Digital Clinical Simulation Suite: Specifications and Architecture for Simulation-Based Pedagogy at Scale

Garron Hillaire
Rick Waldron
Joshua Littenberg-Tobias
Meredith Thompson
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Sara O'Brien
G. R. Marvez
Justin Reich
jreich@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

ABSTRACT

Role-plays of interpersonal interactions are essential to learning across professions, but effective simulations are difficult to create in typical learning management systems. To empower educators and researchers to advance simulation-based pedagogy, we have developed the Digital Clinical Simulation Suite (DCSS, pronounced "decks"), an open-source platform for rehearsing for improvisational interactions. Participants are immersed in vignettes of professional practice through video, images, and text, and they are called upon to improvisationally make difficult decisions through recorded audio and text. Tailored data displays support participant reflection, instructional facilitation, and educational research. DCSS is based on six design principles: 1) Community Adaptation, 2) Masked Technical Complexity, 3) Authenticity of Task, 4) Improvisational Voice, 5) Data Access through "5Rs", and 6) Extensible AI Coaching. These six principles mean that any educator should be able to create a scenario that learners should engage in authentic professional challenges using ordinary computing devices, and learners and educators should have access to data for reflection, facilitation, and development of AI tools for real-time feedback. In this paper, we describe the architecture of DCSS and illustrate its use and efficacy in cases from online courses, colleges of education, and K-12 schools.

CCS CONCEPTS

• **Applied computing → Interactive learning environments**.

KEYWORDS

Simulation, roleplay, coaching, AI, learning at scale, professional learning

ACM Reference Format:
Garron Hillaire, Rick Waldron, Joshua Littenberg-Tobias, Meredith Thompson, Sara O'Brien, G. R. Marvez, and Justin Reich. 2022. Digital Clinical Simulation Suite: Specifications and Architecture for Simulation-Based Pedagogy at Scale. In *Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S '22), June 1–3, 2022, New York City, NY, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3491140.3528276

L@S '22, June 1–3, 2022, New York City, NY, USA.

ACM ISBN 978-1-4503-9158-0/22/06.
https://doi.org/10.1145/3491140.3528276

1 INTRODUCTION

Practice is essential to learning, and some domains and disciplines are more amenable to online practice than others. Rehearsing computer programming in online learning environments is straightforward: online learning environments replicate integrated development environments (IDEs) that are functionally very similar to the IDEs that professional software developers might use. This is not the case for many forms of managerial and professional education, especially those that involve substantial synchronous interpersonal engagement. With the current instructional palette of typical learning management systems—text, short videos, multiple choice questions, and discussion forums—it can be difficult to create environments that allow people to have authentic learning experiences in skills like leading teams, providing one-on-one coaching, or teaching.

To address this "missing pigment" in the instructional design palette, we have developed the Digital Clinical Simulation Suite (DCSS, pronounced "decks"), an open-source platform for rehearsing for improvisational interactions. For learners, DCSS is experi-

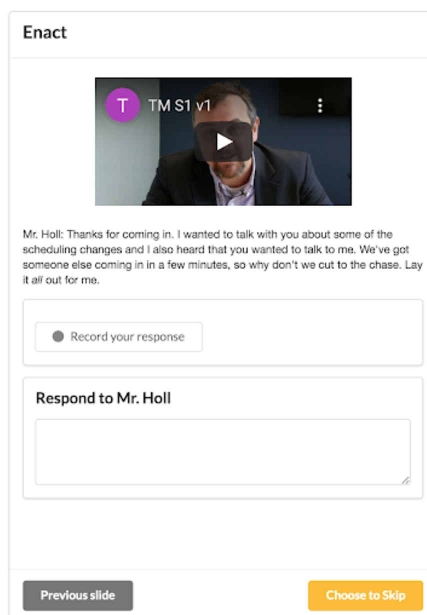

Figure 1: View of a sample "slide" from a DCSS scenario.

enced as a mobile app that can function as a stand-alone "mini-LMS" or be integrated into existing LMSs, where learners rehearse for and reflect upon important decisions in professional settings. Learners are immersed in vignettes of professional life through text, images, and video, and they are called upon to improvise responses through text, radio buttons, multiplayer chat, and recorded audio. In a typical scenario (see Figure 1), participants might read a briefing, watch videos that present conversational turns with a

professional colleague, record audio responses to these conversational turns, indicate a follow-up action through a radio button selection or text input, and then debrief their decisions through reflection questions or by reviewing samples of their own text and audio responses. Scenarios can branch such that participant choices lead to different pathways in the simulation, and an extensible system of AI coaches can "listen" to participant responses and provide feedback and scaffolding as appropriate.

For educators, DCSS is experienced as a publishing and facilitation environment for simulations. DCSS instances maintain a bank of simulations that can be reused and remixed. The authoring environment (see Figure 2) for editing existing or authoring new simulations combines features from slide presentation software (such as Google Slides) and survey tools (such as Qualtrics). The main scenario editing screen is rendered in three panels: the left panel presents a slide sorter of the various slides that constitute the simulation, the central panel allows the editing of one slide at a time, and the right panel includes a menu of interactive elements that can be added to a slide. Instructors can create participant cohorts and assign playlists comprised of one or multiple simulations. A robust set of data tools provide tailored data displays to support instructional facilitation and educational research.

For education systems, DCSS is an open-source software product that can be white-labeled for use in different industries and institutions. At present, the primary deployment of DCSS is Teacher Moments: a simulation bank with over 600 scenarios, developed by a community of 200 instructions, and used by over 13,000 participants with a focus on teacher professional learning. A second deployment is Interpret Me, used by professional social workers to train stakeholders in the criminal justice system about how youth use and interpret social media. DCSS simulations have also been used in the medical community for nursing education and physician education. In what follows, we introduce some of the key literature on simulation-based pedagogy that defines our problem space and informs the ongoing development of DCSS. We then elaborate on our key design principles, describe the architecture of DCSS as it supports those design principles, and describe several use cases for the implementation of DCSS in a variety of professional learning settings. We conclude with directions for future work. Throughout, we provide evidence from a variety of early studies using DCSS that simulation-based pedagogies can support authentic professional education and generate new avenues for measuring learning and behavior change.

2 BACKGROUND AND CONTEXT

2.1 Drills and Scrimmages: Approximations of Practice

Simulations are integral to the history of education technology; in a 1969 survey of teacher education simulations, Cruickshank could point to earlier examples from aviation, medicine, business leadership, and management [6]. Cruickshank's appraisal of the benefits of simulations used in teacher education from the mid-20th century holds up well today: "it is possible, for example, to have the trainee encounter in just two weeks the most critical problems he will face in his first year of teaching, in a threat-free, failure-free environment unlike that of student teaching." (p.24) The most sophisticated simulations of Cruickshank's era were remarkable in their complexity: student teachers stood in front of a projection screen, and for a single simulation, the operator had a bank of 60 prerecorded filmstrips with multiple projectors that allowed for a kind of "choose your own adventure" media experience. A film strip was played, participants responded orally, and the projectionist selected the next film in a sequence based on the participant's response.

In the last two decades, simulations have been framed as "approximations of practice" [12]; opportunities for less-skilled practitioners to rehearse and receive feedback on teaching within low-stake settings [19, 30]. Approximating practice in simulation requires confronting pedagogical dilemmas and tradeoffs. Professional practices can be immensely complex and highly situationally contingent. A teacher, for instance, must simultaneously provide instruction, recall substantive content, track her place in a lesson plan, track the time remaining in the period, observe the reactions of students, and respond with improvisation to questions or behaviors, potentially making thousands of micro-decisions throughout a single class period. The science of complex learning suggests that novices will struggle to simultaneously practice this whole complex assemblage while making improvements in discrete parts. Simulations are often intentionally designed to focus some components of the scenario, but not others, directing the novices' attention to key instructional features [2, 24]. However, isolating elements of the complex assemblage for rehearsal can feel inauthentic, which can reduce motivation and transfer for simulation participants. Managing this risk of inauthenticity is a key challenge to simulation designers.

This dilemma is intuitively well understood by athletic coaches and music teachers, who use an instructional combination of practices such as drills, scrimmages, and rehearsal in their instruction. Scrimmages allow athletes to rehearse for the complex assemblage of sports in low-stakes but realistic settings. Drills abstract away the complexity of a sport to allow athletes to develop fluency and automaticity in particular situations that can then be integrated back into a complex assemblage. These drills can be at varying levels of complexity: free throw drills in basketball can build automaticity in a kinesthetic motion, while two-on-one breakaway drills can help athletes develop heuristics and techniques to deploy in uncommon but high-stakes situations. A typical scrimmage might only include a few two-on-one breakaway situations distributed idiosyncratically among players, but a set of drills can

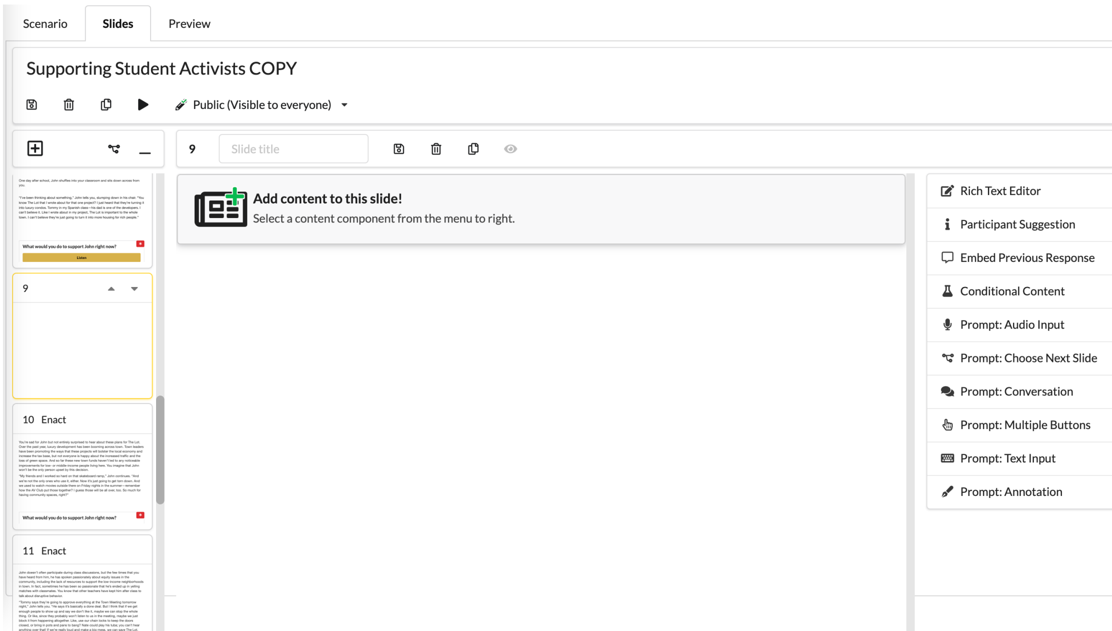

Figure 2: Authoring environment for DCSS scenario.

allow a whole team multiple opportunities to rehearse for these situations in time-efficient ways. Generally speaking, DCSS supports professional practice opportunities that are more akin to drills than scrimmages, emphasizing opportunities for repeated rehearsal of key interactions rather than trying to comprehensively recreate an entire setting. This kind of repeated practice is akin to Ericsson's deliberate practice (although strictly speaking, deliberate practice requires a well-established target expert performance—like a perfectly rendered recital piece—and professional situations often lack well-defined correct answers) [10].

2.2 Digital Adaptations of Live-Actor Roleplay

Two sets of simulation researchers have been particularly influential in the development of DCSS: Ben Dotger at Syracuse [7] and Elizabeth Self and Barbara Stengel [27] at Vanderbilt. Dotger is a special education professor who developed live-actor simulations using facilities available at the neighboring medical school. He trained actors who typically played the role of patients to act as parents of school children, often children with disabilities. In these simulations, Dotger trained the actors to respond improvisationally to participants, but also to guide the conversation towards certain trigger phrases. The actor playing a parent might have a variety of interactions with novice teachers, and the actor's job is to ensure that the conversation always includes the phrase, "I was a teacher once. I will be talking to the principal about your 'unreasonable expectations' of your students." All participants then have to respond to this cue, allowing for a collaborative debrief experience where participants can reflect on their varying decisions in this situation.

If learners benefit from interactions with human actors that guide conversations towards defined trigger phrases, then we theorized that they could potentially benefit from a digital clinical simulation where those trigger phrases are "hard-coded" into a narrative interaction. In adapting Dotger's simulations into DCSS, we might take the kinds of things actors typically say, record an actor saying these phrases (using video, text, and/or images), and then intersperse these recordings with moments where DCSS participants need to provide an audio response using an audio recorder. Certainly, much might be lost in digitizing these interactions, in particular the real-time, improvisational responses of the actors in the simulation. There are substantial potential gains in scalability, however. Coordinating an entire class or program of teacher education candidates to travel to a simulation room to have scheduled one-on-one recorded interactions with trained actors is logistically complex. If that same interaction can be imperfectly digitized and learners can participate in a similar situation on their mobile phones, at their convenience, and as often as they would like to practice, then repeatability, adaptability, and administrative conveniences might offer a compelling value proposition to professional educators.

2.3 Conditionally Inclusive Ideology and Disequilibrium

A key principle that emerges from Dotger's simulation work is the idea of "conditionally inclusive ideology" [8]. Dotger's program emphasizes inclusive special education, where as much as possible, students with diagnosed learning disabilities are kept in mainstream classrooms rather than pulled out into separate learning tracks and spaces. Dotger found that his students were strong advocates for inclusive principles in seminar discussions. However, when those same students were placed in simulations, he found them strikingly willing to consider and adopt "pull-out" practices in simulated settings. Students who expressed a strong ideological commitment to inclusivity in a seminar might not act on that commitment when

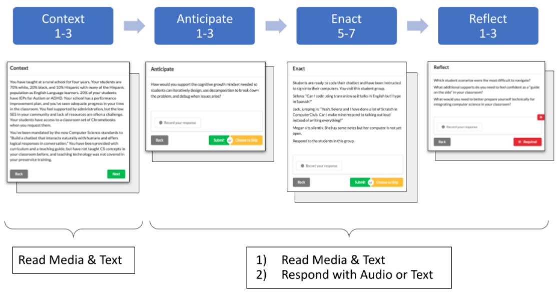

Figure 3: Typical sequence of slides in a DCSS scenario (numbers refer to typical number of slides in a section)

faced with simulations where various trade-offs seem more salient. Simulations allow educators to surface and address these gaps between expressed beliefs and professional choices.

Students do not necessarily immediately recognize this conditionally inclusive ideology—this gap between stated and enacted beliefs, so Dotger developed a routine of debriefing and reflection that highlighted some of these tensions. This routine was formalized into the SHIFT protocol by Self and Stengel. Self and Stengel focus on what they call "critical incident scenarios" that relate to issues of racism, bias, and prejudice in school settings. In the SHIFT protocol, participants get a briefing about a scenario, engage in a simulation with a trained actor, and then participate in an immediate reflection where they process their experience and typically select some piece of evidence from their performance to share with a larger community of learners (see the example template in Figure 3). In the collaborative debrief, individual participants bring together these pieces of evidence in a collective discussion. Guiding this process, skilled facilitators can use evidence from simulations to help groups of learners understand the range of potential choices in a professional dilemma and reflect on the appropriateness of their own decisions.

In this model, cognitive disequilibrium is essential to participant learning. Scenarios include challenging moments where participants are "pulled up short" [27] by challenging situations that could be high stakes in the real world. In debriefs, participants review their actions and recognize conflicts with their own beliefs or the beliefs of others.

2.4 Cognitive Dissonance, Reflection, and Change

What Dotger calls disequilibrium is closely tied to cognitive dissonance. In classical behaviorist theory, individuals change their habituated response to a circumstance when conditioned by rewards or punishment. The literature on cognitive dissonance proposes an alternative approach to behavior change [15]. People find it quite unpleasant when they realize that their held or stated beliefs conflict with their revealed actions, and they are often strongly motivated to reconcile this conflict. If you, the reader, are strongly concerned with climate change, reflect for a moment on how often you drive and fly. Your belief that ecological and human systems will be severely harmed by actions like driving and flying might lead you to feel uncomfortable. People deal with this discomfort in a variety of ways—minimization is one approach ("my driving and flying doesn't contribute that much greenhouse gas compared to other sources"), changing beliefs is another ("actually, driving and flying is so great that it's worth simply adapting to the forthcoming planetary damage"), and so is changing actions ("to align my actions and values, I will driving less"). In the context of professional education, simulation-based pedagogies are designed to create scenarios that generate cognitive dissonance and debriefing experiences that are designed to emphasize the possibilities, options, and techniques of constructive behavioral change.

Since cognitive dissonance is both a cognitive and emotional process[14]—where a cognitive realization is linked to emotional discomfort—educators, instructional designers, and researchers need to understand how participants are thinking and feeling during simulations to support effective learning design. The DCSS platform is instrumented to capture different kinds of data about student experiences, including participants' own reflections about their simulated actions, beliefs, and learning.

In summary, practice is essential to managerial and professional learning, and DCSS is designed to empower educators to create scenarios for simulation-based pedagogies. Our focus is not on trying to completely recreate complex situations (like a scrimmage) but to support educators in developing "drills" that allow participants to practice targeted skills, abstracting away complexity to allow for more focused, repeatable practice. We encourage educators to embed our simulations in an instructional sequence of briefings, enactments, and debriefings. Cognitive dissonance is a key design feature throughout this process. When scenarios and debriefing experiences allow participants to productively experience cognitive dissonance and explore actions and pathways that can reduce that dissonance, then DCSS opens pathways to professional learning and behavioral change.

3 DESIGN PROBLEMS, PRINCIPLES, AND EVIDENCE.

3.1 Problem Statement

From these considerations of approximations of practice, cognitive dissonance, and reflection, we define our design problem as working toward a system that provides: authentic, context-relevant, targeted, repeatable, low-cost simulations where participants experience challenging professional decisions with real-time feedback followed by opportunities for collaborative, data-informed discussion and reflection.

3.2 Design Principles

The software architecture of DCSS and related support materials that addresses this problem statement enacts six design principles, which together aspire to allow community educators to create powerful simulation-based learning experiences at scale.

Community Adaptation means that educators with strong domain knowledge but limited experience with programming or software design should be able to easily develop scenarios, AI coaches, and other instructional elements in DCSS. Our commitment to Community Adaptation emerges from two sources. First, digital decolonization refers to efforts to recenter power in historically marginalized and displaced communities [11]. Large scale learning technologies should support local adaptation and design rather than trying to distribute canonical experiences developed in elite institutions. This focus on adaptation aligns with a growing understanding of heterogeneous context effects, that interventions that work very successfully in one context will often fail in other contexts [5, 20]. Scaling effective education technologies means creating tools that communities can adapt.

Masked Technical Complexity means that participants should have a seamless front-end experience when using typical, low-cost computing devices. DCSS offers a rich and complex set of tools for designers and researchers. Back-end elements of the platform—for instance, our extensible AI infrastructure or data recording systems—can be technically sophisticated. However, participants can click a link on their phone and easily play through the simulations without encountering this complexity.

Authenticity of Task means that DCSS design focuses on challenging participants to practice real communication and actions in professional or managerial settings. There are three other kinds of authenticity that are less emphasized- authenticity of the setting (we do not try to comprehensively recreate workplace settings, for instance, using AR or VR), authenticity of the role (scenarios can ask participants to play different roles in a scenario, not just their own professional roles), and authenticity of complexity (we abstract away the complexity of reality to support focused practice). Research suggests that simulations built from low-fidelity elements like text and images are perceived as highly authentic if participants engage in professional tasks that feel authentic [13].

Improvisational Voice supports participants in actually enacting communication rather than describing "what they would" do in a circumstance. Recorded audio is an essential feature of DCSS and is the primary feature that merits the design of a new platform (in many respects, DCSS is very similar to survey software like Qualtrics). When participants are required to improvisationally talk in response to scenarios, they find the challenge engaging, and reflecting on their own words can be a key source of cognitive dissonance [31].

The 5Rs of Data Access [29] are right stakeholder, right data, right format, right context, and right timing (adapted from the 5 Rs of medication administration: right drug, right patient, right dosage, right route, and right time). Extensive data is collected in DCSS, and we aspire to provide stakeholders with appropriate data for reflection, facilitation, design and research. Participants are consented in each scenario. Instructors have real-time access to participant data to facilitate classes and manage debriefs. Participants can see their own data from the scenarios, and it is straightforward to "pipe" data forward in a scenario; for instance, so that participants can record audio responses in an improvisational enactment and then listen to it again later as part of a debrief.

Extensible AI coaching allows an infrastructure for automated real-time feedback in scenarios. Designers can develop "listeners" that review recorded audio and text inputs and classifiers that detect key features across scenarios (like an audio classifier that detects confusion) or within scenarios (like a text classifier that can recognize whether a participant evaluating a specific website in a scenario about search literacy recognizes a particular site as satirical). These classifiers can trigger "coaches" that provide feedback and scaffolds, move participants down different scenario branches, or let them try an interaction again. Combined, these agents can provide real-time feedback as a complement to instructor-led debriefs.

3.3 Evidence of Efficacy

These design problems and principles suggest that evaluating DCSS requires three primary sources of evidence:

- Evidence that participants in DCSS simulations improve in the skills, techniques, and heuristics of professional practice
- Evidence that DCSS simulations can be integrated into typical professional education settings at reasonable cost, and efforts devoted to integration
- Evidence that DCSS simulations can be adapted and authored by typical professional educators

4 SYSTEM ARCHITECTURE AND WEB SERVICES

Conceptually, DCSS can be understood as four modules–1) authoring, 2) rendering, 3) data storage, and 4) extensible AI. DCSS sends JSON strings across these modules to organize interactive elements on slides into scenarios, render those slides and elements in browsers, deploy AI coaches appropriately, and accurately record the history of all actions in the system (see Figure 4).

4.1 Authoring

In the authoring module, educators choose from a set of interactive elements and place them onto slides. Interactive elements include features like rich text display, video display, audio input, text input, and button selection input. Input elements are configurable to assign AI classifiers, define branching pathways, and customize data recording parameters. Each interactive element, stored as a JSON string, is assigned a PromptID. Each scenario is assigned an identifier that can function as a URL, and a series of scenarios can be assigned to a "cohort" so that a specific group of participants can be assigned a specific sequence of scenarios. Representing and storing scenarios as a series of editable, modular elements supports our design principle of Community Adaptation, allowing users to simply create new scenarios or adapt and remix existing scenarios that are tailored to a local context.

4.2 Rendering and Data Storage

When participants navigate to a scenario, the rendering module rehydrates the JSON strings such that the participant sees the interactive elements and slides in the intended order. Each time an interactive element is rendered, and the participant inputs a response (such as recorded audio, typed text, or radio button selection), a new PromptResponseID is generated that captures a snapshot of the participant action and the state of the prompt at that time. Thus, if an educator iteratively improves a scenario by changing the text or options of a prompt, the system records the state of the responses and the contemporary state of the interactive element simultaneously. Each input then potentially passes to several additional modules and subservices for transcripts, AI classification, and storage. While this complex activity happens on the back end, the user experience

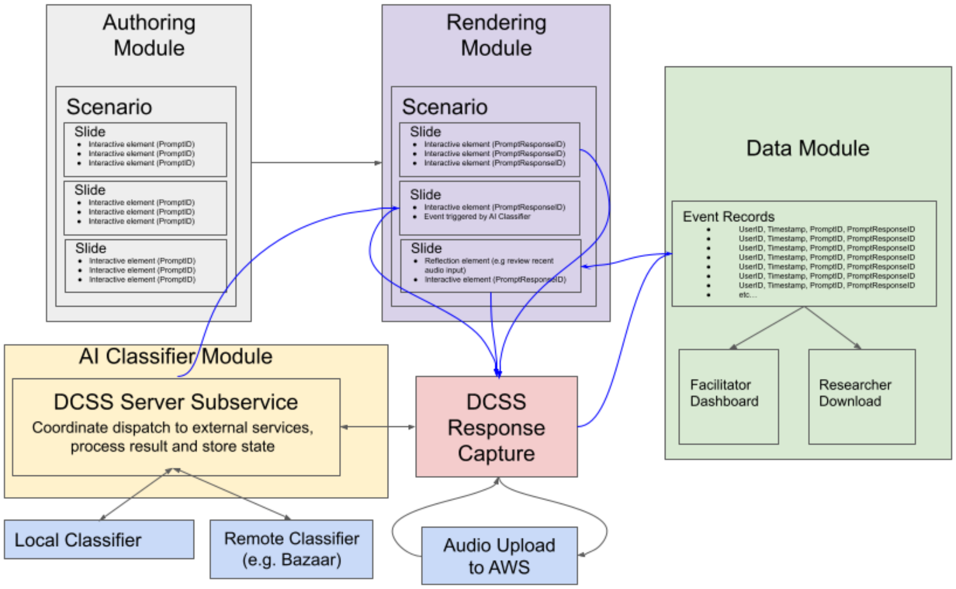

Figure 4: Architectural schematic of DCSS modules.

goal for the rendering module is masking technical complexity: the user should be simply able to proceed through the scenario with any basic mobile device.

The signature interactive element of DCSS is the recorded audio input. In some respects, nearly every other basic feature of DCSS could be generated in survey software like Qualtrics, except for the recorded audio input. When generated by participants, the recorded audio files are stored in Amazon Web Services lockers, and those files are referenced as links in the JSON string associated with a PromptResponseID. These files are also passed to IBM Watson for transcription in near real time, and when transcripts are returned, that data is also associated with a PromptResponseID. Recorded audio supports our design principles of Authenticity of Task and Improvisational Voice, allowing professionals to simulate their authentic responses to challenging interactions.

When a PromptResponse is generated, the JSON string is stored in a postgres database, that stores timestamped (as closely as possible) records of all participants' browser and server actions, including information about the state of each prompt at the time of the response, that allows for a full history and playback of all participant actions in the system.

Scenarios can be designed so that these records can be recalled during a scenario run in the form a "piped response." After participants record the improvised audio in response to a prompt, a subsequent slide can ask participants to listen to their audio, read their transcript, and reflect on their choices and actions. Our granular data storage supports our 5Rights approach to data collection and usage, ensuring that we can provide specific, accurate data with the necessary context to different system stakeholders as needed.

4.3 AI Artchitecture

As participants respond to prompts, these responses are then passed to the extensible AI system. The extensible AI system assumes that all classifiers will operate remotely, and the API defines a specification for making web socket connections between individual classifiers and the DCSS AI subservice. Our recommended path for developing new AI classifiers is a simple protocol that involves developing a classifier in a Jupyter notebook, exporting the classifier as a compiled Python program, and then forking our example repo to store the program on our Heroku instance. However, any classifier on any remote server that meets the API specifications can be included. For instance, classifiers developed in the CMU DANCE lab's Bazaar repository have been integrated into DCSS [1]. New classifiers are checked into the DCSS repository (or any of the downstream white labeled instances), and they are then available for the scenario authors to select for any relevant interactive element.

When classifiers are activated in DCSS's interactive elements, responses to those elements are passed to the remote classifiers and classifiers send back information about whether the classifier triggers an action or not on any given interactive element. For some classifiers, the state is recorded over multiple prompts, so that, for instance, a user might need to be classified as confused across three prompts before an AI agent determines that the participant should receive feedback. A classifier can trigger actions on any subsequent slide in a DCSS scenario, such that participants might get immediate feedback, be sent down a different branch, or receive customized reflection questions at the end. This system supports our design principle of extensible AI coaching, where just as educators can compile a bank of scenarios for use with different learners, so can developers create a bank of AI classifiers that can be integrated across different scenarios.

4.4 DCSS Summary

To summarize, scenarios consist of a sequence of slides populated by interactive elements, encoded as JSON strings. When participants navigate to a scenario in a browser, the browser rehydrates the JSON to render the scenario as intended, and each action the participant takes is recorded to a postgres database with both a snapshot of the interactive element and data about the participant response. These responses are also served to the embedded AI classifiers to evaluate whether participant responses should trigger feedback or other scenario interventions. The modular structure of interactive elements allows educators to easily remix and adapt existing scenarios, or construct new ones, and facilitates data recording where researchers can reconstruct the action history of every consented user in every scenario.

5 DCSS: USE CASES IN TEACHER EDUCATION

The most extensive testing and deployment efforts with DCSS have been in support of professional learning for practicing educators through the Teacher Moments white-label implementation. Below, we describe three use cases for implementing digital clinical simulations in authentic professional learning settings. In evaluating our use cases, we highlight three kinds of evidence: that DCSS 1) supports the development of professional skills and mindsets, 2) can be adapted and implemented in typical professional learning settings, and 3) that DCSS scenarios can be authored and implemented by typical professional educators.

5.1 Simulations Embedded in MOOCs: Teacher Moments and Online Professional Development

One advantage of *Teacher Moments* seamless front-end experience is that it is easy to integrate the application into existing learning management systems. In our work, we have embedded *Teacher Moments* scenarios in two large-scale online courses on edX: *Becoming a More Equitable Educator: Mindsets and Practice* [1] — a course on how educators can incorporate more equitable approaches into their instruction— and *Sorting Truth from Fiction: Civic Online Reasoning* [2] — a course on teaching students how to effectively evaluate online information. Combined, these courses enrolled over 23,000 registrants. Each course was organized around units of instruction with a specific core concept or competency. Simulations were used to help illustrate specific concepts or methods, prompt reflection on previously held beliefs, and encourage integration of new concepts into participants' teaching practices. At the conclusion of the course, participants were encouraged to move from learners to facilitators, and to share the simulations from the course with their colleagues, so the impact of the learning experience could scale beyond the participants directly registered in the course. Becoming a More Equitable Educator launched from March - June 2020 and was run a second time from January - August 2021. Sorting Truth From Fiction ran from September - November 2020 and then from May - August 2021. In the following paragraph, we report data from both runs of each course.

In both courses, participants reported increases in learning after completing the online courses with embedded simulations. Participants in the *Becoming a More Equitable Educator: Mindsets and Practices* course reported large statistically significant increases in self-efficacy for teaching with an equity perspective after completing the course ($d = 0.90$, $df = 503$, $p < 0.001$) and were more likely to report using equity practices such as reflecting on their own identity, sharing equity resources with colleagues, and participating in educational equity networks ($d =0.26$, $df = 504$, $p < 0.001$). Additionally, participants' mindsets shifted toward more equity-oriented perspectives about teaching after completing the course ($d =0.68$, $df = 439$, $p < 0.001$). In Sorting Truth From Fiction, course participants also reported large statistically significant increases in self-efficacy for helping students evaluate online information ($d = 0.85$, $df = 397$, $p < 0.001$). Course participants were also substantially more likely to correctly identify effective and ineffective instructional strategies for teaching students about evaluating online information ($d = 1.57$, $df = 347$, $p < 0.001$). Participants described the digital clinical simulations and post-simulation video debriefs as key components of their learning. In post-course surveys, 90% of participants across courses reported that digital clinical simulations were 'very important' to their learning experience. Participants also described the digital clinical simulations as authentic representations of the instructional challenges they faced in their own teaching. Most (82%) participants agreed that they 'found themselves thinking about the real-life situation' that the simulation represented and many (72%) agreed the simulations were 'good representations of real-life teaching situations. These findings suggest that digital clinical simulations were useful tools for encouraging individual reflection and learning within large-scale asynchronous courses.

5.2 Simulations in Higher Education: The Innovative New Practice and Rehearsal in Teacher Education (INSPIRE) Program

While MOOCs allow for rapid, large-scale dissemination and testing of digital clinical simulations, partnerships with teacher educators open the possibility for a different mechanism for scaling and distribution. When College of Education faculty adopt digital clinical simulations, they can use the DCSS platform to develop their own scenarios or adapt existing scenarios to their own settings. We have run several iterations of INSPIRE fellowship programs that are designed to provide initial training for teacher educators to author and implement scenarios in their courses. In INSPIRE-CS-AI, we worked with a network of 22 Computer Science (CS) teacher educators in 9 states across the U.S. as a proof of concept that illustrates how community involvement can create simulations for teacher professional development [9, 16] and at the same time collected community perspectives to train AI leveraging the extensible AI architecture of Teacher Moments [17]. In the program, teacher educators created their own simulations about equity issues that they anticipate their pre-service teachers will face [21, 22, 26, 28].

For instance, one program participant published their efforts to create a scenario about recognizing and reacting to student

[1] Becoming a More Equitable Educator, https://bit.ly/3uwPXIA
[2] Sorting Truth from Fiction, https://bit.ly/3B7ak0f

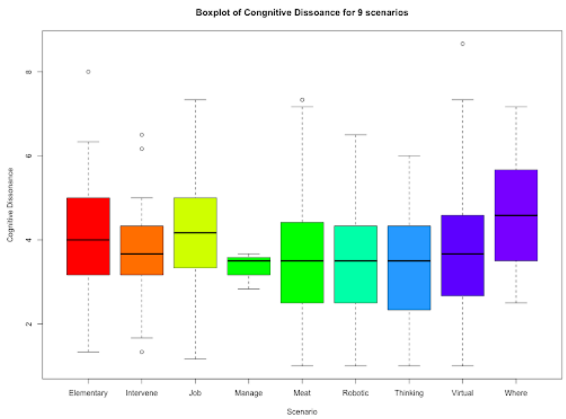

Figure 5: Distribution of self-reported cognitive dissonance by participants in nine community created DCSS scenarios

trauma in the CS classroom [28]. Although the program started pre-pandemic, Sullivan used their scenario with CS teacher education students during an early pandemic period when field placements and other clinical learning opportunities had been canceled. Participants played through the scenario where a student becomes unexpectedly disengaged, and the scenario encourages participants to frame their thinking not as "what's wrong with this student?", instead as the question "what happened to this student?" Participants responded to a set of debriefing questions reflecting on what they noticed, interpretations of what they noticed, and how they interacted with the disengaged student. INSPIRE CS-AI participants played through each other's scenarios, and many chose to integrate a sequence of scenarios created by multiple participants in their own courses throughout the following years. By sharing their scenarios with each other, participants built a community of teacher educators using simulation-based pedagogies. Embedded at the end of several simulations from the INSPIRE CS-AI project was a survey designed to measure the extent to which they experienced cognitive dissonance - an anticipated important signal for learning in simulations.

The results from using a consistent exit survey, the cognitive dissonance questionnaire, demonstrated that the simulations created by teacher educators from the INSPIRE CS-AI project had very comparable results (see Figure 5). Twelve INSPIRE CS-AI fellows implemented at least 1 of the 9 scenarios which included the cognitive dissonance survey. The results indicated a rather consistent score with means scores consistently near 4 on a scale from 1-9. This suggests that with support and guidance, typical educators can author simulations that provoke the kind of cognitive dissonance theorized to support reflective learning and change.

An ongoing effort that began during the INSPIRE CS-AI project is to demonstrate how the system can support AI development by coordinating communities to explore a single measurement goal. Fellows created scenarios that included opportunities for participant audio input (roleplaying responses to students, parents, and colleagues), and during the reflection phase of the scenario, participants were asked to reflect on their responses and to label their own audio data by answering the question "Do you sound confused?" The question allows participants to reflect on their performance, while supporting the development of classifiers that can detect confusion in vocal expression. We generated labeled data from over 10,000 audio files that integrated participants' perspectives into defining the ground truth, and we are developing AI coaching agents that can detect participant confusion in future scenarios and provide feedback through scaffolds and support.

5.3 Simulations in Workplace Settings: K-12 Schools

Teachers typically receive their initial training over an undergraduate degree or Masters program in colleges of education, but then most of the rest of the professional learning that they receive throughout their careers is through districts. We partnered with teacher leaders and instructional coaches in the Boston Public Schools who worked with middle years (3rd-9th grade) math teachers to develop scenarios and practice skills related to facilitating inclusive discussions. Our workshops included instruction on facilitation techniques, playing scenarios created by our instructional designers, and then helping fellows create their own scenarios to use with colleagues.

In our professional learning with these educators, we found an instance of "conditionally inclusive ideology." One major theme in the professional learning in the program was including the voices of students in classroom discussion when they do not have the most efficient or completely correct answer and finding ways of including their thinking in productive discourse. In our seminar discussions, participating teachers echoed the importance of this view, as one said "I never choose the group who has it all correct to answer first. Never. Because then, I think they [rest of the class] stop listening." [3]

However, when educators participated in simulated class discussions, we found that they regularly chose to start conversations either by calling on students with correct, efficient answers, or by calling on students with incorrect and inefficient answers, but then simply correcting them rather than trying to draw out their ideas and the potential value of their contributions. Using a pre-post study design, we found that at the beginning of our professional development intervention, teachers in simulations rarely called on students with inefficient or incorrect answers to start a conversion. In our post-test, we found that teachers had become more likely to include diverse perspectives in small group discussions, but they made substantially less progress in calling on the different perspectives presented in a simulation about a whole class discussion. By identifying these differences of perspective in professional simulation-based learning, we can refine future versions of our programming to provide more support to teachers in facilitating whole class discussions that include diverse voices.

5.4 Discussion

Combined, our three use cases demonstrate that DCSS can be implemented in a wide range of professional learning settings, from online courses, to in-person colleague courses, to professional workplace settings. Our fellowship programs demonstrate that with support, typical educators can author, adapt, and implement DCSS simulations in professional learning settings. DCSS simulations collect data about participant behaviors, attitudes, and performance that can be used to formatively assess learning and inform instructional design and refinement. In fields like teaching where measuring professional behavior change (by observing participants in the field) is costly, simulations offer new research possibilities for measuring the effectiveness of professional learning experiences.

6 FUTURE WORK IN DIGITAL CLINICAL SIMULATIONS

From these early case studies and evaluations, we see three important dimensions of growth for the DCSS project: expanding our community of educators, empowering educators to develop AI coaches, and supports, and collecting rigorous evidence of efficacy in changing practice in the field.

6.1 Sustaining a Community of Practice

The DCSS includes features designed to support networks of educators. Because scenarios are shared and can be copied, edited, and remixed, educators can adapt scenarios created by others to fit their own students' needs. The current platform does not have social features to connect educators, so to complement the platform we have created a community of practice that meets regularly. Through the Community of Practice, educators learn best practices for using simulations from colleagues beyond their own institution. With the Community of Practice meetings, we hope to highlight the distributed knowledge about DCSS, rather than positioning ourselves as the sole experts on the system.

The Community of Practice currently meets four times per month: every other week on Wednesday mornings and Thursday afternoons, to accommodate users in different time zones. In addition to these regular meetings, we are working to create other opportunities for users to collaborate with one another. For example, we are currently planning a series of workshops in February 2022, offered jointly with the Learning Analytics Learning Network. These workshops will help users create and train their own audio and text classifiers, which they can use to build dynamic supports in their simulations.

6.2 Community Created AI Classifiers

Another future direction of this project is to develop functionality so educators can create their own AI classifiers that automatically evaluate and provide feedback on responses within simulations. Just as authoring a DCSS scenario currently requires no special technical or programming skills, we hope to develop tools so that any educator can create AI coaching agents. One pathway towards this goal builds on the existing feature that allows participants to self-label their own responses after they complete a simulation. This effort focuses on training AI by centering students through providing data labels on their own responses [18]. We plan to prototype systems where scenario authors define binary reflection questions ("did you sound confused in that response?"), participants label their own data, and then we use libraries like Ludwig [3] and Aequitas [4] to develop support for automated classification using self-labels.

Second, we are examining whether we can support authors in developing their own classifiers using a small number of labeled examples. Few-shot learning are a subclass of machine learning approaches that draw on a small number of labeled examples. In the last few years, the emergence of large-scale language models such as BERT and GPT-3 have changed the landscape for few-shot learning, allowing the possibility of developing classifiers within only a small number of labeled examples and without any prior fine-tuning of the models [4, 23, 25]. In this proposed system, the authors of the scenarios would write prompts, write sample responses, and write feedback on those sample responses. The system would then automatically match participant responses to sample responses and determine whether the match is close enough to warrant sending feedback. An interface would allow the authors of the scenarios to offer feedback on the decisions of these coaching agents, to identify new categories of participant responses, and to develop appropriate feedback on responses in those categories.

Through these approaches, we hope that future versions of DCSS will support simple systems for community development of AI coaching agents.

6.3 Rigorous Efficacy Evaluations

While early evidence from field trials suggests that participants enjoy digital clinical simulations, find them engaging, and learn from these experiences, more work remains to rigorously demonstrate that clinical simulations for professionals can be linked to improved performance in the field. In future work, we hope to conduct randomized trials of implementation; for instance, in the teacher education field, we hope to randomly assign DCSS-embedded online courses to educators in schools, and then collect multiple measures of teacher performance, including classroom observations, student surveys, and measures of student academic performance to assess the efficacy of our professional learning methods. In particular, if we can identify strong correlations between professional performance in simulation and professional performance in real workplace settings, then it opens the door for DCSS to serve as a powerful research tool for studying professional learning—flexible and nimble enough to be used in rapid cycles but linked to real gains in professional performance in authentic settings.

7 ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under grant number # 1917668 and the Bill and Melinda Gates Foundation.

REFERENCES

[1] David Adamson and Carolyn Penstein Rosé. 2012. Coordinating Multi-dimensional Support in Collaborative Conversational Agents. In *Intelligent Tutoring Systems*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Stefano A. Cerri, William J. Clancey,

[3] Ludwig, https://ludwig-ai.github.io/ludwig-docs/
[4] Aequitas, https://github.com/dssg/aequitas

Giorgos Papadourakis, and Kitty Panourgia (Eds.). Vol. 7315. Springer Berlin Heidelberg, Berlin, Heidelberg, 346–351. DOI:**http://dx.doi.org/10.1007/978-3-642-30950-2_45** Series Title: Lecture Notes in Computer Science.

[2] Clark Aldrich. 2009. Learning online with games, simulations, and virtual worlds: strategies for online instruction. *JosseyBass guides to online teaching and learning Variation JosseyBass guides to online teaching and learning* (2009). ISBN: 9780470438343.

[3] Gregory Benoit, Rachel Slama, Roya Madoff Moussapour, Nancy Anderson, and Justin Reich. 2021. Simulating More Equitable Discussions: Using Teacher Moments and Practice Based Teacher Education in Mathematical Professional Learning. In *Proceedings of the Forty-Third Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.*

[4] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *arXiv:2005.14165 [cs]* (July 2020). **http://arxiv.org/abs/2005.14165** arXiv: 2005.14165.

[5] Christopher J. Bryan, Elizabeth Tipton, and David S. Yeager. 2021. Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour* 5, 8 (Aug. 2021), 980–989. DOI:**http://dx.doi.org/10.1038/s41562-021-01143-3** Number: 8 Publisher: Nature Publishing Group.

[6] Donald R. Cruickshank. 1969. The Use of Simulation in Teacher Education: A Developing Phenomenon. *Journal of Teacher Education* 20, 1 (March 1969), 23–26. DOI:**http://dx.doi.org/10.1177/002248716902000105**

[7] Benjamin Dotger. 2012. *I Had No Idea: Clinical Simulations for Teacher Development.* Age Publishing. **https://www.amazon.com/Had-No-Idea-Simulations-Development/dp/1623961955**

[8] Benjamin Dotger and Christine Ashby. 2010. Exposing Conditional Inclusive Ideologies Through Simulated Interactions. *Teacher Education and Special Education* 33, 2 (May 2010), 114–130. DOI:**http://dx.doi.org/10.1177/0888406409357541** Publisher: SAGE Publications Inc.

[9] Ritam Dutt, Garron Hillaire, Alison Fang, Laura Larke, Carolyn Rosé, and Justin Reich. 2021. Investigating Adoption and Collaboration with Digital Clinical Simulations by Teacher Educators. Association for the Advancement of Computing in Education (AACE), 1209–1217. **https://www.learntechlib.org/primary/p/219277/**

[10] K. Anders Ericsson and Kyle W. Harwell. 2019. Deliberate Practice and Proposed Limits on the Effects of Practice on the Acquisition of Expert Performance: Why the Original Definition Matters and Recommendations for Future Research. *Frontiers in Psychology* 10 (2019). **https://www.frontiersin.org/article/10.3389/fpsyg.2019.02396**

[11] Rebecca Ferguson, Tim Coughlan, Kjetil Egelandsdal, Mark Gaved, Christothea Herodotou, Garron Hillaire, Derek Jones, Iestyn Jowers, Agnes Kukulska-Hulme, Patrick McAndrew, Kamila Misiejuk, Ingunn Johanna Ness, Bart Rienties, Eileen Scanlon, Mike Sharples, Barbara Wasson, and Martin Weller. 2019. *Innovating Pedagogy 2019: Open University Innovation Report 7.* Technical Report. The Open University, Milton Keynes.

[12] Pamela Grossman, Christa Compton, Danielle Igra, Matthew Ronfeldt, Emily Shahan, and Peter Williamson. 2009. Teaching practice: A cross-professional perspective. *Teachers College Record* 111, 9 (Sept. 2009), 2055–2100. **https://www.tcrecord.org/Content.asp?ContentId=15018**

[13] Stanley J. Hamstra, Ryan Brydges, Rose Hatala, Benjamin Zendejas, and David A. Cook. 2014. Reconsidering Fidelity in Simulation-Based Training. *Academic Medicine* 89, 3 (March 2014), 387–392. DOI:**http://dx.doi.org/10.1097/ACM.0000000000000130**

[14] Eddie Harmon-Jones. 2000. A cognitive dissonance theory perspective on the role of emotion in the maintenance and change of beliefs and attitudes. In *Emotions and Beliefs: How Feelings Influence Thoughts.* Cambridge University Press, Cambridge, 185–211.

[15] Eddie Harmon-Jones. 2019. *Cognitive dissonance: reexamining a pivotal theory in psychology* (second edition ed.). American psychological association, Washington DC.

[16] Garron Hillaire, Laura Larke, and Justin Reich. 2020. Digital Storytelling through Authoring Simulations with Teacher Moments. In *Society for Information Technology & Teacher Education International Conference.* Publisher: Association for the Advancement of Computing in Education (AACE), Online, 1736–1745. **https://www.learntechlib.org/p/215950/**

[17] Garron Hillaire, Laura Larke, Justin Reich, and Alison Fang. 2021. Technology Adoption Concerns and Help-Seeking from Teacher Educators Authoring, Implementing, and Researching Equity Focused Simulations. (April 2021).

[18] Garron Hillaire, Bart Rienties, Mark Fenton-O'Creevy, Zdenek Zdrahal, and Dirk Tempelaar. 2022. Incorporating student opinion into opinion mining. In *Open World Learning* (1 ed.). Routledge, London, 171–185. DOI:**http://dx.doi.org/10.4324/9781003177098-15**

[19] David Kaufman and Alice Ireland. 2016. Enhancing teacher education with simulations. *TechTrends* 60, 3 (May 2016), 260–267. DOI:**http://dx.doi.org/10.1007/s11528-016-0049-0**

[20] René F. Kizilcec, Justin Reich, Michael Yeomans, Christoph Dann, Emma Brunskill, Glenn Lopez, Selen Turkay, Joseph Jay Williams, and Dustin Tingley. 2020. Scaling up behavioral science interventions in online education. *Proceedings of the National Academy of Sciences* 117, 26 (June 2020), 14900–14905. DOI:**http://dx.doi.org/10.1073/pnas.1921417117** Publisher: National Academy of Sciences Section: Social Sciences.

[21] Laura R Larke, Garron Hillaire, Hao Chen, Ritam Dutt, Carolyn Penstein Rose, and Justin Reich. 2020. Cognitive Dissonance and Equity: Designing Digital Simulations for K-12 Computer Science Teacher Education. 2.

[22] A. T. Leftwich and Hao Chen. 2021. Using Online Simulations to Uncover Preservice Teachers' Perceived Level of Responsibility Regarding Computer Science Integration. (April 2021).

[23] Lajanugen Logeswaran, Ann Lee, Myle Ott, Honglak Lee, Marc'Aurelio Ranzato, and Arthur Szlam. 2020. Few-shot sequence learning with transformers. *arXiv:2012.09543 [cs]* (Dec. 2020). **http://arxiv.org/abs/2012.09543** arXiv: 2012.09543.

[24] Robert J. Mislevy. 2013. Evidence-centered design for simulation-based assessment. *Military medicine* 178, 10 Suppl (Oct. 2013), 107–114. DOI:**http://dx.doi.org/10.7205/milmed-d-13-00213** Publisher: Oxford Academic.

[25] Milad Moradi, Kathrin Blagec, Florian Haberl, and Matthias Samwald. 2021. GPT-3 models are poor few-shot learners in the biomedical domain. *arXiv:2109.02555 [cs]* (Sept. 2021). **http://arxiv.org/abs/2109.02555** arXiv: 2109.02555.

[26] Chrystalla Mouza. 2021. Critical Perspectives on using Simulations to Broaden Participation in Computer Science through K-12 Teacher Education. (2021).

[27] Elizabeth A. Self and Barbara S. Stengel. 2020. *Toward anti-oppressive teaching: Designing and using simulated encounters.* Harvard Education Press, Cambridge, Massachusetts.

[28] Florence Sullivan, Garron Hillaire, Laura Larke, and Justin Reich. 2020. Using Teacher Moments During the COVID-19 Pivot. *Journal of Technology and Teacher Education* 28, 2 (2020), 303–313. **https://www.learntechlib.org/primary/p/216171/** Publisher: Society for Information Technology & Teacher Education.

[29] G. H. Sullivan. 1991. Five "rights" equal 0 errors. *RN* 54, 6 (June 1991), 65–66, 68.

[30] H. Theelen, A. van den Beemt, and P. den Brok. 2019. Classroom simulations in teacher education to support preservice teachers' interpersonal competence: A systematic literature review. *Computers & Education* 129 (Feb. 2019), 14–26. DOI:**http://dx.doi.org/10.1016/j.compedu.2018.10.015**

[31] Meredith Thompson, Kesiena Owho-Ovuakporie, Kevin Robinson, Yoon Jeon Kim, Rachel Slama, and Justin Reich. 2019. Teacher Moments: A Digital Simulation for Preservice Teachers to Approximate Parent–Teacher Conversations. *Journal of Digital Learning in Teacher Education* 35, 3 (July 2019), 144–164. DOI:**http://dx.doi.org/10.1080/21532974.2019.1587727** Publisher: Routledge _eprint: https://doi.org/10.1080/21532974.2019.1587727.