Online Urbanism: Interest-based Subcultures as Drivers of Informal Learning in an Online Community [21]	Empowerment
PeerStudio: rapid peer feedback emphasizes revision and improves performance [46]	Empowerment
Towards adapting to learners at scale: integrating MOOC and intelligent tutoring frameworks [2]	Efficiency
Graders as Meta-Reviewers: Simultaneously Scaling and Improving Expert Evaluation for Large Online Classrooms [34]	Efficiency
Improving the Peer Assessment Experience on MOOC Platforms [68]	Empowerment
Scaling Expert Feedback: Two Case Studies [33]	Both
Scaling Short-answer Grading by Combining Peer Assessment with Algorithmic Scoring [47]	Efficiency
Self-evaluation in Advanced Power Searching and Mapping with Google MOOCs [74]	Unclear
Gradescope: A Fast, Flexible, and Fair System for Scalable Assessment of Handwritten Work [66]	Efficiency
Writing Reusable Code Feedback at Scale with Mixed-Initiative Program Synthesis [26]	Efficiency
AT OTHER VENUES (IJAIED, CHI, SIGCSE, LAK, CSCW, TOCHI, Science, Psychonomic Review, J. Ed. Psych).	
Building empathy, compassion, and achievement in the jigsaw classroom [3]	Empowerment
Closing global achievement gaps in MOOCs [38]	Efficiency
Codechella: Multi-user program visualizations for real-time tutoring and collaborative learning [24]	Efficiency
Codeopticon: Real-time, one-to-many human tutoring for computer programming [23]	Efficiency
Cognitive Tutor: Applied research in mathematics education	Efficiency
Data-driven interaction techniques for improving navigation of educational videos	Efficiency
Escaping the echo chamber: ideologically and geographically diverse discussions about politics [13]	Empowerment
Interactive Guidance Techniques for Improving Creative Feedback [55]	Empowerment
Juxtapeer:Comparative Peer Review Yields Higher Quality Feedback and Promotes Deeper Reflection [12]	Empowerment
Modelling human teaching tactics and strategies for tutoring systems [16]	Efficiency
Off-task behavior in the cognitive tutor classroom: when students game the system [5]	Efficiency
Online python tutor: embeddable web-based program visualization for cs education.	Empowerment
Pedagogical patterns: advice for educators [8]	Empowerment
Peer and self assessment in massive online classes [45]	Efficiency
Promoting learning mindsets in schools: Lessons from educators' engagement with the PERTS Mindset Kit [7]	Empowerment
Providing support for adaptive scripting in an on-line collaborative learning environment [25]	Empowerment
Responding to children's mathematical thinking in the moment: an emerging framework of teaching moves [32]	Empowerment
Stupid tutoring systems, intelligent humans [4]	Empowerment
Talkabout: Making distance matter with small groups in massive classes [44]	Empowerment
Teaching machines [67]	Efficiency
The ASSISTments Ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching [27]	Empowerment
The classroom as a dashboard: co-designing wearable cognitive augmentation for K-12 teachers [29]	Empowerment
Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor [1]	Efficiency
Towards academically productive talk supported by conversational agents [18]	Empowerment
Using Design Patterns for Math Preservice Teacher Education [31]	Empowerment
Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school [77]	Efficiency
Virtual teams in massive open online courses [73]	Empowerment
Watching MOOCs together: Investigating co-located MOOC study groups [49]	Empowerment

Table 1: Systems studied to develop design perspectives. The inferred design perspective embodied by each system is on right (Efficiency: *scaling through efficiency*, Empowerment: *scaling through empowerment*).

the intervention was first introduced allows designers to reduce the “heterogeneity of irrelevancies” [64] that are embedded in a contextual small-scale intervention, and scale it to other contexts. The design goal is generally not to find universal generality [77], but to transfer an intervention to a setting with “proximal similarity” [64].

For example, consider Yeager et al’s design of a website for fostering a growth mindset among school students at scale [77]. The research team had studied mindset interventions at a smaller scale in primary schools for decades (see [17]). To scale to a larger context (primary school and high-school, on a website that might serve thousands), the team conducted “Qualitative inquiry and rapid, iterative, randomized ‘A/B’ experiments with 3000 participants ... to inform intervention revisions for this [more general] population” [77].

Specifically, the qualitative inquiry allowed the researchers to hypothesize which aspects of their intervention were contextual and would need to be modified from primary schools to high school (such as replacing descriptions of studies on ants with studies on teenagers), and which ones could work across contexts (e.g. describing how intelligence is malleable, rather than simply prescribing that students have a growth mindset.)

The large-scale testing and iteration then helped refine the intervention that worked across contexts. As a result of this process, Yeager et al. focused their website on labeling and explaining benefits of the growth mindset (instead of alternative messages that refuted misinformation, or told students directly to have a growth mindset.)

In the case of mindset interventions, generalization also helped with improving “discriminant validity” [14], i.e. discriminating which constructs exactly are effective across populations. The ease with which scaling for efficiency yields generalizable and more exact scientific findings is a large benefit and the reason why some researchers prefer to design systems with this perspective.

Technique 2: Use models and summarized data

To efficiently scale learning, systems distill data from learner interactions into simplified models. For instance, to replicate the pedagogical and tutorial strategies used by expert human tutors, researchers built a model of a learner and of expert human tutors [54]. Subsequent research has further improved how closely computers can mimic the expert human tutor [16].

Early ITS models were grounded deeply in learning theory, but this is not necessarily the case for all *scaling through efficiency* systems. For example, Bayesian Knowledge Tracing is explicitly based on a learning model. However, other tracing models are based on Long Short-term Memory (LSTM), a

model originally designed for language translation. In practice, both approaches yield useful predictive abilities [53].

Furthermore, while an explicit model is particularly powerful for mimicking expert behavior, it is not essential to improving efficiency. For example, Kim et al created interaction techniques for improving navigation of educational videos by highlighting points where previous learners frequently paused or revisited [36]. While there is no explicit model of learner interest, data from learner behavior is distilled into a useful interaction.

A key consideration about whether derived data/models can be used to scale efficiency are how much data is required. For many systems, derived data can only yield benefits above a critical threshold (for example, see Discussion in [36].)

Technique 3: Select rather than scaffold crowdsourced contributions

Scaling through efficiency benefits from techniques that allow an instructor to select appropriate student contributions from hundreds of learners automatically or with minimal effort. For example, Williams et al introduce AXIS, a system that crowdsources instruction, specifically explanations from learners (e.g. for solutions to word problems), and uses machine learning to determine which explanations are most helpful to other learners [75]. AXIS demonstrates a way to efficiently improve instructional materials. The benefit of operating at large-scale is having a larger number of explanations to choose from. Several other systems have similarly leveraged learner-sourcing [35].

AXIS also highlights that as a design perspective *scaling through efficiency* emphasizes some techniques, but other design techniques may also be minimally present. For instance, AXIS scaffolds learners as they write explanations. However, this scaffolding is minimal. Learners are told to “imagine explaining to another learner, if the two of you were sitting at a computer working at it together.”

By contrast, scaffolding in a system that took the *scaling through empowerment* (where selection is not feasible) would likely use more elaborate scaffolding. For instance, conversational agents in peer tutoring use elaborate scaffolding strategies such as academically productive talk [18].

Technique 4: Exert Managerial Control and Oversight Efficiently

AXIS also shows how systems for *scaling through efficiency* must have strong “managerial control” [61] over instruction, such as controlling how and when explanations are shown. When systems scale through efficiency, they allow a small number of people to serve a large number of learners. In doing so, they must create ways to easily monitor the large number of learners (e.g. [23]) and guide them by efficiently controlling their activity and scaffolding it. AXIS focuses on

guidance through controlling when hints are shown; other efficient scaffolding mechanisms cluster student solutions and suggest corrective feedback [10].

Technique 5: Multiplex learner activities

Systems for *scaling through efficiency* can look at the learning process itself to find opportunities to use fallow time, or multiplex activities such that limited instructor attention can be directed where it is most effective. Codeopticon [23], built on the extremely popular Python Tutor [22] allows tutors to observe and help a large number of learners as they write and debug their code. Codeopticon takes advantage of natural pauses in tutoring interactions when a learner is following tutor’s instruction or discovering a solution, and directs tutor attention to other learners who need help. Similarly, orchestration tools can help teachers direct their attention around a classroom that is solving problems using an online tool (see [71] for examples).

Key considerations for multiplexing systems is to design them such that instructors or other experts can quickly immerse themselves in enough context to be helpful to learners, while minimizing their attention-switching and managing fatigue.

5 DESIGN TECHNIQUES FOR SCALING THROUGH IMPROVING EMPOWERMENT

The *scaling through empowerment* perspective seeks to increase the number of people who can help in a learners’ education, and scaffold them so that they can do so effectively. Unlike *scaling through efficiency* systems that target learners directly, these systems work through a large number of intermediaries (such as peers, mentors, and other non-experts).

Note that this design perspective is different from concrete models such as “train-the-trainer”, where learners are simultaneously taught a skill and how to train others in it [56]. While some systems with the *scaling through empowerment* perspective might train intermediaries, their focus is usually not on making intermediaries experts themselves.

Instead, *scaling through empowerment* systems focus on designing ongoing *computational scaffolding* to let intermediaries help learners at large scale. Furthermore, the distinction between learners and intermediaries is often blurred as in the case of reciprocal peer feedback and tutoring.

A key focus of these systems is to create technology and instructional strategies that a large diversity of participants can use easily. The design techniques such systems use reflect this focus.

Technique 1: Create guides and contextual aids

One common method of empowerment is to create contextual guides to scaffold instructional or interactional patterns.

For instance, Kotturi et al suggest creating “recipes” to help peers manage their online interaction. Bergin et al. suggest creating *pedagogical patterns* [8]. These patterns, may be second nature to seasoned educators who have used them so often may still be useful to beginning educators, peer mentors, and tutors [8]. The idea of *teaching moves* structures these patterns into pairs of environmental cues (such as a mentee struggling to ask the right question) and a productive response [32].

Despite all this success in codifying patterns into guides, it is difficult to discover effective instructional or interactional strategies in the first place [31]. Thus, there is an opportunity to use computational methods to improve this process. Fortune cookies [45] suggest how even a minimal data-driven approach might be fruitful. Kulkarni et al developed fortune cookies by looking at common errors learners make and writing feedback fragments that target those errors [45]. The CritiqueKit system takes this further, showing reusable feedback suggestions that dynamically adapt based on criteria a peer reviewer’s feedback is classified to currently lack [55].

Technique 2: Computationally match participants

When a large number of participants are available, *scaling through empowerment* can match participants with those who are best suited to scaffold learning. Systems have shown that temporal matching (who’s available right now?), matching based on past matching performance, or on other trace data (such as contents of forum posts) are all beneficial.

PeerStudio, a system for peer feedback matches learners to peer reviewers dynamically. It recruits reviewers with the objective of minimizing the time a learner must wait to have their work reviewed. This rapid feedback then improves learning and revision [46]. A similar system for dynamic matching is used by Codechella, which allows learners to write code together, collaboratively understand its execution using multiple mouse cursors, and chat with each other [24]. Similar to PeerStudio, it suggests matches by availability, but allows learners to choose among multiple potential matches. Matching does not have to be temporal. For instance, using the large amount of forum data allows Wen et al to form better learner teams for collaborative projects [73].

Technique 3: Emphasize interdependence

Systems for *scaling through empowerment* bring many more people into the technosocial system, and so they must also manage expectations of benefits and create norms of interdependence. Kotturi et al describe how a reminder message to learners participating in an online discussion system “saying that their peers were counting on them to show up to the discussion” improved how many participated [42]. Many peer feedback systems emphasize reciprocity [12, 46]. Emphasizing interdependence is not always straightforward,

especially when it must combat social loafing, such as when most community members do in fact contribute [41].

While the above examples of interdependence are all peer-based systems, similar interdependence may hold between teachers and parents, or mentors and peers, for instance.

Technique 4: Distribute managerial control and support contextual decisions

Systems built for *scaling through empowerment* are reliant on a large number of people engaging in productive behaviors to effectively target complex learning goals. Thus, in contrast to *scaling through efficiency* systems, the emphasis is not on exerting managerial control efficiently, but rather to distribute managerial control so that people can make contextual decisions.

The PERTS Mindset Kit project [7] offers one example of this strategy in action. Mindset interventions are generally planned by expert psychologists. The previous section described how such experts scaled such interventions efficiently, by creating a website. By contrast, the Mindset Kit project allows non-experts in psychology, i.e. school teachers, to foster a growth mindset in their own classrooms. Because teachers ultimately decide what to do in their classroom the project catalogues simple tweaks to classroom practice that foster growth mindsets. For example, instead of having different expectations for high and low students, teachers could set the expectation that all students will contribute to math.

Creating classroom norms is a complex contextual process, which can be seen as a *situated action* [69] that cannot be adequately scaffolded by a pre-created plan (Technique 1). Therefore, in addition to lesson plans, the Mindset Kit project (www.mindsetkit.org) offers aids such as teacher stories to help teachers to take actions in their specific contexts.

6 DESIGN PERSPECTIVES IN ACTION

Design perspectives affect all aspects of the design process, not only the techniques used. Below, I outline a few cases that demonstrate how design perspectives do so.

Implementation and personalization

Skinner’s *Teaching Machine* [67] individualized learning by automatically adapting instruction and practice opportunities. Similarly, in many *scaling through efficiency* systems, personalized learning is achieved through automated means, such as selecting the appropriate next problem or remedial instruction.

In contrast, personalization in *scaling through empowerment* systems generally occurs through human intermediaries. For example, a system supporting students working out problems in pairs could encourage a student to elaborate their own explanations or to ask their partner to contribute [25].

Challenges to Fidelity

The threat to effectiveness for a *scaling through efficiency* system is that it may not implement a sufficiently powerful and de-contextualized mechanism. For example, to scalably foster growth mindsets, researchers had to try a large number of intervention designs until they narrowed down to a set that was effective across contexts [76, 77].

In contrast, *scaling through empowerment* may suffer from a game of telephone. For instance, educators want to spread the word to their colleagues about fostering a growth mindset, but sometimes use metaphors that may be confusing or dilute the message. Making this attenuation worse, teachers downplay how they learned from their mistakes[7]. Therefore, a key concern while creating empowerment cookbooks is to prevent the effectiveness from being diluted.

Evaluation

Scaling through efficiency systems tend to focus on interventions that work similarly across contexts, and their scientific evaluation also tends to focus on the main-effect, i.e. how well they work in general. For instance, Kim et al., in analyzing the benefits of video lectures, highlight main effects in their paper abstract: “shorter videos are much more engaging, that informal talking-head videos are more engaging, that Khan-style tablet drawings are more engaging” [37].

In contrast, *scaling through empowerment* systems tend to focus on how context varies their effectiveness. In fact, this difference is often desirable. For instance, Gelman et al. highlight how interest-based subcultures emerged as membership of the Scratch online community grew over time [21].

7 DISCUSSION

The preceding sections suggest how design perspectives affect the design techniques available, and design concerns they emphasize. I now turn to how design perspectives might help us address larger questions around learning at scale.

Will it really scale?

Some believe that only completely automated systems can “really” scale [6, 48]. But of course, no system is completely automated: someone still has to design it, train it, turn it on, or maintain it. Therefore, automated systems are better described as systems that *scaling through efficiency*. And as we have seen, these represent only one of the possible design perspectives on scale.

It follows then that how well such systems scale depends largely on how much they can improve efficiency. Section 4 suggests the possible techniques to do so, and their corresponding challenges. One interesting corollary to *scaling through efficiency* is that systems that scale up easily may have trouble scaling *down*, to accommodate fewer users. For

example, Kim et al’s video interaction techniques [36] will yield useful data only above a threshold number of learners.

Similarly, how well systems for *scaling through empowerment* scale also depends on how they can use design techniques for empowerment (Section 5). For example, peer feedback systems have been limited by how well they can create recipes for writing good feedback [45]; recent systems that overcome these bottlenecks [55] may scale better.

But can empowerment systems *really* scale? Because some participants (such as peers) are not paid in money, it is tempting to wonder if their participation is fleeting, and will only last until the novelty of the experience wears off. Empowerment systems may also seem disruptive as they replace formerly paid work such as grading answers with the “free labor” such as peer feedback.

Well designed empowerment systems can address both these criticisms by better articulating the value flows [9]. For instance, while peers may not be paid in money, giving peer feedback improves confidence [19] and learning [12], and peers enjoy looking at their classmates work and find it inspiring [45]. *Scaling through empowerment* systems should design such value flows deliberately and emphasize interdependence (Section 5) to ensure that they really scale.

Opportunities for new computational tools

Here I highlight areas for further research. A/B testing has been effective at generalizing small-scale solutions to pinpoint one scalable solution, and machine learning models like *contextual multi-arm bandits* offer an opportunity to find multiple *different* solutions that work in different contexts instead [60]. Similarly, while some prior work investigates multiplexed learner activity, current techniques scale only to a few hundred learners at once [23]. With more simultaneous learners, multiplexing should become more predictable, and so larger scale is likely possible.

Similarly, empowerment systems rely on matching algorithms to connect participants. Better matching algorithms which consider more contextual factors may help [79]. Furthermore, it is likely existing big-data approaches for efficiently scaffolding learning in *scaling through efficiency* systems (such as next step recommendations) could be adapted to creating better guides and contextual aids. Finally, supporting situated action has remained remarkably immune to technological improvement, though recent techniques from CSCW [58] could perhaps be usefully adapted.

What about other perspectives?

In this paper, I focus on two perspectives for scale that are embodied a large number of research systems. This focus allows a detailed characterization of their design techniques and concerns. That said, other perspectives are also present. In particular, the *connectivist* perspective holds that learning

is not an individual activity, and that the goal of learning systems should be to build communities for learning [52]. While these perspectives are certainly important, the literature I studied does not offer enough guidance about how best to design for such a goal.

Limitations

The design perspectives I articulate are based on the systems I have studied. Because systems for learning at scale are published in many venues, the perspectives are likely skewed to work that appears in the Learning at Scale, CHI, CSCW, AIED, and LAK conferences (where most of the systems I studied were published). Also, as noted previously, because many systems speak about scaling learning broadly, and do not articulate an explicit design perspective, my inferences may not capture the designers’ original intents.

Further, this paper looks at perspectives of *scale*. However, systems for learning may also be designed with other benefits in mind, which I do not discuss here. *Scaling through efficiency* has increased access to education and at low cost [39], and can close global achievement gaps when designed well [38]. Similarly, *scaling through empowerment* has often been used for purposes other than simply learning skills, for example for empathy across political differences [13, 57].

Finally: a word about terminology. While “efficiency” and “empowerment” may seem value-laden terms, the intent is to use them descriptively. As design perspectives, I do not argue that there an inherently better way to scale. If particular systems designers (or readers) believe that one way to scale is “better” than another, I suggest this may well reflect an implicit design perspective.

8 CONCLUSION: DESIGNING WITH PERSPECTIVE

An anecdote (by way of Ken Holstein) illustrates how a deliberate perspective might lead designers to new insights. One of the schools Holstein and collaborators were studying used an intelligent tutoring system (ITS) for math during class time. Walking around the classroom, one teacher noticed a student had been flagged for “bottoming out” on hints and the ITS had decided to withhold further hints to prevent gaming [5]. But when the teacher started a conversation with the student, he learned the student wasn’t unmotivated, but heartbroken: his first love had broken up with him. This observation led Holstein to think about using ITSs as *teachers’ aides* [30]. This deliberate perspective then led to creating dashboards and augmented reality systems that empower teachers to start productive conversations [29].

Changing perspectives for persistent problems: Because design perspectives affect design decisions both explicitly and *implicitly*, relatively fewer systems borrow techniques from both perspectives. Shifting design perspective might help address persistent critiques of some systems

for learning at scale. For instance, critics see MOOCs as “an integral expression of a global logic” [20], with a predominantly Western control of knowledge [20]. Looking at MOOCs through a *scaling through empowerment* perspective instead might suggest some alternatives. For instance, could an online adaptation of the Jigsaw classroom method [3] create an online classroom that is more inclusive as well?

Similar benefits could be achieved taking systems which have typically taken a *scaling through empowerment* perspective and employing a *scaling through efficiency* perspective to their shortcomings. For instance, the ASSISTments project has a *scaling through empowerment* perspective, with the goal of creating an ecosystem which combines educational research and practice allows any teacher to create assessments with built in assistance [27]. But such an approach has its drawbacks. As Heffernan et al note “A single teacher may not be the most apt at explaining all topics to all [their] students” [28]. They then augment ASSISTments with PEERAssist, which adapts the learner-sourcing technique and applies it for teacher-sourcing, selecting appropriate instructional content personalized to each student from a set of submissions by teachers.

Overlooked techniques from the same perspective

Incorporating a different perspective may not be necessary. It may be sufficient to simply use an overlooked design technique from the same perspective. For example, preserving the current design perspective of MOOCs, de-contextualized psychological interventions for belonging have made these systems more inclusive [38]. Similarly, many empowerment systems have improved their initial design to emphasizing interdependence [42].

Applications beyond scaling Teacher professional development programs (TPDs) are methods that improve teachers’ instructional practice through ongoing training. While TPDs do not scale through empowerment, as they do not increase the number of people who help learners, several parallels exist. Given the generally poor improvements seen in TPD programs in practice (one meta-study found only nine of 1,300 studies analyzed to have a positive effect on student learning [78]), it is possible that techniques introduced in *scaling through empowerment* systems may help TPD applications as well.

Taken together, this paper suggests how design perspectives pervade the design of systems for learning at scale. In articulating them explicitly, I hope that as system designers, we ask ourselves not only whether we can design systems for learning at scale, but rather, what perspective should we bring to designing “at scale”, and how best to achieve it.

ACKNOWLEDGMENTS

I would like to thank Matt Easterday at Northwestern University for conversations that sparked this paper, my graduate

students Julia Cambre and Yasmine Kotturi for thoughtful suggestions on my manuscript, and anonymous reviewers whose critique and feedback improved this paper immeasurably. The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180277. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education.

REFERENCES

- [1] Vincent Aleven, Bruce McLaren, Ido Roll, and Kenneth Koedinger. 2006. Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education* 16, 2 (2006), 101–128.
- [2] Vincent Aleven, Jonathan Sewall, Juan Miguel Andres, Robert Sottolare, Rodney Long, and Ryan Baker. 2018. Towards adapting to learners at scale: integrating MOOC and intelligent tutoring frameworks. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. ACM, 14.
- [3] Elliot Aronson. 2002. Building empathy, compassion, and achievement in the jigsaw classroom. *Improving academic achievement: Impact of psychological factors on education* (2002), 209–225.
- [4] Ryan S Baker. 2016. Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education* 26, 2 (2016), 600–614.
- [5] Ryan Shaun Baker, Albert T Corbett, Kenneth R Koedinger, and Angela Z Wagner. 2004. Off-task behavior in the cognitive tutor classroom: when students game the system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 383–390.
- [6] Stephen P Balfour. 2013. Assessing Writing in MOOCs: Automated Essay Scoring and Calibrated Peer Review. *Research & Practice in Assessment* 8 (2013), 40–48.
- [7] J Beaubien, L Stahl, R Herter, and D Paunesku. 2016. Promoting learning mindsets in schools: Lessons from educators’s engagement with the PERTS Mindset Kit.
- [8] Joseph Bergin, Jutta Eckstein, Markus Volter, Marianna Sipo, Eugene Wallingford, Klaus Marquardt, Jane Chandler, Helen Sharp, and Mary Lynn Manns. 2012. *Pedagogical patterns: advice for educators*. Joseph Bergin Software Tools.
- [9] Andrea Bonaccorsi, Gionata Carmignani, and Francesco Zammori. 2011. Service value stream management (SVSM): developing lean thinking in the service industry. *Journal of Service Science and management* 4, 04 (2011), 428.
- [10] Michael Brooks, Sumit Basu, Charles Jacobs, and Lucy Vanderwende. 2014. Divide and correct: using clusters to grade short answers at scale. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 89–98.
- [11] J Cagan and CM Vogel. 2002. What drives new product development.
- [12] Julia Cambre, Scott Klemmer, and Chinmay Kulkarni. 2018. Juxtapeer: Comparative Peer Review Yields Higher Quality Feedback and Promotes Deeper Reflection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 294.
- [13] Julia Cambre, Scott R Klemmer, and Chinmay Kulkarni. 2017. Escaping the echo chamber: ideologically and geographically diverse discussions about politics. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2423–2428.
- [14] Thomas D Cook. 1990. *The generalization of causal connections: Multiple theories in search of clear practice*.
- [15] Kees Dorst. 2010. The nature of design thinking. In *Design thinking research symposium*. DAB Documents.

- [16] Benedict Du Boulay and Rosemary Luckin. 2001. Modelling human teaching tactics and strategies for tutoring systems. *International Journal of Artificial Intelligence in Education* 12, 3 (2001), 235–256.
- [17] C Dweck. 1999. Self-Theories: Their role in motivation, personality and development. 1999.
- [18] Gregory Dyke, Iris Howley, David Adamson, Rohit Kumar, and Carolyn Penstein Rosé. 2013. Towards academically productive talk supported by conversational agents. In *Productive multivocality in the analysis of group interactions*. Springer, 459–476.
- [19] Lauren Eskreis-Winkler, Ayelet Fishbach, and Angela L Duckworth. 2018. Dear Abby: Should I Give Advice or Receive It? *Psychological science* 29, 11 (2018), 1797–1806.
- [20] Rachel Flamenbaum, Manduhai Buyandelger, Greg Downey, Orin Starn, Catalina Laserna, Shreeharsh Kelkar, Carolyn Rouse, and Tom Looser. 2014. Anthropology in and of MOOCs. *American Anthropologist* 116, 4 (2014), 829–838.
- [21] Ben U Gelman, Chris Beckley, Aditya Johri, Carlotta Domeniconi, and Seungwon Yang. 2016. Online Urbanism: Interest-based Subcultures as Drivers of Informal Learning in an Online Community. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 21–30.
- [22] Philip J Guo. 2013. Online python tutor: embeddable web-based program visualization for cs education. In *Proceeding of the 44th ACM technical symposium on Computer science education*. ACM, 579–584.
- [23] Philip J Guo. 2015. Codeopticon: Real-time, one-to-many human tutoring for computer programming. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 599–608.
- [24] Philip J Guo, Jeffery White, and Renan Zanelatto. 2015. Codechella: Multi-user program visualizations for real-time tutoring and collaborative learning. In *Visual Languages and Human-Centric Computing (VL/HCC), 2015 IEEE Symposium on*. IEEE, 79–87.
- [25] Gahgene Gweon, Carolyn Rose, Regan Carey, and Zachary Zaiss. 2006. Providing support for adaptive scripting in an on-line collaborative learning environment. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 251–260.
- [26] Andrew Head, Elena Glassman, Gustavo Soares, Ryo Suzuki, Lucas Figueredo, Loris D’Antoni, and Björn Hartmann. 2017. Writing reusable code feedback at scale with mixed-initiative program synthesis. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*. ACM, 89–98.
- [27] Neil T Heffernan and Cristina Lindquist Heffernan. 2014. The ASSISTments Ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education* 24, 4 (2014), 470–497.
- [28] Neil T Heffernan, Korinn S Ostrow, Kim Kelly, Douglas Selent, Eric G Van Inwegen, Xiaolu Xiong, and Joseph Jay Williams. 2016. The Future of Adaptive Learning: Does the Crowd Hold the Key? *International Journal of Artificial Intelligence in Education* 26, 2 (2016), 615–644.
- [29] Kenneth Holstein, Gena Hong, Mera Tegene, Bruce M McLaren, and Vincent Aleven. 2018. The classroom as a dashboard: co-designing wearable cognitive augmentation for K-12 teachers. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. ACM, 79–88.
- [30] Kenneth Holstein, Bruce M McLaren, and Vincent Aleven. 2017. Intelligent tutors as teachers’ aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. ACM, 257–266.
- [31] Paul Salvador Inventado, Sharris Gayle Francisco Inventado, Noboru Matsuda, Yeping Li, Peter Scupelli, Korinn Ostrow, Neil Heffernan, Shihfen Tu, Craig Mason, Mary Logue, and Pat McGuire. 2018. Using Design Patterns for Math Preservice Teacher Education. In *Proceedings of the 23rd European Conference on Pattern Languages of Programs (EuroLoP ’18)*. ACM, New York, NY, USA, Article 31, 8 pages. <https://doi.org/10.1145/3282308.3282340>
- [32] Victoria R Jacobs and Susan B Empson. 2016. Responding to children’s mathematical thinking in the moment: an emerging framework of teaching moves. *ZDM* 48, 1-2 (2016), 185–197.
- [33] David A Joyner. 2017. Scaling Expert Feedback: Two Case Studies. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*. ACM, 71–80.
- [34] David A Joyner, Wade Ashby, Liam Irish, Yeeling Lam, Jacob Langston, Isabel Lupiani, Mike Lustig, Paige Pettoruto, Dana Sheahen, Angela Smiley, et al. 2016. Graders as Meta-Reviewers: Simultaneously Scaling and Improving Expert Evaluation for Large Online Classrooms. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 399–408.
- [35] Juho Kim. 2015. *Learnersourcing: improving learning with collective learner activity*. Ph.D. Dissertation. Massachusetts Institute of Technology.
- [36] Juho Kim, Philip J Guo, Carrie J Cai, Shang-Wen Daniel Li, Krzysztof Z Gajos, and Robert C Miller. 2014. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 563–572.
- [37] Juho Kim, Philip J Guo, Daniel T Seaton, Piotr Mitros, Krzysztof Z Gajos, and Robert C Miller. 2014. Understanding in-video dropouts and interaction peaks in online lecture videos. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 31–40.
- [38] René F Kizilcec, Andrew J Saltarelli, Justin Reich, and Geoffrey L Cohen. 2017. Closing global achievement gaps in MOOCs. *Science* 355, 6322 (2017), 251–252.
- [39] Kenneth R Koedinger, John R Anderson, William H Hadley, and Mary A Mark. 1997. Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education (IJAIED)* 8 (1997), 30–43.
- [40] Kenneth R Koedinger, Jihee Kim, Julianna Zhuxin Jia, Elizabeth A McLaughlin, and Norman L Bier. 2015. Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the second (2015) ACM conference on learning@ scale*. ACM, 111–120.
- [41] Yasmine Kotturi, Andrew Du, Scott Klemmer, and Chinmay Kulkarni. 2017. Long-Term peer reviewing effort is anti-reciprocal. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*. ACM, 279–282.
- [42] Yasmine Kotturi, Chinmay E Kulkarni, Michael S Bernstein, and Scott Klemmer. 2015. Structure and messaging techniques for online peer learning systems that increase stickiness. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 31–38.
- [43] Sean Kross and Philip J Guo. 2018. Students, systems, and interactions: synthesizing the first four years of learning@ scale and charting the future. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. ACM, 2.
- [44] Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S Bernstein, and Scott R Klemmer. 2015. Talkabout: Making distance matter with small groups in massive classes. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1116–1128.
- [45] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R Klemmer. 2013. Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 6 (2013), 33.
- [46] Chinmay E Kulkarni, Michael S Bernstein, and Scott R Klemmer. 2015. PeerStudio: rapid peer feedback emphasizes revision and improves

- performance. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 75–84.
- [47] Chinmay E Kulkarni, Richard Socher, Michael S Bernstein, and Scott R Klemmer. 2014. Scaling short-answer grading by combining peer assessment with algorithmic scoring. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 99–108.
- [48] Tamar Lewin. 2013. After setbacks, online courses are rethought. *The New York Times* 10 (2013).
- [49] Nan Li, Himanshu Verma, Afroditi Skevi, Guillaume Zufferey, Jan Blom, and Pierre Dillenbourg. 2014. Watching MOOCs together: Investigating co-located MOOC study groups. *Distance Education* 35, 2 (2014), 217–233.
- [50] Panagiotis Louridas. 1999. Design as bricolage: anthropology meets design thinking. *Design Studies* 20, 6 (1999), 517–535.
- [51] Robert F Lusch, Stephen L Vargo, and Gregor Wessels. 2008. Toward a conceptual foundation for service science: Contributions from service-dominant logic. *IBM systems journal* 47, 1 (2008), 5–14.
- [52] Jenny Mackness, Marion Waite, George Roberts, and Elizabeth Lovegrove. 2013. Learning in a small, task-oriented, connectivist MOOC: Pedagogical issues and implications for higher education. *The international review of research in open and distributed learning* 14, 4 (2013).
- [53] Ye Mao, Chen Lin, and Min Chi. 2018. Deep Learning vs. Bayesian Knowledge Tracing: Student Models for Interventions. *JEDM/ Journal of Educational Data Mining* 10, 2 (2018), 28–54.
- [54] Douglas C Merrill, Brian J Reiser, Michael Ranney, and J Gregory Trafton. 1992. Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences* 2, 3 (1992), 277–305.
- [55] Tricia J Ngoon, C Ailie Fraser, Ariel S Weingarten, Mira Dontcheva, and Scott Klemmer. 2018. Interactive Guidance Techniques for Improving Creative Feedback. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 55.
- [56] Jennifer Pearce, Mala K Mann, Caryl Jones, Susanne van Buschbach, Miranda Olff, and Jonathan I Bisson. 2012. The most effective way of delivering a Train-the-Trainers program: a systematic review. *Journal of Continuing Education in the Health Professions* 32, 3 (2012), 215–226.
- [57] Justin Reich, Brandon Stewart, Kimia Mavon, and Dustin Tingley. 2016. The civic mission of MOOCs: Measuring engagement across political differences in forums. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 1–10.
- [58] Daniela Retelny, Michael S Bernstein, and Melissa A Valentine. 2017. No workflow can ever be enough: How crowdsourcing workflows constrain complex work. *ACM, New York, NY, USA* (2017).
- [59] Steven Ritter, John R Anderson, Kenneth R Koedinger, and Albert Corbett. 2007. Cognitive Tutor: Applied research in mathematics education. *Psychonomic bulletin & review* 14, 2 (2007), 249–255.
- [60] Adam C Sales, Anthony Botelho, Thanaporn Patikorn, and Neil Heffernan. 2018. Using Big Data to Sharpen Design-Based Inference in A/B Tests. In *Proceedings of the 11th International Conference on Educational Data Mining*. ACM.
- [61] Gregory D Saxton, Onook Oh, and Rajiv Kishore. 2013. Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management* 30, 1 (2013), 2–20.
- [62] D. Schön. 1985. *The Design Studio: An exploration of its traditions and potential*. London: Royal Institute of British Architects (1985).
- [63] John Self. 1998. The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International Journal of Artificial Intelligence in Education (IJAIED)* 10 (1998), 350–364.
- [64] William R Shadish. 1995. The logic of generalization: Five principles common to experiments and ethnographies. *American Journal of Community Psychology* 23, 3 (1995), 419–428.
- [65] Valerie J Shute. 1991. Rose garden promises of intelligent tutoring systems: Blossom or thorn. (1991).
- [66] Arjun Singh, Sergey Karayev, Kevin Gutowski, and Pieter Abbeel. 2017. Gradescope: a fast, flexible, and fair system for scalable assessment of handwritten work. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*. ACM, 81–88.
- [67] Burrhus Frederic Skinner. 1960. Teaching machines. *The review of economics and statistics* (1960), 189–191.
- [68] Thomas Staubitz, Dominic Petrick, Matthias Bauer, Jan Renz, and Christoph Meinel. 2016. Improving the peer assessment experience on MOOC platforms. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 389–398.
- [69] Lucy Suchman. 1987. *Plans and situated action*. Cambridge: Cambridge University Press.
- [70] Katherine Liu Sun. 2015. *There’s no limit: Mathematics teaching for a growth mindset*. Ph.D. Dissertation. Stanford University.
- [71] Anouschka van Leeuwen, Nikol Rummel, Kenneth Holstein, Bruce M McLaren, Vincent Aleven, Inge Molenaar, Carolien Knoop-van Campen, Baruch Schwarz, Naomi Prusak, Osama Swidan, et al. 2018. Orchestration tools for teachers in the context of individual and collaborative learning: what information do teachers need and what do they do with it? International Society of the Learning Sciences, Inc.[ISLS].
- [72] Stephen L Vargo and Robert F Lusch. 2004. The four service marketing myths: remnants of a goods-based, manufacturing model. *Journal of service research* 6, 4 (2004), 324–335.
- [73] Miaomiao Wen, Diyi Yang, and Carolyn Penstein Rosé. 2015. Virtual teams in massive open online courses. In *International Conference on Artificial Intelligence in Education*. Springer, 820–824.
- [74] Julia Wilkowski, Daniel M Russell, and Amit Deutsch. 2014. Self-evaluation in advanced power searching and mapping with google MOOCs. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 109–116.
- [75] Joseph Jay Williams, Juho Kim, Anna Rafferty, Samuel Maldonado, Krzysztof Z Gajos, Walter S Lasecki, and Neil Heffernan. 2016. Axis: Generating explanations at scale with learnersourcing and machine learning. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 379–388.
- [76] David S Yeager, Ronald E Dahl, and Carol S Dweck. 2018. Why interventions to influence adolescent behavior often fail but could succeed. *Perspectives on Psychological Science* 13, 1 (2018), 101–122.
- [77] David S Yeager, Carissa Romero, Dave Paunesku, Christopher S Hulleman, Barbara Schneider, Cintia Hinojosa, Hae Yeon Lee, Joseph O’Brien, Kate Flint, Alice Roberts, et al. 2016. Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of educational psychology* 108, 3 (2016), 374.
- [78] Kwang Suk Yoon, Teresa Duncan, Silvia Wen-Yu Lee, Beth Scarloss, and Kathy L Shapley. 2007. Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues & Answers. REL 2007-No. 033. *Regional Educational Laboratory Southwest (N71)* (2007).
- [79] Sharon Zhou, Melissa Valentine, and Michael S Bernstein. 2018. In Search of the Dream Team: Temporally Constrained Multi-Armed Bandits for Identifying Effective Team Structures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 108.
- [80] John Zimmerman and Jodi Forlizzi. 2014. Research through design in HCI. In *Ways of Knowing in HCI*. Springer, 167–189.