Supporting Crisis Response with Dynamic Procedure Aids

Leslie Wu **Jesse Cirimele** **Kristen Leach** **Larry Chu** **T Kyle Harrison** **Stuart Card** **Scott R Klemmer**
{lwu2 cirimele krleach}@cs.stanford.edu *{lchu Kharriso}@stanford.edu* *scard@cs.stanford.edu* *srk@ucsd.edu*
Stanford University, Stanford, CA 94305-9035
UC San Diego
La Jolla, CA 92093-0436

ABSTRACT

Checklist usage can increase performance in complex, high-risk domains. While paper checklists are valuable, they are static, slow to access, and show both too much and too little information. We introduce *Dynamic Procedure Aids* to address four key problems in checklist usage: ready access to aids, rapid assimilation of content, professional acceptance, and limited attention. To understand their efficacy for crisis response, we created the *dpAid* software system. Its design arose through a multi-year participation in medical crisis response training featuring realistic team simulations. A study comparing Dynamic Procedure Aids, paper, and no aid, found that participants with Dynamic Procedure Aids performed significantly better than with paper or no aid. This study introduces the narrative simulation paradigm for comparatively assessing expert procedural performance through a score-and-correct approach.

Author Keywords: checklists; medicine; cognitive aids

ACM Classification Keywords: H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

CHECKLISTS FOR COMPLEX HIGH-RISK PROCEDURES

With high-risk procedures like surgery and crisis response, errors are easy to make, but consequences are severe. The core problem is complexity. To avoid harm, highly skilled teams working tightly together under significant time pressure must execute many tasks almost perfectly [24]. One study counted 178 tasks per day that must be perfectly executed for the average patient in an Intensive Care Unit [12]. The vast number of known medical conditions also increases complexity. One classification lists 13,600 diagnoses, 6000 drugs, and 4000 medical procedures [46].

The number of surgeries performed globally is about 234 million per year and rising [22]. About half of adverse outcomes in U.S. crisis care are estimated to be preventable, including over 400,000 preventable deaths per year [25]. Many errors arise from complexity-induced breakdowns: missed steps, timing errors, lack of shared mental model, poor resource management [15].

DIS 2014, June 7–11, 2014, Vancouver, BC, Canada.
ACM 978-1-4503-2902-6/14/06.
http://dx.doi.org/10.1145/2598510.2598565

To rein in this expanding complexity, doctors have adopted risk-management techniques from aviation, such as training in simulation, crew resource management, and checklists [16]. Introducing checklists has demonstrably reduced errors for both routine [28,38] and emergency [2,20] tasks. Even simple checklists can substantially reduce adverse events. Introducing checklists in Michigan hospitals decreased infection rates by 66%, saving $175 million and more than 1500 lives in the first 18 months [38]. Across diverse hospitals, checklists led major complications from surgery to drop 36 percent and deaths 47 percent [18,22].

The performance of checklists does not, however, always live up to the promise. Retrieving information can demand additional time and attention [31], and the perception that checklists make medical procedures take longer has slowed their adoption. As Fourcade *et al* put it, “time governs willingness” [43]. There is also a cultural skepticism of checklists as an externally imposed disruption that interferes with medicine [14,18,45]. Medicine is also different from aviation in its team dynamics. Aviation features highly regulated ergonomics: cockpit crews work in teams of two or three with similarly trained participants. In operating rooms (OR), sensors, displays, and interaction points are spread throughout the environment [32,39]. Teams comprise half a dozen or more specialties, each with their own cultures, roles, and equipment.

With organizational support, well-designed checklists can manage complexity, increase safety, and help error recovery [22,44], but poorly designed ones can make things worse. Even Gawande, one of checklists’ foremost promoters, noted the usability failure of his first attempt [18]. In other words, designing good checklists is hard. How can we *reliably* harvest and amplify this potential? To this end, this article makes the following contributions:

- A participatory design process with anesthesiologists to suggest intervention opportunities, cost/benefit tradeoffs.
- *Dynamic Procedure Aids* generalize checklists to expand their benefits, introducing four design concepts: *shared displays* for ready access, *step-at-a-glance* for rapid assimilation, *resources-at-a-glance* for professional acceptance, and *attention aids* for limited attention.
- *dpAid*, a software system for crisis medicine that presents dynamic procedure aids across multiple displays.
- *Narrative simulation* for comparatively assessing expert performance through a score-and-correct approach.
- Empirical results that dynamic aids outperform paper-based aids and non-use of aids.

PARTICIPATORY DESIGN PROCESS AND CONCEPTS

Working with doctors, we attended live surgeries and observed over 50 hours of simulated crises at a state-of-the-art, high-fidelity, operating room simulator. High-fidelity medical simulation places students with confederate nurses and doctors, supported by simulationists who remotely control a patient mannequin [16]. From a control room, we observed dozens medical residents responding to crises and unexpected events, followed by debriefing with instructors.

Our observation focused on OR anesthesiologists, who are responsible for managing emergent events during perioperative patient care. They are trained to recognize and respond to emergencies, taking the role of the crisis team leader. Like pilots, anesthesiologists prepare for the beginning of surgery (takeoff), keep an eye on controls during the procedure (flight), and monitor its completion and initial recovery (landing). Their job is often characterized as hours of boredom punctuated by moments of terror [17].

Participatory design work spanned 16 months and generated more than 60 prototypes at various fidelities. In general, we held weekly design reviews with 3-4 computer scientists and 2-3 doctors or medical professionals. We initially explored tablet-based, general-purpose OR designs. To evaluate designs in a concrete domain, we focused on Advanced Cardiac Life Support (ACLS) [33], because it is important, widespread, and required for professional certification.

The prototypes expressed four stages in our thinking:

➔**Tablet on crash cart…** improves access to aids	➔**Shared large displays showing resources…** improve team acceptance and coordination	➔**Perceptually structured text and graphics…** improve assimilation	➔**Dynamic progressive disclosure…** manages attention and access to information

Initial prototypes were tablet-based. However, reading required walking to the tablet, and only one person could see it. A second set of prototypes added large, mirrored displays. Mindful of acceptance issues, these prototypes integrated doctor-requested resources, such as blood availability, test results, and vitals with history. Doctors reacted negatively to the resource-rich displays, which showed too much. Instead, doctors wanted fast-to-assimilate information. The third series of prototypes, therefore, simplified the display. While successful, this lead to fragmented information, obscuring the larger picture. Therefore the final prototype series used attention-reactive, focus+context techniques to dynamically shift the detail on the display.

Figure 1 Doctor uses dynamic procedure aid (right) in simulator

These participatory design sessions and prototype experience in high-fidelity team simulations highlighted four problem areas (Table 1): *ready access, rapid assimilation*, *professional acceptance*, and *limited attention*. For each, we identified a design concept that reframed the aid to reduce cost or increase benefit. This process yielded dpAid, a Web application showing procedure aids and resources, synchronized across multiple displays. This section and table describe key experiences with prototypes, how dpAid implements each concept, and rationale for why it should help.

1. Ready Access

Problem: Invisible Resources Go Unused

Open-loop communication, misaligned mental models, and invisible work cause many medical errors [34]. Common examples include requests without a specified recipient, lack of acknowledgement or follow-up. For example, "we need to get the crash cart" rather than "Jon can you call for the crash cart", Jon—"yes, I'll get the crash cart". We observed one doctor inform another of an important change in vitals. The other doctor failed to hear, but the first did not notice. As a result, neither realized they held different mental models. This is exacerbated at large hospitals, where team members commonly do not know all their colleagues' names.

Shared artifacts like paper aids [20] and whiteboards [47] facilitate coordination. However, poor ergonomics and static content discourage use. We observed doctors responding to crises would start using paper aids until another task required attention. Then, they would put down the aid, where it would invariably get covered and never picked up again. Other times, doctors would hold a binder of aids in one hand, without a convenient way to make it visible or accessible to others. Consequently, aids were often invisible, hidden physically, or held by only one team member.

Prototype Experience: Multiple Displays

The patient's body provides an important coordination focus [26]. Early prototypes featured a single wall-mounted display. However, because surgical teams are often arranged in a circle around the patient, any single location has blind spots. To address this, we added a second, mirrored display. These displays can be permanently mounted, and/or wheeled in on emergency crash carts (Figure 1).

Emergent Concept: Shared Displays

Large, shared displays can improve awareness and visibility. They provide consistent physical locations for checklists, legible from most locations, supporting common ground [9]. Research in aviation [30] and medicine [29] has connected team performance to shared mental models. Shared visual referents to the procedure, its state, and the resources involved may increase the shared understanding.

2. Rapid Assimilation

Problem: Too Much Information, Much Too Slow

Checklists must be fast to use by someone attending to something *else*. Checklists are rejected when they are slow to use and compete with time and attention needed for the patient [43,45]. Conceptually, it is useful to distinguish rare procedures from common ones. For rare events, checklists provide new or poorly recalled information. Here, checklists must be easy to read. By contrast, for common events, checklists cover routine and familiar material, serving as a reminder to not skip steps or make assumptions too quickly. Here, checklists should be easy to skim, and remind effectively. In between, checklists are used to look-up or confirm facts (*e.g.*, drug doses). In all cases, aids must work well as part of a multi-tasking workflow.

Emergent Concept: Step-at-a-Glance.

A useful way of designing for multi-tasking procedures is to design secondary tasks (in this case, extracting the next step or other information from the checklist) so that either the complete secondary task or a unit step of work on it can be done during the short interval that the main task (in this case, attending to the patient) can be neglected. We call this the *step-at-a-glance* concept. Three design techniques for accomplishing this come out of our prototyping experience.

Prototype Experience: Focus on the current context

In reviewing prototypes, doctors preferred clear, simple presentation of the current step, even when that sacrificed peripheral information. Like turn-by-turn directions, the whole screen can be focused on the current protocol step, both increasing relevant information and reducing cognitive load. While paper is restricted to a static display, software can dynamically change the emphasis of information.

Emergent Concept: Object/Action checklist language.

Early checklists were presented as full sentences with little visual structure [48]. These were slow to read and scan. Because checklists have a highly-constrained structure, visual design can carry more of the information load and improve usability [8]. To continue in this vein, we extracted the basic procedural structure from written descriptions and represented it graphically as appropriate [23]. Increasing visual structure and shortening text speeds reading and improves scanning. We designed a stylized language for re-expressing medical procedures in a compressed object/action format. This language, loosely inspired by aircraft configuration checklists [11], reduces the number of words in a checklist, sometimes substantially. Whenever possible, each step begins with an *object* followed by an *action* or state setting to be achieved for the object. For example, the checklist steps

> Increase FiO_2 to 100%
> Verify ischemia with 12 lead EKG if possible

can be re-expressed as

> **FiO_2:** ↑100%
> **Ischemia:** Verify (Use 12-lead EKG)

We further exploit structure by listing the object to the left, in larger, bold type. This leads to consistent information mapping between content and form. dpAid expands the

Problem	**1. Ready Access**: Hard to find and share	**2. Rapid Assimilation**: Too slow; Hard to multi-task with patient care	**3. Professional Acceptance**: Mixed acceptance discourages use	**4. Limited Attention**: Narrow, scarce attention under stress
Prototype Experience	Paper got put in a corner and ignored. A digital display helped, but not all could view. Usage still limited.	Doctors preferred clear, simple presentation of the current step, even when that sacrificed peripheral information.	Doctors found prototypes showing all potentially useful information to be cluttered.	Prototypes with drug timers and alerts immediately garnered enthusiasm.
Emergent Concept	**Shared Display**: Make aids visible to team through multiple large screens. *Design Shift: Paper* → Multiple shared displays	**Step-at-a-Glance:** Simplify display: Focus on what to do now in context. Speed reading and search. *Design Shift:* Text → Object / Action + Information mapping	**Resources-at a-Glance:** Reframe aids as part of resource management system. *Design Shift:* Checklist → resource management	**Attention Aids:** Direct interface focus dynamically. *Design Shift:* Attention regulator → Attention Aid; Focus+Context
Design Instantiation	Mirror display & interaction across multiple large screens & tablets.	Formulate steps to be found & read in bursts. Progressive protocols. Object/Action, compressed checklist language.	Rapid resource access: personnel, supplies, reference. Aids transition from routine to crisis	Automated drug timers and attentional prompts.
How It Addressed Problem	Provides shared context, facilitates finding checklist, provides more detail	Faster read/skim, search due to: fewer words, stereotyped syntax Procedure step processable in one multi-tasking cycle	Integration incentivizes use; familiarizes and habituates practitioners	Cognitive/procedure aid also serves as attentional aid

Table 1: The four key issues; their induced design shifts, and proposed solution components

focus steps to reveal additional details. Collectively, these treatments seek to increase speed for the several types of procedure reading: direct reading, skimming, and searching.

3. Professional Acceptance

Problem: Bridging the Gap between Promise and Adoption

In addition to increasing speed and reducing error, checklists should foster a safety culture, supporting quality-control and coordination through standardization [10,43]. However, highly-skilled professionals rarely welcome the oversight implied by standardization, despite improved outcomes. Consequently, checklists are underused because some perceive an unfavorable cost/benefit ratio or an unwelcome restriction on professional autonomy. Even in aviation, where checklists are standard, excess checklists reduce compliance [19]. Medical professionals seek better, timely, resource and personnel information [4]. About 39% of surveyed anesthesiologists admitted to having made errors due to lack of medical information found in handbooks, and 74% reported a need for real-time medical knowledge at least monthly [35]. We believe that improving adoption requires tackling these issues head on: reduce the usage costs, expand and emphasize benefits to practitioners.

Emergent Concept: Resource-at-a-Glance

Existing information resources can impede rather than encourage tight collaboration. In one simulation, we observed a resident pull out a smartphone to search for information about a competing diagnosis: malignant hyperthermia vs. thyroid storm. Because the form factor of the information was ill-suited for the device and task, he spent about 5 minutes out of a 20-25 minute crisis reading his device. This illustrates the importance and the difficulty of considering multiple options with current resources.

To address these perceived and actual cost/benefit problems, dynamic aids reframe checklists as a centerpiece of an integrated resource view. To foreground available resources, dpAid shows pictures, names and roles of the team and those on their way, andsupplies like blood that can be requested. Integrated resource visibility may improve decision-making and communication, because as participants gather information, they look to the same screen.

Prototype Experience:
Balance simplicity and amount of information.

Early prototypes presented a laundry list of information resources desired by the doctors, including: inventories of blood, medicine, and supplies; expected availability of laboratory tests; patient identification, medical record highlights and images, procedure site, and plan. This led to a display where, in principle, everything was available but in practice little was findable. Information rich domains face this tension. The challenge was exacerbated by the wall-scale form factor, which requires legibility at a distance.

When the medical team members saw all of this information together during design reviews, they very reasonably found it to be overwhelming. To balance the access/overload tension, the revised dpAid design shows some of this information only when relevant or on request.

4. Limited Attention

Problem: Complex Setting Fragments Team Attention

Crisis response is attention-limited [42]. Anesthesiologists may split visual attention between the patient, vitals and preparing a drug, while ensuring that others continue high-quality CPR. Medical personnel must re-orient physically to attend cognitively and socially. This physically-distributed attention [41] differs from desktop and mobile work.

Prototype Experience: Alerts as a Hook

Administering recurring drugs provides a frequent and important example. Frenetic pacing and multiple responsibilities cause teams to miss doses, forget prior doses or re-dose too often. Some OR personnel rely completely on memory, others use clipboards or whiteboards. Precisely timed attention to multiple activities is difficult for people, but easy for software. We saw timers as a clear, high-value draw to engender broader use. In participatory design sessions, doctors were extremely enthusiastic about integrating timers and other reminders. How might dpAid effectively present these alerts in the chaotic context of the operating theater?

We initially explored audio alarms, as they are agnostic to orientation. However, ORs are extremely noisy. Medical alarms are unregulated, so tones, volume, and frequency vary. Crises make matters worse; the number of genuine and false alarms increases. Con-

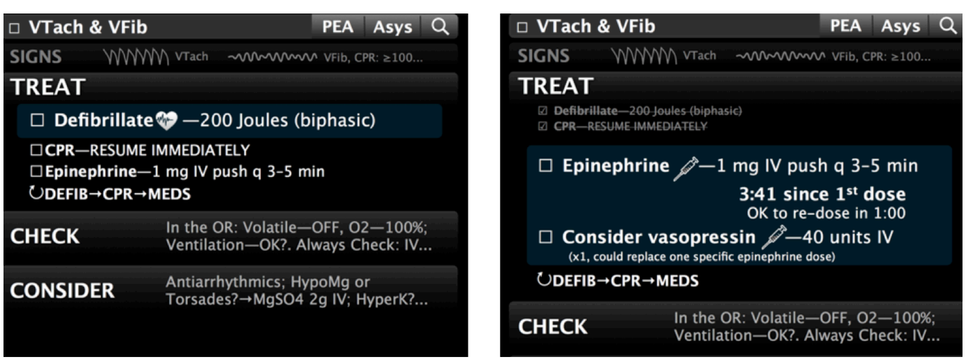

Figure 2: Focus+Context: selecting element in the overview (*left*) reveals additional details (*right*).

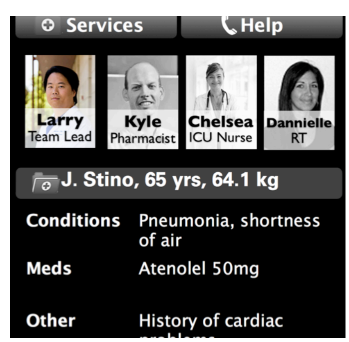

Figure 3. Resources, like patient & team information, make dpAid a "one-stop shop", streamlining attention.

sequently, "demanding" attention through audio is often fruitless. However, medical professionals (like pilots) are trained to cycle rapidly through displays they are monitoring, and a visual alert can be ready for them when they do.

Emergent Concept: Attention Aid

Given these complexities, the design shifted from checklists as attention *regulators* to checklists as attention *aids*. To foreground current state, speed the path to action, and reduce errors, dpAid provides context-specific drug timers and alternate diagnoses to consider. The timers embed a dose and countdown at the relevant aid step, concentrating relevant information where it is needed (see Figure 2). Suggestions such as "consider…" flag similar diagnoses and diagnoses the current condition may evolve into. These suggestions lower the cost of switching to another aid and discourage fixation on initial diagnosis, a common issue under duress [7,16]. Like the timers, dpAid places these suggestions within the aid at the relevant action step.

The dpAid system embodies these design shifts to proactively aid attention and support a rich, shared mental model across a medical team. It facilitates adoption by serving as a resource management system and reduces load through selective emphasis and rapid-read checklists. Here is an example of how dpAid might be used in practice.

Katherine is a resident anesthesiologist paged to the OR. Entering the room, she sees the crash cart with a mounted large-screen display. As she approaches her colleague Justin, he reports they have a 65-year-old patient who came in for laparoscopic knee surgery. They look at dpAid, which displays patient information and personnel (Figure 3).

As they review the vitals and history, the patient's pulse becomes erratic and blood pressure drops. Eventually, the patient is pulseless, resulting pulseless electrical activity (PEA). Katherine asks a nurse to bring up the PEA aid. dpAid reminds her to switch to 100% O_2 and ventilate at 10 breaths/minute. Katherine moves away and gives epinephrine, triggering an on-screen timer to ensure redosing every 3-5 minutes. Meanwhile CPR begins as Justin monitors compression quality. After these immediate actions, Justin and Katherine review possible causes, such as anaphylaxis. They rule out several diagnoses and review other options. Katherine calls for an arterial blood gas, and notices an important electrolyte abnormality. She uses dpAid to verify these numbers and see what resources she can call upon.

DO DYNAMIC AIDS IMPROVE CRISIS DECISIONS?

There are many ways to present medical information aids. To understand the impact of interface presentation style on medical decision-making, we need both better theory and better empirical tools. In particular, we sought a technique that supported rapid, controlled experiments of alternative presentations. This paper addresses this goal through the introduction of the narrative simulation paradigm.

Narrative simulation—inspired by video training [1]—presents a consistently unfolding scenario to all participants. The scenario asks participants questions, records their response, and then reveals recommended best practice. The scenario continues from that action. This cross-participant consistency enables rapid, controlled experiments of how presentation affects medical performance. Participants are asked to verbalize proper procedure under attentional stress and time limits. These scenarios placed single participants in the role of team-leader for cardiac arrest crisis, the role typically responsible for using the aids to support decision-making. This allowed us to test aids and verify their merit before requiring investment into larger teams and expensive simulation. Though not as realistic as a high-fidelity team simulation, narrative simulation incorporates narrative elements that emulate teamwork, for example, involving virtual team members that report vitals and ask for next steps.

Method

37 people (28 MDs, 9 medical students) were recruited from our university to participate in a one-hour study: 20 female and 17 male. Common specialties included: Internal Medicine (8), Anesthesia (7), and Emergency Medicine (7). All were trained in ACLS, which requires re-certification every 2 years. The distribution of recertified participants was: two years ago (4), one year ago (13), in the current year (16), and not yet certified (4). In this hospital, residents run cardiac arrest response teams. There are 2-4 "codes" per month (a "code blue" is used in hospitals to alert staff that a patient requires resuscitation or other immediate attention). On average, each resident participates every few months.

Materials

A pre-study survey asked participants for (expected) graduation year from medical school, specialty, date of first (and most recent) ACLS certification. Participants were counterbalanced based on number of certifications (0, 1, 2+).

This within-subjects experiment compared speed and quality of medical responses in three conditions: with paper aids, with dpAid, and with no aids. We hypothesized that narrative simulation would reflect the attention and time-limited nature of crises, and that dynamic aids would improve participant response quality relative to other conditions.

Paper Cognitive Aids. This condition provided participants with paper ACLS aids. We chose widely-used aids that have been shown to support crisis teams in high-fidelity simulations [48]. These aids were not standard in our hospital, so none of the participants used these aids in their regular work. We printed the paper aids on 8.5"×11" paper and laminated them so they would be sturdy and easy to handle. They were placed on a table nearby, a common practice.

Dynamic Procedure Aids. The dynamic aid, shown on a screen adjacent the scenario display, appeared to respond to scenario events as they happened. These pre-timed interfaces slides were synchronized with the scenario slides, advancing automatically as if a nurse or reader were controlling the interface via a mirrored tablet. The medical content in this condition was substantively equivalent to the paper condition, but presented using Step-at-a-Glance. Content from existing paper aids was divided into 2-4 steps, and the dynamic aid changed the focus step to match the scenario.

Scenario Design and Slide Simulators. This study used narrative encapsulations of authentic medical scenarios, enabling fast and inexpensive medical challenges. Scenarios were designed to test participants' medical knowledge and crisis management decisions under time pressure. These medical scenarios were adapted from online training videos [1] and updated by our medical collaborators. This simulation approach focuses on psychological fidelity over physical fidelity, and is used widely in training [5].

The scenario advanced slides every 5 seconds, revealing information about the patient and unfolding crisis. Each scenario contained 20 to 30 questions like "What is the next important step?" or "What is this [EKG] rhythm?" Participants had 10 seconds to verbally answer each question. Responses after 10 seconds were not counted; speed had no other impact on score. Regardless of response, scenarios revealed a fixed narrative.

Scoring comprised three steps. First, we defined a rubric with the help of a doctor collaborator who teaches medical crisis response. Second, two authors jointly graded 1/3 of participants to align expectations, and split the other 2/3 equally. Partial credit was given as appropriate (*e.g.*, for incorrect dosage but appropriate drug or defibrillation). Finally, answers with non-obvious grades were re-evaluated with the doctor who helped create the rubric.

Experimental Setting & Apparatus. The experimental room was configured with an empty patient bed, a secondary task display, and a scenario screen showing the simulation narrative and questions (Figure 4). In the dynamic condition, an external display showed the dpAid. In the paper condition, participants received laminated paper aids on a table.

Secondary Task. To simulate the additional cognitive load and multi-tasking required in crises, participants had to attend to a secondary task. On a separate screen, a filled circle randomly changed colors from gray to red, yellow, or blue approximately 50 times each scenario. Participants had 10 seconds to press a matching color-labeled key, reverting the color to gray. This induced an additional load on the participant's attention, since they had to turn physically to see the secondary task display. The difficulty of this task was chosen such that participants would uniformly do well.

Procedure

Experimental Sequence. The experiment comprised the following steps: consent form, pre-study survey, training, 3 scenarios, post-scenario surveys, post-study survey, and debriefing. Total study time was 1 hour. Simulation runs were video recorded. Participants were alone—nurses and other doctors were implicitly present in the scenario design.

Training (10 mins). Participants were guided through a 10-minute training period to familiarize them with simulation slides, secondary task, paper cognitive aids, and the dynamic checklists. Participants ran through two abbreviated versions of ACLS slide simulations, first with paper cognitive aids and next with a synchronized dynamic checklist.

Scenarios (3×8 mins). All participants responded to three simulations, always in the same order. These were the progression of medical conditions for each:

Male, 65, Pneumonia: Bradycardia, Asystole, Ventricular Fibrillation (25 questions)

Male, 65, Syncope: Unstable Supraventricular Tachycardia, Ventricular Fibrillation (25 questions)

Female, 78, Unresponsive: Ventricular Fibrillation, Asystole, Ventricular Tachycardia (24 questions)

Conditions. Each participant saw three conditions: dpAid, paper aids, and no aid. Participants saw each condition once; order was counterbalanced using a Latin square design. In the aid conditions, participants were told, "In this condition you will be given access to an aid. It will be located here." They were told aid use was discretionary.

Post-Scenario Self-Assessment (3×1 min). After each scenario, participants filled out a survey on their perceived performance for the scenario and secondary task:

- How many times do you feel like you selected the incorrect color or missed one entirely?
- How many questions do you feel like you missed?
- If you used a cognitive aid/checklist, how much do you feel it changed your score on the questions?

Post-Study Survey & Debrief (10 mins). Participants filled out a survey including demographic information and open response questions about ACLS and checklist experience.

All materials used in the experiment, including the secondary task, surveys, scenarios, aids, and experimental protocols are available at *https://github.com/icogaid/study-2013.*

Statistical Analysis and Data Cleaning

Scores are reported as the percentage of correct trials. Results were compared in R using the lm fixed effects model, a type of linear regression. Unlike the t-test and similar to

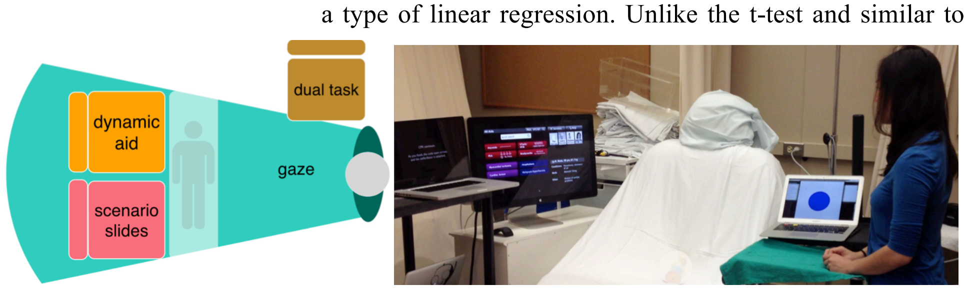

Figure 4: Overhead view of experimental setup with scenario & aid screens (left). Participant uses dynamic aid while responding to questions (right), with color task visible and adjacent.

the ANOVA, linear regression accounts for the probability of multiple pair-wise tests being simultaneously true. Regression models have two benefits over repeated-measures ANOVA. First, fixed-effects linear models are strictly more powerful than an ANOVA because they can handle unbalanced or missing data, but are otherwise equivalent to a multivariate ANOVA. Second, random effects can be added to account for factors such as participant and scenario differences that in practice cannot be exhaustively sampled [3]. This paper primarily uses fixed-effects regression models. Each result report comprises three pieces: first, per-condition averages; second, the effect-size β, indicating the slope difference reported by the mixed effects model; third, the key statistic and *p*-value. Note that β is slightly different than simply subtracting the condition averages because β incorporates the model's estimate of underlying variation in random and fixed effects.

Data Cleaning. 29 of 37 starting participants had usable data for all scenarios: 6 had at least one scenario removed due to synchronization issues; 2 saw incorrect conditions. In the *Pneumonia* scenario, we removed questions 16 to 24 from the analysis after discovering that for many participants, a software bug caused Dynamic Aids not to advance with the scenario. We report results after this data cleaning.

Results

Aid type. Dynamic Aids reduced medical procedure errors. Participants responded correctly significantly more often in the Dynamic condition than in the unaided condition (79.6% vs. 69.1% correct; β=9.46, t(82)=3.3, p<.01); the paper condition was not statistically better than unaided (70.0% vs. 69.1%; β=.30, t(82)=.104, p=.92) (see Figure 5). Moreover, more use of Dynamic Aids correlated with fewer errors (Adj R^2=0.28, F(4,82)=8.01, p<.001).

Analyzing only the first scenario creates a between-subjects comparison that avoids the risk of priming or fatigue effects. With this first-scenario analysis, the effect of Dynamic Aids was even stronger: those using Dynamic Aids responded correctly significantly more often than unaided participants (80.0% vs. 63.6%; β=16.4, t(26)=4.3, p<0.01). Again, there was no significant difference between paper and no aids (67.6% vs. 63.6%; β=3.95, t(26)=.974, p=.34).

Significant factors & interaction effects. To determine what factors were important in predicting scores, we compared several different models. To compare two *lm* models, we used R's *ANOVA* function on pairs of model outputs. A significant ANOVA indicates the two models differ. Incrementally adding and testing factors and interaction effects revealed that scenario, experience level, and experimental condition were all important. There were no significant interaction effects between scenario and experience level, between experience level and experimental condition, and between experimental condition and scenario.

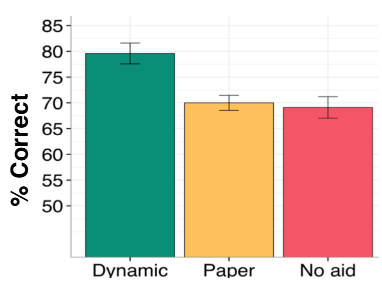

Figure 5. Medical participants scored higher in the *dynamic* condition than in the *paper* and *no aid* conditions.

Scenarios varied in difficulty, as measured by error rate. The *Pneumonia* and *Syncope* scenarios did not differ significantly (β=-1.2, t(82)=-.042, p=.67), but *Unresponsive* was easier than *Pneumonia* (β = 9.1, t(82) = 3.17, p < .01).

Experience. As might be expected, advanced medical personnel (residents and fellows) had more correct trials than medical students when controlling for condition and scenario (74% vs. 67%) (β = 8.3, t(80) = 2.81, p <.01).

Secondary task. Across all scenarios, participants successfully responded to 92% of colors. There was a learning effect: response rates improved as scenarios progressed (88%, 93%, 97%). There was a marginally significant effect of condition on total missed responses on the color task (85 dynamic, 88 none, 115 paper, χ^2(2, n=30)=5.7, p=0.06).

Perceived Utility. In a post-test survey, participants reported both paper and dynamic aids as beneficial. However, participants perceived a larger score increase with Dynamic Aids (15.3%) than paper (4.4%) (t=-4.52, df=56.0, p<.001).

Discussion: Benefits of Dynamic Aids

Dynamic Procedure Aids focus on four key problems: ready access, rapid assimilation, professional acceptance, and limited attention. We discuss observations for each in turn.

Ready Access: paper aids can be tough to find, easy to lose, and inconvenient to hold. Dynamic aids address this through a shared display with context-relevant information and resources. The study found that indeed participants used dynamic aids more than paper ones (mean 22.9 vs. 18.1 times per participant. t=-2.2, df=54, p < .05).

Rapid Assimilation: Current aids are slow to read and search, diverting attention away from the patient. Dynamic aids address this through "step-at-a-glance": cuing attention to the current step, and displaying relevant information. To make steps glanceable, aid content was expressed in a consistent object/action language and layout. Peripheral steps were summarized. Selecting a step as the focus dynamically expands it to present additional details. The secondary task simulated doctors' multiple attentional demands. This dual-task methodology converts attentional load into errors. Consequently, dynamic aids' lower error rate suggests that step-at-a-glance reduced attentional load.

Professional Acceptance: dpAid integrates multiple resources. This presentation appears to have succeeded: participants estimated that Dynamic Aids improved their score by 15.3%; paper aids by 4.4%. This difference is significant (t=-4.52, df=56.0, p<.001). It is important that procedure aids both improve performance and are perceived to do so. These actual and perceived benefits suggest that dynamic aids can facilitate aid acceptance. Because this study relied on volunteers, future work should assess perceived efficacy and directly measure acceptance in the broader community.

Limited Attention: Crises have multifarious activities competing for scarce resources. Attentional overload acutely affects people with less experience, because tasks requires more conscious effort [13]. Consequently, a change in the novice/expert performance spread may indicate a change in the attentional bandwidth required. Improving newcomers' performance is especially important because they commit more errors [36]. In this study, unsurprisingly, doctors had a higher accuracy rate than students (74.5% vs. 67.0%, Figure 6). However, students' performance increased far more in the dynamic condition (21% for students, 7.5% for doctors). This suggests that dynamic aids are more attentionally efficient, providing more headroom for intrinsic task demands. Note that students seemed to outperform residents in the dynamic condition. We hypothesize that students relied more on aids, while residents relied more on experience. When designed well, external representations can be faster and more reliable.

(When) do paper aids help?

Notably, the study found no significant advantage of paper compared to no aid. In contrast, prior studies *have* found increases in team performance and adherence [2,20,48]. We posit three factors for this difference: usage, teams, and training. First, the study assessed discretionary use; aid use was not required. Everyone referred to the dynamic aid at least ten times. By contrast, five participants used paper fewer than ten times–essentially placing themselves in a no-aid scenario. Second, prior work measured teams. This experiment studied individuals. To supplement this, we have run high-fidelity team simulations, finding that teams made crisis response decisions with dpAid. Third, prior work may have provided more training on aids used. This study provided two minutes of training for each aid style. In the debrief, participants reported lack of familiarity as a major impediment to using paper aids. Many had experience with other aids. Given this, it is striking that participants used the digital aid well with minimal training. Aggregating these results with prior work suggests that paper aids are valuable when used, underuse minimizes impact, and that dynamic aids can encourage adoption.

DYNAMIC PROCEDURE AIDS

The study finds that a checklists' design influences effectiveness. We note examples of how digital aids helped.

Dynamic aids track changes in best practices. Medical best practices change frequently, so even a doctor who perfectly remembers medical school may not be up-to-date. Prior to 2010, best practice was to check for pulse and rhythm changes immediately after shock. In newer versions, responders immediately perform post-shock CPR for *all* patients in cardiac arrest, even if they have a pulse [27].

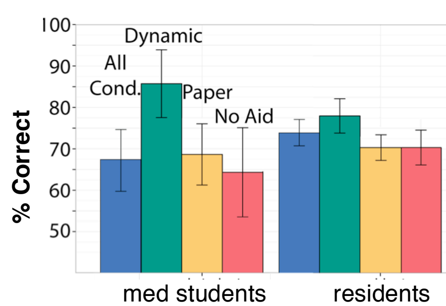

Figure 6: While residents outperformed students, students benefitted more from dynamic aids. Bars are 95% confidence intervals.

Performing CPR before checking for a pulse (the hoped-for outcome of the shock) was counter-intuitive and contrary to prior training for many participants. 24 participants studied ACLS before 2010, learning a dated protocol. The results reflect this: 9 of the 11 participants who saw this in the dynamic condition responded correctly; only 3 of 10 in the paper condition and 2 of 8 in the no aid condition responded correctly. One benefit of digital aids is that revisions can instantly propagate globally as knowledge evolves.

Digital aids provide access to more information. Participants often forgot protocol specifics such as dosing, timing, joules, and appropriate ordering. A dynamic aid provides appropriate detail when needed, with less clutter.

Digital aids can reduce costs & variability of access. Paper aids can be tough to find, easy to lose, and inconvenient to hold. Two different participants dropped paper aids on the floor while trying to use them. Multiple participants missed questions while reading paper aids. Some became so frustrated after first use that they put them down permanently.

Digital aids (and simulation) help the low performers more. An important goal of medical crisis response—and many technology scaffolds—"is to raise up the lowest performers to the level of the average performers" [21]. As we saw, medical students without aids performed the worst, and aids helped their performance dramatically.

Digital aids combine with simulation for effective training. This paper introduced narrative simulation to evaluate time-constrained behavior. Three attributes suggest this approach. First, consistent scenario structure enables comparison across participants. Second, enforced pacing provides an element of realism, and assesses performance under tight time demands. Third, narrative simulation is relatively fast and cost-effective. Our experience is that simulation provides an excellent venue for introducing and evaluating aids. This builds on decades of research in simulation [10,16] and we hope other researchers find it valuable.

Some may worry: do checklists and aids de-skill experts? People as far back as Socrates have worried that knowledge recorded on paper and elsewhere will become a crutch that de-skills memory [37]. However, with checklists as with books, this is not a zero-sum game. People delegate the memory of knowledge to recorded media (when they believe they can access it later) [40]. Given the fragile nature of memory, this is often wise. Concurrently, people strengthen their information search, assessment, and integration skills—improving quality of diagnosis and treatment.

Another worry is that checklists, whether paper or software-based, could increase errors, or change the kinds of errors made. One could overfocus on an aid and respond slowly to unexpected events. A low-ranking staff member charged with reading checklists aloud [6] may feel uneasy speaking up, leading to missed steps or diagnoses. Social challenges aside, checklists have shown to be broadly useful, even though best practices have yet to be formalized. In crises, both paper and software aids have the benefit of being non-blocking, that is, practitioners can chose to attend to other matters if usage is too slow or otherwise non-functional.

Generalizing Dynamic Aids

This paper addresses complex, high-risk procedures, but the principles can be used more broadly. While the focus has been on medicine, the Dynamic Aid interface paradigm is broadly useful for real-time assistive interfaces. For example, driving is also a paced, perilous task. Using Dynamic Aids to analyze a GPS display shows how the same components combine to reduce drivers' attentional burden.

Abstraction	*Surgery*	*Driving*
Shared Display	Mirrored stadium displays w/ crash cart	Car GPS display
Steps-at-a-Glance	Simplify display, focus on current step	Turn by turn instructions
Resource-at-a-Glance	Team names, supplies, lab results	Roads, arrival time, shop locations
Attention Aids	Drug timers	Location-driven display and speech

GPS navigation, unlike paper maps, provides a quickly findable display visible to drivers and passengers. Input is best delegated to those in a support role (passenger). Turn-by-turn reveals directions with step-at-a-glance. Displays provide resources-at-a-glance: estimated arrival time, distance, nearby shops (for gas, cash, or caffeine).

CONCLUSIONS AND LOOKING FORWARD

Deploying aids through software has broad benefits for authoring and distributing best practices. Creating effective checklists requires both medical and design expertise. Encoding best layout practices in software would enable more experts to create and revise checklists. Digital aids also provide a mechanism for automatic logging and recording.

Designing tools to support crisis response can be a challenge given the pace, risk, multi-tasking and team nature of medicine. Dynamic aids offer the ability to reduce the impedance between a doctor's needs and the information shown, improving adoption and adherence to best practice.

Following on these promising results, further work should be done to look at the effects of dynamic aids on teamwork in high-fidelity simulations including the social impact on team communication and the possibility of distraction. Additional work should look at interaction issues, and in-situ professional acceptance. Finally, practical issues of technology availability and security are worth exploring.

REFERENCES

1. ACLS-Algorithms. MegaCode. 2012. http://acls-algorithms.com/acls-megacode-simulator.
2. Arriaga, A.F. et al. Simulation-based trial of surgical-crisis checklists. *N Engl J Med.*, (2013).
3. Baayen, R.H., Davidson, D.J., and Bates, D.M. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, (2008).
4. Bardram, J.E., Hansen, T.R., and Soegaard, M. AwareMedia: A shared interactive display supporting social, temporal and spatial awareness in surgery. *Proc. CSCW*, (2006).
5. Beaubien, J.M. and Baker, D.P. The use of simulation for training teamwork skills in health care: how low can you go? *Qual Saf Health Care*, (2004).
6. Burden, A.R. et al. Does every code need a "reader?" improvement of rare event management with a cognitive aid "reader" during a simulated emergency: a pilot study. *Simulation in Healthcare*, (2012).
7. Burian, B.K. Design Guidance for Emergency and Abnormal Checklists in Aviation. *Proceedings of the Human Factors and Ergonomics Society*, (2006).
8. Chu, L. and Fuller, A. *Appendix C: Crisis Management Cognitive Aids in Manual of Clinical Anesthesia*. Lippincott, 2011.
9. Clark, H.H. and Brennan, S.E. Grounding in Communication. In L.B. Resnick, J.M. Levine and S.D. Teasley, eds., *Perspectives on Socially Shared Cognition*. American Psychologcal Association, Washington, D.C., 1991.
10. Degani, A. and Wiener, E.L. Cockpit Checklists: Concepts, Design, and Use. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, (1993).
11. Degani, A. *On the typography of flight-deck documentation*. NASA-Ames Research Center, Moffett Field, CA, 1992.
12. Donchin, Y. et al. A look into the nature and causes of human errors in the intensive care unit. *Crit Care Med.*, (1995).
13. Ericsson, K.A. and Lehmann, A.C. Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, (1996).
14. Fourcade, A. et al. Barriers to staff adoption of a surgical safety checklist. *BMJ Quality & Safety*, (2012).
15. Gaba, D.M., Fish, K.J., and Howard, S.K. *Crisis Management in Anesthesiology*. Churchill Livingstone, New York, 1994.
16. Gaba, D.M. et al. Simulation-Based Training in Anesthesia Crisis Resource Management (ACRM): A Decade of Experience. *Simulation & Gaming*, (2001).
17. Gaba, D.M. Out of this nettle, danger, we pluck this flower, safety: healthcare vs. aviation and other high-hazard industries. *Simulation in Healthcare*, (2007).

18. Gawande, A. *The Checklist Manifesto: How to Get Things Right*. Metropolitan Books, NY, 2009.
19. Hales, B.M. and Pronovost, P.J. The checklist--a tool for error management and performance improvement. *Journal of Critical Care*, (2006).
20. Harrison, T.K., Manser, T., Howard, S.K., and Gaba, D.M. Use of Cognitive Aids in a Simulated Anesthetic Crisis. *Anesthesia & Analgesia*, (2006).
21. Harrison, T.K. personal communication. 2012.
22. Haynes, A.B. et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med.*, (2009).
23. Horn, R.E. *Mapping hypertext: Analysis, linkage, and display of knowledge for the next generation of on-line text and graphics*. The Lexington Institute of Hospitality Careers, 1989.
24. Hunziker, S. et al. Teamwork and leadership in cardiopulmonary resuscitation. *Journal of the American College of Cardiology*, (2011).
25. James, J.T. A new, evidence-based estimate of patient harms associated with hospital care. *Journal of Patient Safety*, (2013).
26. Kyng, M., Nielsen, E.T., and Kristensen, M. Challenges in designing interactive systems for emergency response. *Proc. DIS*, (2006).
27. Link, M.S. et al. Part 6: Electrical Therapies Automated External Defibrillators, Defibrillation, Cardioversion, and Pacing. In *2010 AHA Guidelines for Cardio-pulmonary Resuscitation and Emergency Cardiovascular Care*. 2010.
28. Makary, M.A. et al. Operating room briefings: working on the same page. *Joint Commission Journal on Quality and Patient Safety*, (2006).
29. Manser, T., Harrison, T.K., Gaba, D.M., and Howard, S.K. Coordination patterns related to high clinical performance in a simulated anesthetic crisis. *Anesthesia & Analgesia*, (2009).
30. Mathieu, J.E. et al. The influence of shared mental models on team process and performance. *The Journal of applied psychology*, (2000).
31. McConnell, D.J., Fargen, K.M., and Mocco, J. Surgical checklists: A detailed review of their emergence, development, and relevance to neurosurgical practice. *Surgical Neurology International*, (2012).
32. Mentis, H.M., O'Hara, K., Sellen, A., and Trivedi, R. Interaction proxemics and image use in neurosurgery. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2012).
33. Neumar, R.W. et al. Part 8: adult advanced cardiovascular life support: 2010 AHA Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*, (2010).
34. Parush, A. et al. Communication and team situation awareness in the OR: Implications for augmentative information display. *Journal of Biomedical Informatics*, (2011).
35. Perel, A. et al. Anaesthesiologists' views on the need for point-of-care information system in the operating room: a survey of the European Society of Anaesthesiologists. *Eur. J. Anaesthesiol.*, (2004).
36. Phillips, D.P. and Barker, G.E.C. A July spike in fatal medication errors: a possible effect of new medical residents. *Journal of general internal medicine*, (2010).
37. Plato, Hamilton, E., and Cairns, H. *The collected dialogues of Plato, including the letters*. Princeton University Press., Princeton, NJ, USA, 1961.
38. Pronovost, P. et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med.*, (2006).
39. Sarcevic, A., Marsic, I., and Burd, R.S. Teamwork Errors in Trauma Resuscitation. *ACM Transactions on Computer-Human Interaction*, (2012).
40. Sparrow, B., Liu, J., and Wegner, D.M. Google effects on memory: cognitive consequences of having information at our fingertips. *Science*, (2011).
41. Srinivasan, N., Srivastava, P., Lohani, M., and Baijal, S. Focused and distributed attention. *Progress in Brain Research*, (2009).
42. Takahashi, Y., Kojima, H., and Okada, K. Injured person information management during second triage. *Proc. UIST*, ACM Press (2011).
43. Verdaasdonk, E.G.G., Stassen, L.P.S., Widhiasmara, P.P., and Dankelman, J. Requirements for the design and implementation of checklists for surgical processes. *Surgical Endoscopy*, (2009).
44. De Vries, E.N. et al. Collaborative Group Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med*, (2010).
45. Winters, B.D. et al. Clinical review: checklists - translating evidence into practice. *Critical Care*, (2009).
46. World Health Organization. *ICD-9-CM International Classification of Diseases*. 2005.
47. Xiao, Y. et al. Cognitive properties of a whiteboard: A case study in a trauma centre. *Proc. ECSCW*, (2001).
48. Ziewacz, J.E. et al. Crisis checklists for the operating room: development and pilot testing. *J. of the American College of Surgeons*, (2011).