Supporting Crisis Response with Dynamic Procedure Aids

Leslie Wu[1], Jesse Cirimele[1], Kristen Leach[1], Stuart Card[1], Larry Chu[2], Kyle Harrison[2], Scott Klemmer[1]

[1]Stanford University HCI Group
Dept. of Computer Science, School of Engineering, Stanford, CA 94305
[2]Stanford University AIM Lab
Dept. of Anesthesiology, School of Medicine, Stanford, CA 94305
{lwu2, cirimele, krleach, scard}@cs.stanford.edu, {lchu, kharriso, srk}@stanford.edu

Abstract

Checklist usage can increase performance in complex, perilous domains. While paper checklists are valuable, they are static, slow to access, and show both too much and too little information. In response, we introduce the Dynamic Procedure Aids approach. Dynamic Procedure Aids address four key problems in checklist usage: ready access to the aids, rapid assimilation of their content, professional acceptance of their use in medical procedures, and the limited attention available to their users. To understand the efficacy of Dynamic Procedure Aids for crisis response, we created dpAid, a software system for crisis medicine. dpAid's design was based on more than a year of observing medical teams responding to simulated crises. We assess our Dynamic Procedure Aids with narrative simulation. A study compared Dynamic Procedure Aids, paper, and no aid conditions, finding that participants with Dynamic Procedure Aids performed significantly better than with paper or no aid.

1. INTRODUCTION

1.1. The Problem of Complex Perilous Procedures

This paper is concerned with *complex perilous procedures* (CPP), such as those arising in surgery and hospital crisis responses. In contrast with the routine cognitive skill of many office tasks [Card, Moran, & Newell 1983], these procedures are at the edge of tractable complexity [Patterson 2007; Rochlin 2005; Gawande 2009]. Errors are easy to make, yet the perilous environment is severely unforgiving of even small errors. Also, complex procedures require intricately coordinated multi-tasking, are team-based, and strongly time-paced. Supporting these procedures requires a different interface paradigm than the classical graphical user interface designed largely for creating and editing office documents.

Arguably, there is no complex perilous procedure domain with more impact than surgery and emergency medical care. The number of surgeries performed globally each year is about 234 million and rising [Haynes 2009]. However, the complexity of surgery and related medical crisis care leads to a higher level of adverse outcomes than necessary [Gawande 2009]. The influential "Harvard study" estimated there are between 44,000 and 98,000 preventable deaths per year [Brennan et al. 1991, Leape et al. 1991]. Other studies estimate about half of adverse outcomes are preventable [Kohn, Corrigan & Donaldson 2000; Davis et al. 2002; Neale, Woloshynowych, & Vincent 2001; Thomas et al. 2000; Vincent, Neale & Woloshynowych, 2000; Dekker 2011]. The core problem is complexity. To avoid harm, many tasks must be executed almost perfectly by highly-skilled teams working tightly together under significant time pressure. One study counted 178 tasks per day for the average patient in an Intensive Care Unit [Donchin et al. 1995]: each task puts the patient at risk. Beyond the need for almost faultless execution, complexity is also driven by the vast number of conditions and remedies. The WHO international disease classification system lists 13,600 diagnoses, 6000 drugs, and 4000 medical and surgical procedures [ICD-9-CM International Classification of Diseases 2005]. Given the expanding complexity associated with medicine doctors have turned to other high-risk domains to look for solutions.

1.2. The Promise of Checklists

To help manage complexity, doctors have begun to adopt risk-management techniques from aviation, such as training in simulation environments, crew resource management, and the use of checklists [Gaba 1994; Gaba 2001]. Aviation, like nuclear power and space flight, is a high-risk, high-tempo CPP domain, where checklists have been studied for decades [Boorman 2001; Burian 2005; Degani 1990; Gawande 2009; Harrison 2006; Haynes 2009; Ziewacz 2011]. Applied to medicine, checklists do, in fact, reduce errors for both simulated medical crisis response [Harrison 2006; Ziewacz 2011] and routine tasks such as pre-surgery setup [Makary 2006] or inserting central lines [Pronovost 2006]. Using even simple checklists can substantially reduce adverse events [Pronovost 2006]. Pronovost documented that without checklists about a third of the time physicians skipped at least one step while putting in a central line (a catheter used to administer fluids). Using a checklist reduced ten-day infection rates from 11% to 0% in his first study. Introducing checklists into Michigan hospitals decreased infection rates by 66%; Pronovost estimated this saved about $175 million and more than 1500 lives in the first 18 months. When extended to hospitals of different types and countries, introducing checklists saw major complications from surgery drop 36 percent and deaths 47 percent [Gawande 2009]. As McConnell writes, checklists are an important "vessel of safety culture" [McConnell 2012]. These benefits generalize to procedurally-organized knowledge, often called cognitive aids [Chu 2011]. Harrison et al. [2006] found that anesthesia care teams improve their performance the more they use such cognitive aids.

Figure 1 Doctor uses dpAid system on a crash cart (right), in high-fidelity operating room simulator.

1.3. Checklist Challenges

The performance of checklists does not, however, always live up to the promise. Cognitively, checklists can sometimes interrupt workflow. For example, finding and searching checklists can induce additional time, attentional demand, and complexity [McConnell, Fargen, & Mocco, 2012]. This procedural interference—and cultural skepticism—has slowed checklists' adoption by medical teams [Gawande 2009; Winters 2009]. As Verdaasdonk et al. [2009] put it, "Time governs willingness and compliance in the use of checklists." In psychological terms, *perceived* time may be the more salient variable.

Though medical checklists have drawn inspiration from aviation, there are important differences between the domains, especially in team composition and work. In aviation, the physical ergonomics are static and highly regulated. Aircrews sit in cockpits where controls and displays are co-designed and co-located [Hutchins 1995]. By contrast, operating rooms (ORs) have sensors, information displays, and interaction points spread throughout the environment [Mentis 2012; Sarcevic 2010]. Cockpit crews work in small teams of two or three, and typically have similar backgrounds. Hospital crisis care teams may comprise surgeons, anesthesiologists, pharmacists, nurses, technicians, and other specialists, arranged *around* the patient, each with its own cultures, roles, and equipment. Not only must staff work under time pressure, risk, and uncertainty, but they must cope with the coordination and communication complexity inherent in team-based crisis care [Hunziker 2011]. These complexities lead to breakdowns in effective crisis care: missed steps, timing errors, lack of a shared mental model, and poor resource management. Checklists are potentially a way to mitigate and recover from these breakdowns, but they must be carefully implemented or they could make things worse.

Even Gawande, one of checklists' foremost promoters, noted the usability failure of his first attempt to make a viable checklist [Gawande 2009]. Furthermore, given that medical checklists are designed as cognitive aids, it is ironic that checklist deployments sometimes give the impression of an externally imposed barrier or disruption, ignorant of skill, wisdom, and context. Thomassen, et al. [2010] note that "Despite the increasing use of checklists in healthcare worldwide, few studies have explored personnel experiences in using this new tool." Those that have (*e.g.*, [Fourcade et al. 2011]) find barriers such as the checklist slowing down or otherwise disrupting the procedure. Though current checklists often provide benefits, their costs have impeded changes in medical practice. This leads us to believe that a stronger benefit:cost ratio could tip the scales.

1.4 Dynamic Procedure Aids

There is considerable evidence that, with organizational support, well-designed checklists can help manage the complexity and increase the safety of medical procedures [Haynes, A. B, et al. 2009; De Vries, E. N. et al., 2010; Haynes, A. B., et al.,2011; & De Vries, E. N., et al., 2011]. How can we *reliably* harvest and amplify this potential? Toward this end, this article makes the following contributions:

- A 16-month participatory design investigation with anesthesiology teams, highlighting intervention opportunities for different roles, and the cost-benefit appeal of integrating checklists with other information resources.
- *Dynamic Procedure Aids* that generalize checklists to expand their benefits and lower their costs. They comprise four key design concepts: *shared displays* for ready access, *steps-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 manifests the key concepts of Dynamic Procedure Aids. dpAid presents a system architecture that includes multiple, mirrored displays and dynamic, software-based cognitive aids.
- The *narrative simulation* paradigm for comparatively assessing expert procedural performance through a score-and-correct approach.
- Empirical results that dynamic aids outperform paper-based aids and non-use of aids.

2. OBSERVATIONS AND DESIGN CONCEPTS

In our 16 months working with doctors, we observed more than 50 hours of simulated crisis scenarios at a state-of-the-art, high-fidelity, medical simulator on campus. We observed medical teams responding to both operating room crises and cardiac arrests. From behind a one-way mirror, we saw dozens of medical residents work with confederate actors to handle complex and unexpected patient crises. We sat in on the post-simulation debriefs organized by the medical teaching staff, as they taught the principles of crisis resource management [Gaba 2001]. Furthermore, we engaged in a number of design critique sessions with participants in these simulated operating rooms and had the opportunity to observe one live surgery.

2.1. Participants and Process

High-fidelity medical simulation offers a unique opportunity to investigate crisis response, without endangering live patients and fewer privacy concerns. These simulations were created to provide a safe, realistic setting for medical education and doctor re-certification [Gaba 2001]. They place one or more students in an operating room with a confederate crew of nurses and doctors are also supported by a team of simulationists behind the scenes, who remotely control patient mannequin responses.

Our observation focused on the practice of OR (Operating Room) anesthesiology. OR anesthesiologists are responsible for managing emergent events during peri-operative patient care. Anesthesiologists are trained to recognize and respond to medical emergencies, and take on the role of team leader once this occurs. Like pilots, anesthesiologists prepare for the beginning of surgery ("take-off"), keep an eye on the controls, the end ("landing"), and have been characterized as having hours of boredom punctuated by moments of terror [Rehmann, et al. 1983; Gaba 2007]. Both fields have different checklists and protocols for routine care versus crisis care.

We created more than 60 different prototypes at various fidelities, sat in on actual surgery, and observed or reviewed video recordings of simulated crises with medical faculty to walk-through user errors and opportunities for software system interventions. To understand the interaction demands, an interactive Web application was developed using HTML and JavaScript. This design probe used WebSockets to synchronize tablet and large-screen displays, and was deployed in two high-fidelity operating room simulations. Initial prototypes addressed general surgery. Later prototypes concentrated on cardiac arrest treatment in a hospital setting as a task domain because of its ubiquity and importance. The medical term for this domain is Advanced Cardiac Life Support (ACLS) [Newmar, R. W., et al. 2010].

2.2. Key Design Concepts

In this process, we identified four problem areas (Table 1) on which to focus: 1) *ready access* (making the aids themselves more rapidly and reliably accessible to the team), 2) *rapid assimilation* (decreasing the time to find and assimilate information from the aid), 3) *professional acceptance* (increasing team acceptance of the aid), and 4) *limited attention* (improving the ability to multitask with the aid). For each of these, we identified a key design concept that shifts the formulation of the aid so as either to reduce the cost or increase the benefit for the medical staff using it. We also give the concrete instantiation of that concept used in dpAid and our design rationale for why it should help.

NEED	KEY CONCEPT	DESIGN INSTANTIATION(S)	HOW IT ADDRESSES PROBLEM
1. Ready Access: Hard to find; Hard to share	**Shared Display**: Make aids visible to team through large-screen display. **DESIGN SHIFT:** Paper to Multiple shared displays	● Mirror display and interaction across multiple large-screens and tablets	● Provides shared context, facilitates finding checklist, provides more detail
2. Rapid Assimilation: Too slow; Hard to multitask with patient care	**Steps-at-a-Glance:** Procedure step processable in one multitasking cycle. Focus on what to do now in abbreviated context. Simplify Display. Speed reading and search. **DESIGN SHIFT:** Text to Object/State + Information mapping	● Reformulation of step to be findable and readable in small bursts. ● Object/Action, compressible checklist language. ● Progressive aid protocols.	● Faster read, skim, search due to: - reduction in number of words - stereotyped syntax - Information mapping ● Processable in small time units for multitasking
3. Professional Acceptance: Mixed acceptance leading to less use	**Resources-at a-Glance:** Reframe checklists as part of a larger, resource management system. **DESIGN SHIFT:** Checklist to Resource Management	● Rapid access to team names, supplies, calculators, reference ● Allow aid to transition from routine to crisis, display additional resources	● Provides incentive to use system, familiarizes and habituates practitioners
4. Limited Attention: Narrow, scarce attention under stress	**Attention Aids:** Direct interface focus dynamically **DESIGN SHIFT:** Attention regulator to Attention Aid Focus+Context	● Automated drug timers and attentional prompts	● Cognitive aid serves as attentional aid

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

2.2.1. Ready Access

Problem: The Invisible Paper Aid.

One fascinating and unexpected observation was that doctors responding to a crisis would often start using a paper cognitive aid until they came across something that needed to be done. Then, they would put the checklist aid down on some flat surface, where almost invariably it would be covered by something else and never picked up again. Other times, doctors would hold the binder containing the aids in one hand, without a convenient place to position the binder so that it was visible and yet accessible. Furthermore, doctors were inconsistent in their use. Given identical scenarios, some doctors never picked up the aid, others looked at it once, and others made personal and/or public use of its information. Consequently, the aid's useful information was often invisible, hidden physically, or held by only one team member.

We also observed "invisible" work practices and mental models. For example, one doctor informed another of an important change in patient vitals, the other doctor failed to hear, but this was not obvious. As a result, neither realized they held different mental models of the situation. Unsurprisingly, coordinated mental models correlate with improved team performance in both aviation [Mathieu 2000] and medicine [Manser 2009].

Key Concept: Shared Displays.

We hypothesize that a large, shared display can mitigate both of these problems. It can provide a consistent physical location, legible from most locations, supporting common ground [Clark and Brennan, 1991; Clark, 1996]—the achievement of "a shared understanding of what is being discussed in a conversation with multiple participants" [Birnholtz et al. 2010]. By providing shared visual referents to the procedure, its state, and the resources involved, the grounding process may be shortened. For example, if the display indicates which drugs were administered, a query about them might be answered with a quick gesture or might not have to be asked at all. In other words, a shared display can make cooperation more *intrinsic* and less *extrinsic,* increasing speed and reducing errors.

In early prototypes our software ran on a single large-screen display mounted on a wall. However, any single location had blind spots for someone. Subsequent prototypes added a second display so that everyone had a clear view. These displays can be permanently mounted in an OR, or brought in on "crash carts" wheeled in during emergency codes. One benefit of a single, shared display is the clarity of what everyone can see—in contrast with personal displays where it's less clear what individuals can see [Wallace et al. 2009]. Synchronizing the two displays retains most of the grounding clarity that a shared display provides. Furthermore, a synchronized view enables input to be driven by an individual, such as a nurse with a tablet. Our process found nurses to be a valuable intervention point because of a professional inclination to process adherence and functional role in organization and support. Doctors could also give verbal commands to nurses for controlling the display.

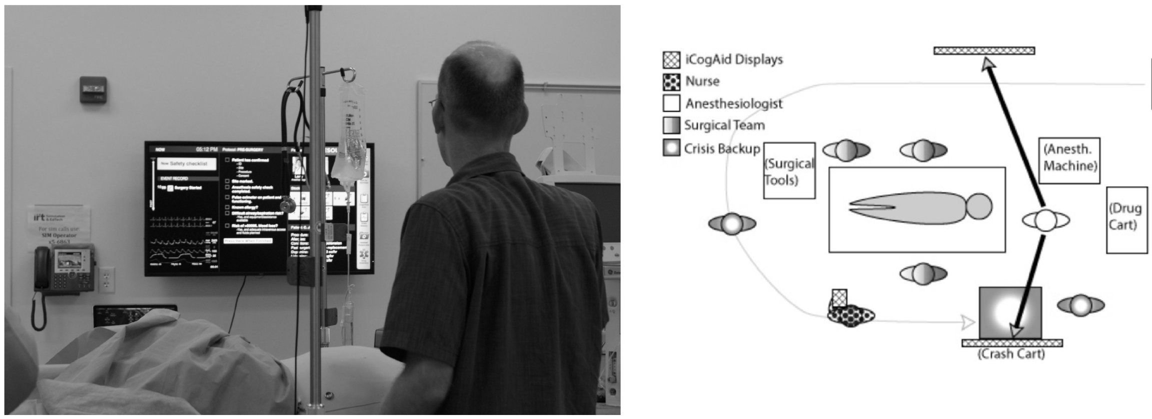
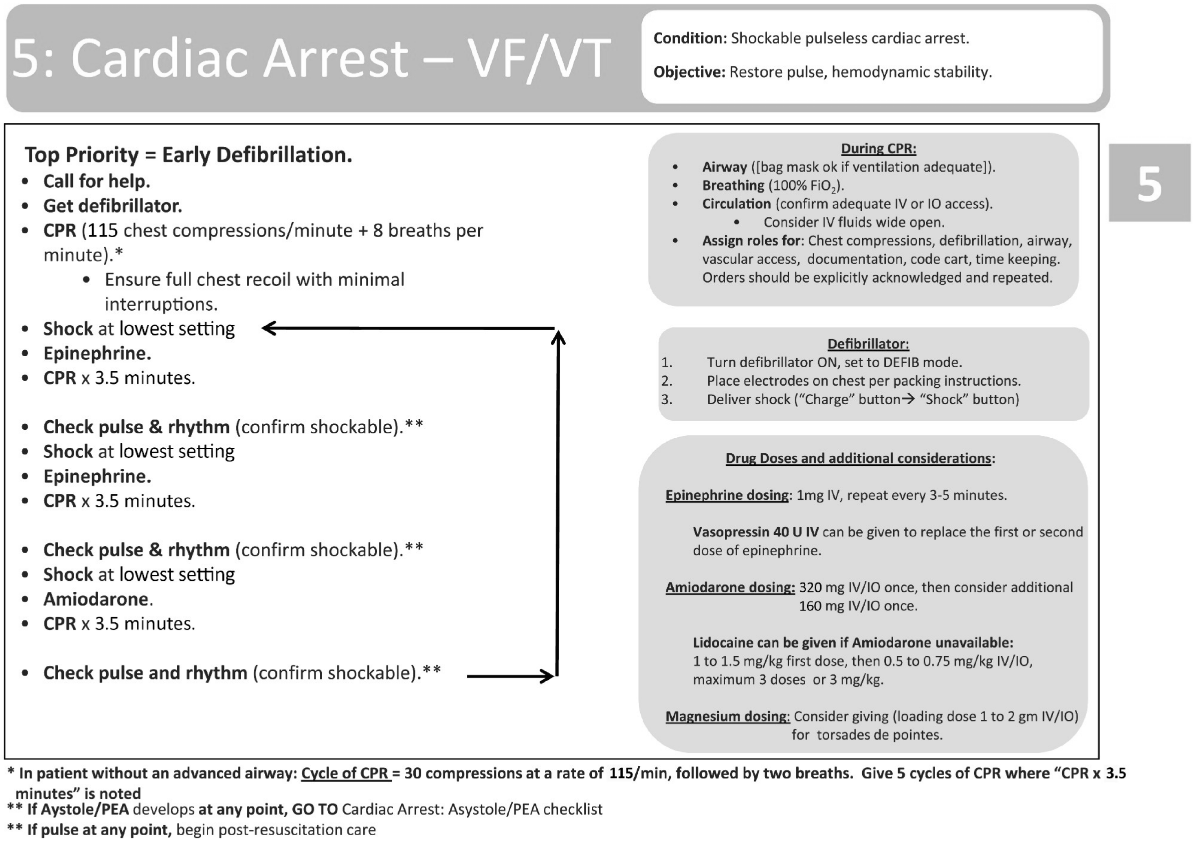


Figure 2. Doctor refers to digital aids on a large-screen display (left); Example OR layout (right)

2.2.2. Rapid Assimilation

Problem: Too Much, Much Too Slow

Many checklist critiques complain that checklists are slow to use and consequently compete with time and attention needed for the patient [Kendell, J. and Barthram, C. 1998; Winters, et al. 2009; Verdaasdonk 2009]. Conceptually, it is useful to distinguish rare procedures from common ones. For rare medical events that a team has never experienced, checklists provide new or poorly recalled information. Here, checklists aids must be easy to read. By contrast, for common events, checklists cover routine and familiar material and serve 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 a particular fact, such as a drug dosage. In all cases, to support rapidly shifting visual attention, steps must also be fast to find and to re-find if the reader looks away to attend to something else.



5: Cardiac Arrest – VF/VT

Condition: Shockable pulseless cardiac arrest.

Objective: Restore pulse, hemodynamic stability.

5

Top Priority = Early Defibrillation.

- **Call for help.**
- **Get defibrillator.**
- **CPR** (115 chest compressions/minute + 8 breaths per minute).*
    - Ensure full chest recoil with minimal interruptions.
- **Shock** at lowest setting
- **Epinephrine.**
- **CPR** x 3.5 minutes.

- **Check pulse & rhythm** (confirm shockable).**
- **Shock** at lowest setting
- **Epinephrine.**
- **CPR** x 3.5 minutes.

- **Check pulse & rhythm** (confirm shockable).**
- **Shock** at lowest setting
- **Amiodarone**.
- **CPR** x 3.5 minutes.

- **Check pulse and rhythm** (confirm shockable).**

<u>During CPR:</u>

- **Airway** ([bag mask ok if ventilation adequate]).
- **Breathing** (100% FiO_2).
- **Circulation** (confirm adequate IV or IO access).
    - Consider IV fluids wide open.
- **Assign roles for:** Chest compressions, defibrillation, airway, vascular access, documentation, code cart, time keeping. Orders should be explicitly acknowledged and repeated.

<u>Defibrillator:</u>

1. Turn defibrillator ON, set to DEFIB mode.
2. Place electrodes on chest per packing instructions.
3. Deliver shock ("Charge" button→ "Shock" button)

<u>Drug Doses and additional considerations</u>:

<u>Epinephrine dosing</u>: 1mg IV, repeat every 3-5 minutes.

Vasopressin 40 U IV can be given to replace the first or second dose of epinephrine.

<u>Amiodarone dosing:</u> 320 mg IV/IO once, then consider additional 160 mg IV/IO once.

Lidocaine can be given if Amiodarone unavailable: 1 to 1.5 mg/kg first dose, then 0.5 to 0.75 mg/kg IV/IO, maximum 3 doses or 3 mg/kg.

<u>Magnesium dosing:</u> Consider giving (loading dose 1 to 2 gm IV/IO) for torsades de pointes.

*** In patient without an advanced airway: <u>Cycle of CPR</u> = 30 compressions at a rate of 115/min, followed by two breaths. Give 5 cycles of CPR where "CPR x 3.5 minutes" is noted**

**** If Aystole/PEA** develops **at any point, GO TO** Cardiac Arrest: Asystole/PEA checklist

**** If pulse at any point,** begin post-resuscitation care

Figure 3. Paper checklists provide valuable information. This checklist [Ziewacz 2011] exemplifies how static information presentation can be hard to skim during crisis response.

Key Concept: Steps At a Glance.

A useful way of designing for multitasking [Salvucci & Taatgen 2011; Brumby, Salvucci, & Howes 2007] is to estimate a typical time interval during which the dominant task can be neglected and to design steps of the secondary task so that they can be completed in this turn length [Green 1999]. We introduce the *step-at-a-glance* concept that information artifacts should be designed so that steps can be assimilated in one glance. This chunking speeds use and facilitates attentional shifts when needed. Our participatory design led to three techniques that reduce the time of assimilating a step.

Balance simplicity and amount of information. Our early designs included nearly every piece of information that participants suggested, and consequently suffered from feature clutter. This led to a display where in principle everything was available but in practice little was findable. Technical, information rich domains often face this tension. Our challenge was exacerbated by the wall-scale form factor, which requires clear legibility at a distance. A lesson we learned repeatedly was that clear presentation of less information was much more important than displaying all potentially useful information.

Focus on current context. In reviewing prototypes, doctors strongly preferred a clear and simple representation of the current context, even when that required sacrificing useful but more peripheral information. Like turn-by-turn map directions, the whole screen can be focused on the current step, simultaneously increasing relevant information and reducing cognitive load. While paper is restricted to a static display, software can emphasize currently-needed information (Figure 4), such as a specific treatment protocol. Information that has already been used, is not yet needed, or provides additional explanation for the curious can be minimized by default and expanded if necessary (2). This approach expands the focus+context layout strategy [Card, Mackinlay, & Shneiderman 1999; Bedersen 2000] to procedural documents.

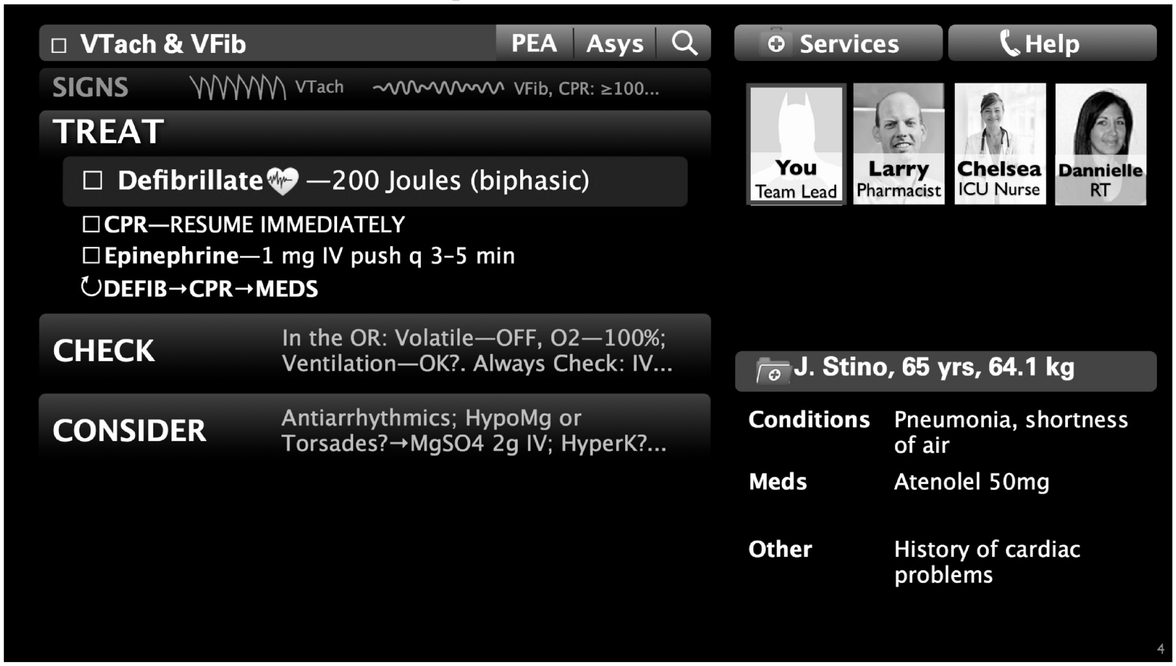


Figure 4. Dynamic Procedure Aid for Ventricular Tachycardia & Ventricular Fibrillation

Object/Action checklist language. Early checklists were presented as full sentences with little visual structure (*e.g.*, Figure 3). As we have mentioned, these were slow to read and slow to scan. Because checklists have a highly-constrained structure, visual design can carry more of the information load and improve usability. Chu's aids leveraged this with richer visual presentation [Chu 2011]. Our work continued in this vein, extracting the basic procedural structure from written descriptions and representing it graphically when appropriate. Increasing visual structure and shortening the text speeds reading and improves scanning. We have designed a stylized language for re-expressing medical procedures in an object/action compressed language (see Figure 4). This language, loosely inspired by configuration checklists for aircraft, reduces the number of words in a checklist, in some cases by half. Whenever possible, each step begins with an object followed by an action or state setting to be achieved for the object. For example, the steps

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

could be re-expressed as

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

We further exploit the structure by listing the object to the left, bold facing it, and giving it larger type, creating a consistent information mapping [Horn 1988] from content to visual form. We furthermore expand the steps of the procedure (see Figure 5) when they are at the point of execution to make available additional subsidiary information. Collectively, these treatments are designed to increase speed for the several types of procedure reading: direct reading, skimming, and searching.

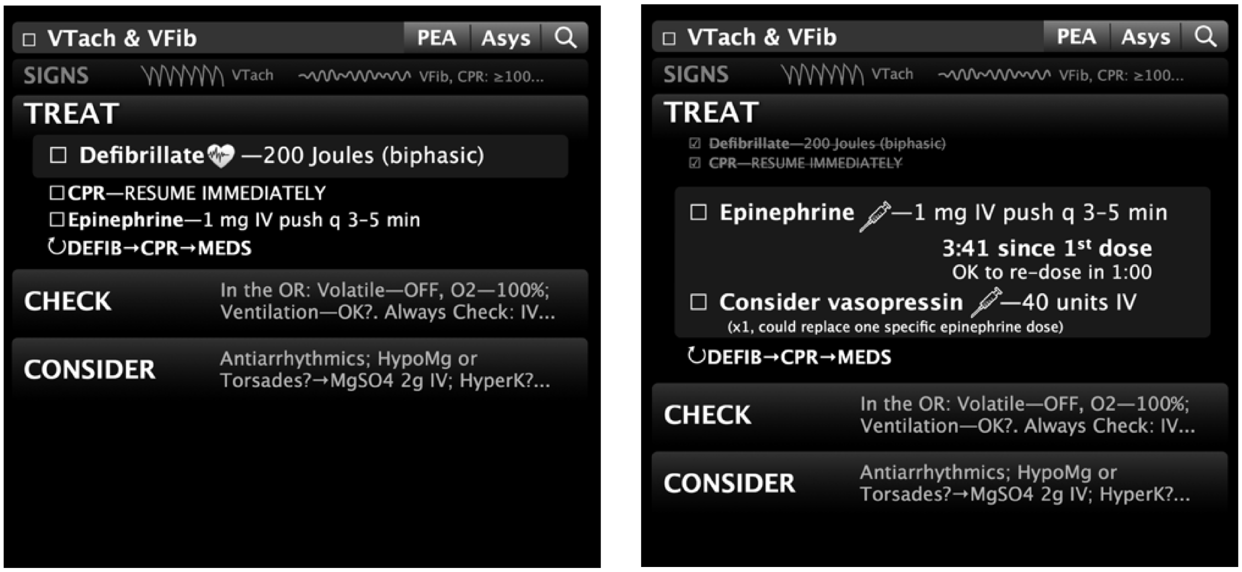


Figure 5. Compressed language combined with variable disclosure: selecting an element in the overview (left) reveals additional details (right).

2.2.3. Professional Acceptance

Problem: Bridging the Gap.

As we have described, current checklist aids often improve outcomes, yet are underused because some perceive an unfavorable cost:benefit ratio or an unwelcome and unwise restriction on professional autonomy. We believe that improving adoption requires tackling these issues head on: expand the benefits, reduce the usage costs, and emphasize the cognitive aid role over the bureaucratic oversight role.

Key Concept: Resource at a Glance

According to literature reviews by [Degani 1990; Degani,1993; Verdaasdonk 2009], checklists should serve the following functions:

- a defense strategy to prevent human errors
- a memory aid to enhance task performance
- standardization of the tasks to facilitate team coordination
- a means to create and maintain a safety culture
- support for quality control by management, government and inspectors

Highly-skilled professionals rarely welcome the oversight implied in the later items of this list, even if this standardization on average improves outcomes. Even in aviation, where checklist use is standardized, too many checklists reduce compliance [Hales 2006]. At the same time, professionals in many fields seek better, timely information. In one simulated crisis we observed, an anesthesia resident pulled out his smartphone to search the Internet for information about a competing diagnosis (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 to 25 minute crisis reading his device. We see this as evidence that bite-sized, contextually-relevant information is a critical need. Therefore, we propose adding to this list another function:

- rapid access to task-relevant information mid-crisis

That is to say, we propose generalizing *checklists* into *procedure aids*.

To address these perceived and actual cost:benefit problems, our work reframes the checklist aid concept to feature them as the centerpiece of an integrated resource view. For example, at a large hospital, team members commonly don't know the names of everyone else on the team, especially when much of their head and face is obscured by their scrubs. This often yields open-loop communication such as "we need to get the crash cart" rather than closed-loop communication (*e.g.*, "Jon can you call for the crash cart", Jon—"yes I will call for the crash cart"). dpAid's shared screen shows pictures and names of people in the room along with information about those on their way to help. This simple cognitive aid makes the social space visible and, potentially, the communication more precise.

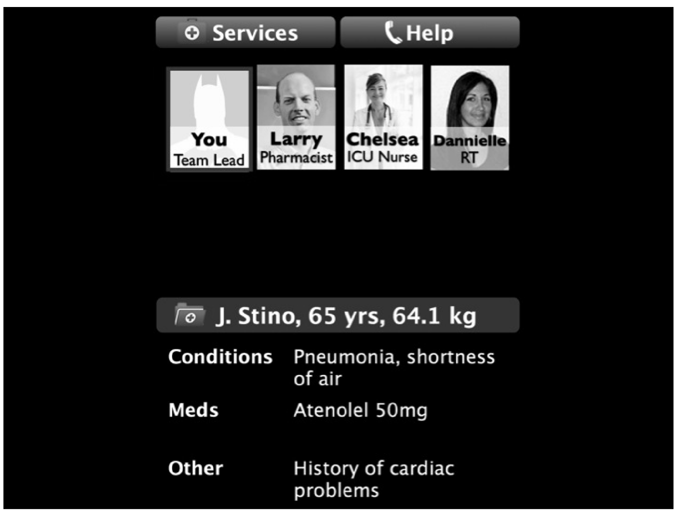


Figure 6. Integrating additional resources, like patient and team information, helps make the dynamic aid a "one-stop shop", encouraging usage.

Our resource manager also explored dashboards showing inventories of blood, medicine, and other supplies available, the expected time to availability of laboratory tests, patient records needed for the procedure, patient identification and procedure site, and the plan of the procedure, names and roles of the operating team, and images useful for the procedure. It can embed medical calculators already initialized to the patient's weight and other parameters. Each time a team member would like to gather resource information, they look to the same screen. By providing an integrated, glanceable view of multiple, commonly-referenced resources, we hope to lower the activation energy for acquiring information, facilitate serendipitous reminding, and create the habit of more frequently consulting these resources.

2.2.4. Limited Attention

Problem: Paced, Multi-surface, Multi-user Attention.

Attention is a major limiting factor during crisis response [Takahashi 2011]. Multiple co-located people work across multiple surfaces on a network of interdependent, important tasks. For example, anesthesiologists may split visual attention between a vitals display, the patient, and a drug vial they are preparing, while simultaneously ensuring that other staff continue high-quality CPR. Medical personnel must orient and attend cognitively, physically and socially. This physically-distributed attention [Srinivasan 2009] differs from desktop [Horvitz 2003] and mobile [Oulasvirta 2005; Iqbal 2010] attentional patterns, and complicates the design of software for crisis teams.

Administering recurring drugs provides a frequent and important example. We observed that frenetic pacing and multiple responsibilities caused medical teams to sometimes forget to miss the time to redose, or forget about a prior dose and redose too often. Some operating rooms rely completely on memory, others have a nurse track dosages on a clipboard or whiteboard. Precisely timed attention to multiple activities is difficult for people, but easy for software. And timers can serve as a clear, high-value draw that in turn engenders broader use.

We initially explored audio alarms, because they are more agnostic to physical orientation. However, operating rooms are extremely noisy: even during routine operation, rock music combines with device alerts, social chat, and work-related discussion. For example, anesthesiologists may be listening to the surgeon while asking a nurse to call for an arterial blood gas, peripherally keeping an ear out for the O2 saturation, but ignoring a false alarm from a different machine. Currently, there are no regulations on how medical alarms should behave, so the alarm tone, volume, and frequency are as varied as the device manufacturers. Crises make matters worse: though social chatter dissipates and music is turned off, the number and frequency of genuine and false alarms increases dramatically, as does the speed and volume of communication. Consequently, "demanding" attention through an audio alert is often fruitless and possibly detrimental. However, medical professionals (like pilots) are trained to cycle rapidly through the dashboard displays they are responsible for, and a visual alert can be ready for them when they do.

In aviation, electronic checklists for routine operation (pre-flight checklists) sometimes mandate step-by-step affirmation. However, the required speed of crisis response make this step-by-step approach unworkable—staff do not want to mark every checklist item when adverse emergent (crisis) events occur. Medical doctors aren't alone in resisting lockstep adherence. Even pilots don't usually check off every item, the so-called READ-DO method [Gawande 2009]. They mostly use the READ-CONFIRM method of performing several items from memory, then consulting the checklist to see if they missed anything. That is to say, professionals naturally modulate their care in response to challenge and risk. For example, to mitigate the extreme hazards of transoceanic flight engender greater diligence, pilots often employ the more cautious READ-DO approach. We hypothesize that enabling this flexibility increases adoption.

Which doesn't mean people always make the right call when left to their own devices, so we must also make the adherence path encouraged and fast. The reader role in medicine [Burden 2012], the WHO surgical time-out [Makary 2006] or the READ-DO or READ-CONFIRM practice in aviation [Gawande 2009] exemplify the attention regulator approach. In this approach, an agent (reader in medicine, co-pilot in aviation) blocks, or regulates progress until some action is taken.

Key Concept: Attention Aid

Given these complexities, our design shifted from checklists as attention *regulators* to checklists as attention *aids*. To help medical teams maintain state, dpAid provides context-specific drug timers and alternate diagnoses to consider. The timers embed dosage and countdown information at the relevant step of the cognitive aid, concentrating relevant information where it's needed (see Figure 5). Suggestions such as "consider ______ aid" flag medically similar diagnoses and diagnoses the current condition may evolve into. These suggestions lower the cost of switching to another aid. Suggestions also seek to discourage fixation on initial diagnosis, a common issue under duress [Gaba 2001; Burian 2006a; Burian 2006b]. Like the timers, dpAid places these suggestions within the aid at the relevant action step to facilitate their use.

2.3. dpAid system design

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 on-call at a large public hospital. She is paged by another anesthesiologist to help an emergent event during a routine knee surgery. As she enters the OR, she sees the crash cart already close to the patient, with a defibrillator and mounted large-screen display. As she approaches her colleague Justin, he tells her that they have a 65-year-old patient who came in for laparoscopic knee surgery. They both look at the dpAid system which displays patient information on the screen (Figure 6). She sees nearby a personnel roster, some of whom she knows and some she does not.*
>
> *As they review the patient vitals and patient history, the patient's pulse becomes erratic and blood pressure drops. Eventually, the patient is pulseless, resulting in a state of pulseless electrical activity (PEA). Katherine asks nurse Kyle to bring up the PEA aid. It reminds her to switch to 100% oxygen and ventilate at 10 breaths/minute. Katherine moves away and pushes epinephrine, which triggers an on-screen timer to make sure it is redosed every 3-5 minutes. Meanwhile CPR begins while Justin monitors compression quality and depth. After these immediate actions are taken, Justin and Katherine begin reviewing possible causes, such as anaphylaxis (allergic reaction) or hypovolemia (loss of blood). They rule out several diagnoses quickly and review several other options to consider. Katherine calls for an arterial blood gas, and later notices an important electrolyte abnormality. She uses dpAid to double-check these numbers and see what additional resources she can call upon.*

3. EVALUATION

To investigate the hypotheses embedded in the dpAid system, we developed the narrative simulation approach which makes use of pre-timed scenario and interface slides. Narrative simulation—inspired by the MegaCode video training materials [MegaCode 2012]—presents scenarios in a linear fashion, no matter how the participant responds. The participant sees a slide-based presentation which automatically advances to tell the patient story. For example they may initially learn that the patient is a 64-year old male with a certain blood pressure and heart rate. Later the heart rate is reported to change. The scenario slide then asks the participant how they will respond. The participants response is recorded and assessed for accuracy. The system then presents the actual action taken in the scenario and continues the story. This linearity and synchronization enables comparison across participants and conditions at each step.

The goal of narrative simulation was to create a fast and inexpensive evaluation that would let us rapidly test cognitive aids. Participants are asked to verbalize proper procedure under attentional stress and time limits. Our narrative simulation scenarios were designed to place the participants in the role of team-lead for cardiac arrest crisis. Participants were asked to make decisions and judgements that team leaders make as part of primary treatment. This allowed us to test aspects of cognitive aid and verify their merit before requiring investment into full medical teams and expensive simulation setup.

3.1. Method

3.1.1. Participants

37 people were recruited from our university to participate in this one hour study: 20 female and 17 male, all trained medical personnel. Participants included 9 medical students, 27 residents, and 2 fellows. Participant specialties were distributed as follows: Internal Medicine (8), Anesthesia (7), Emergency Medicine (7), Undecided (3), Surgery (2), Dermatology (1), Radiology (1), and Urology (1). All were trained in Advanced Cardiac Life Support (ACLS), which requires re-certification is required 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 are responsible for running the cardiac arrest response teams. These skills have a moderate number of opportunities to practice: 2-4 cardiac arrests (codes) per month, in which residents may be involved in only one of these every couple of months.

3.1.2. Materials.

Pre-Study Survey. The pre-study survey asked participants for their (expected) graduation year from medical school, medical specialty, date of first ACLS certification, and date of their most recent ACLS certification. Participants were counterbalanced based on their amount of ACLS training (group 1: zero or one certifications, group 2: two or more).

Paper Cognitive Aids. In this condition, participants were provided with the paper ACLS checklist aids developed by Gawande et al. [Ziewacz 2011] (see Figure 3). These aids were chosen because they had been previously validated in the literature [Ziewacz 2011] and shown to support crisis teams responding to ACLS scenarios in high-fidelity simulations. We printed the paper cognitive aids out on 8.5" x 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. To synchronize with the narrative simulation, the dpAid interface was presented pre-timed interface slides that advanced sequentially. These slides synchronized with the scenario slides, advancing as if a nurse or reader was controlling the interface. The medical content in this condition was substantively equivalent to the paper condition.

Scenario Design and Slide Simulators. This study created narrative encapsulations of authentic medical scenarios, enabling fast and inexpensive medical challenges. Scenarios were designed to test participants' medical knowledge and crisis management. These medical scenarios were adapted from the MegaCode online training videos [MegaCode 2012] and updated by our medical collaborators.

In this slide simulator, scenario slides advance automatically every 5 seconds and reveal information about the patient and how the crisis progresses. During the scenario between 20 and 30 questions appear and participants are given 10 seconds to answer them verbally. Answers given after 10 seconds were not counted but otherwise speed didn't affect score. Regardless of how the user chose to act, a the scenarios remain on a predetermined narrative path. In the dynamic condition, the slide simulator was augmented with the dynamic cognitive aid slides. Since scenarios were created with a fixed sequence of events, these slides could be easily synced with a matching, non-interactive Keynote prototype. The prototype then appeared to respond to scenario events as they happened. This slide simulation approach allowed us to quickly evaluate a proposed interface design with minimal implementation complexity. This resulted in a simpler component-level test that could be easily run and evaluated.

In order to determine correct/incorrect we took three steps. First we generated a rubric with the help of our medical doctor collaborator who regularly teaches and evaluates the crisis response material we were testing. Second two of the authors graded one third of the participants together to align their expectations, and split the other two thirds equally. Finally, participant answers which were not obvious for grading were re-evaluated with the doctor who helped create the rubric.

Experimental Setting and Apparatus. The experimental room was configured with a non-realistic patient mannequin, a laptop displayed the scenario, and a secondary task apparatus. The scenario screen (Figure 6) presented the simulation narrative and prompted the participant with questions. Audio and video were recorded. In the dynamic condition, an external display showed the dpAid. One laptop ran the checklist; a second ran the simulator questions; a third ran the secondary task. In the paper condition, participants were provided with laminated paper aids on a table.

Figure 7: Overhead view of experimental setup with scenario [A] and displayed aid [B] (left). Participant uses dynamic aid while responding to questions (right), with color task visible and adjacent.

Secondary Task. To simulate the additional cognitive load and multitasking required in many crises, the simulation included a secondary task to which participants had to attend . On a separate screen, a filled circle randomly changed colors from gray to red, yellow, or blue approximately 50 times during each scenario. Participants had 10 seconds to press the color-labeled keyboard key corresponding to the correct color, changing it back to gray. The required multitasking created an additional load on the participant's attention, since the participant had to turn physically to see the secondary task display. The difficulty of the color task was chosen such that participants would uniformly do well.

3.1.3. Procedure

Experimental Sequence. The experiment comprised the following steps: consent form, pre-study survey, training,scenarios (*Pneumonia*, *Syncope*, *Unresponsive*), followed by post-scenario surveys, post-study survey, and final debriefing. Total participant time for the study was 1 hour. Simulation runs were audio and video recorded. All participants were exposed to three screen-based simulations, always in the same order. Unlike in a standard Code Blue, participants were alone—nurses and other doctors were implicitly present in the scenario design.

Training (10 mins. Participants were guided through a ten-minute training period to familiarize themselves with the 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.

User Scenarios (3 x 8 mins). The following outlines the sequence of medical conditions presented in each scenario.

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. We had three conditions: The participant could receive dpAid to support the scenario, they could receive the paper aids to support the scenario, or they received no aid for the scenario. Participants saw each condition once. Participant condition order was counterbalanced using a latin square design. Participants in the conditions with aids were told "In this condition you will be given acess to an Aid. it will be located here". They were not explicitly told they had to use the aids.

Post-Scenario Self-Assessment (3 x 1 min). After each scenario participants filled out a survey on their perceived performance for the secondary task and medical scenario response. They were asked to respond numerically to these three questions:

- 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 and Final Debriefing (10 mins). After the three scenarios, participants filled out a final survey, including demographic information (gender), open response questions about ACLS expertise, and about checklist experience. A grading rubric was used to score valid and invalid responses to scenario questions such as “What is the next important step?” or “What is this [EKG] rhythm?” Partial credit was given depending on the timing of the response, and specificity (reduced credit for incorrect dosage but appropriate drug or defibrillation). Graders were research assistants familiar with the scenarios and the appropriate ACLS response. Two different researchers reviewed grades to ensure consistency and accuracy. All materials required to replicate the experiment, including the secondary task, surveys, scenarios, paper aids, Dynamic Procedure Aids, and experimental protocols are available online for download at https://github.com/icogaid/study-2013

3.1.4 Statistical Analysis and Data Cleaning

All scores are reported as the percentage of correct trials. Results were compared using fixed effects modeling, also called linear regression. Our analysis was done in R using the "lm" function. Unlike the t-test and similar to the ANOVA, linear regression is able to account for the probability of multiple pair-wise tests being true at the same time. In addition, these models have two benefits over a repeated measures ANOVA. First, a linear model with only fixed effects can handle unbalanced or missing data, and is otherwise equivalent to a multivariate ANOVA used for repeated measures analysis. Fixed effects linear models are strictly more powerful than an ANOVA. Second, random effects can be added in to account for factors such as participant and scenario differences that in practice cannot be exhaustively sampled [Bayaan 2008]. In this paper we primarily use fixed-effects regression models.

Each reported result 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 the underlying variation in the random and fixed effects.

Data Cleaning. 29 of 37 starting participants participants had data that we could analyze across all scenarios: six had at least one of their scenarios removed due to synchronization issues; two were exposed to incorrect conditions. In the *Pneumonia* scenario, we removed questions 16 to 24 from the analysis after discovering that a software bug that caused Dynamic Procedure Aids to get stuck in the wrong state on those questions for all participants. We report the results after this data cleaning.

3.2. Results

Aid type. Dynamic Procedure Aids reduced medical procedure errors. Participants using Dynamic Procedure Aids responded correctly significantly more often than unaided participants did (79.6% vs 69.1% correct; $\beta = 9.46$, $t(82) = 3.3$, $p<.01$), but those using paper aids were not statistically better than unaided (70.0% vs 69.1%; $\beta = .30$, $t(82) = .104$, $p = .92$) (see Figure 7). Moreover, more use of the Dynamic Procedure Aid correlated with fewer errors (Adj $R^2 = 0.28$, $F(4,82) = 8.013$, $p < .001$).

Looking only at the first experimental scenario creates a between-subjects comparison that avoids the risk of priming or fatigue affecting the data. When we use data only from the first scenario, the effect of Dynamic Procedure Aids becomes even stronger: those using Dynamic Procedure Aids responded correctly significantly more often than unaided participants (80.0% vs. 63.6% correct; β=16.4, $t(26) = 4.3$, $p < 0.01$). Again, there was not a significant performance difference between paper aids and no aids (67.6% vs. 63.6%; $\beta = 3.95$, $t(26) = .974$, $p = .34$).

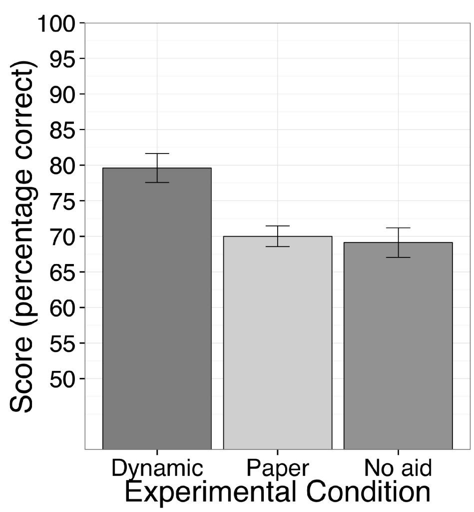


Figure 8. Participants using Dynamic Procedure Aids responded correctly significantly more often than those using paper aids or no aid.

Significant factors and interaction Effects. To determine what factors were important in predicting score we compared several different models. In R, to compare two "lm" models we used the "anova" function on pairs of model outputs. If the anova was significant that indicated that the two models were different. By incrementally adding in factors and interaction effects and testing for significance we found that scenario, experience level, and experimental condition were all important to predicting score but the interaction between scenario and experience level, the interaction between experience level and experimental condition, and the interaction between experimental condition and scenario were all significant. This indicates that there were no significant interaction effects for these factors.

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

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%) ($\beta = 8.3$, $t(80) = 2.81$, $p < .01$).

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 the total missed responses on the secondary color task (85 dynamic, 88 none, 115 paper, $\chi^2(2, n=30)=5.7$, $p=0.06$).

Perceived Utility of Aids. Both paper aids and digital dynamic aids were *perceived* as beneficial by participants, according to the post-test survey. However, participants perceived a larger increase in score when using Dynamic Procedural Aids (15.3%) than paper aids (4.4%) ($t = -4.52$, $df = 56.0$, $p < .001$).

Interaction Effects

3.3. Experimental Discussion

3.3.1. Exploring the benefits of dynamic aids

Overall, participants using Dynamic Procedure Aids responded correctly significantly more often in the simulated medical procedure than those using paper checklists or no aids at all (79.6% to 70.0% and 69.1%). Dynamic Procedure Aids focus on four problem areas of medical checklists: ready access, rapid assimilation, professional acceptance, and limited attention. We discuss observations related to each.

Ready Access: paper aids can be tough to find, easy to lose, and inconvenient to hold. Dynamic aids sought to mitigate this problem by giving participants a shared display that always showed a relevant aid 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-value $< .05$).

Rapid Assimilation: Current aids are slow to read and search, and this diverts important attentional resources away from the patient. Dynamic aids sought to mitigate this problem through its step-at-a-glance design pattern of cuing attention to the current step, and placing all step-relevant information in that one location. To achieve glanceability, we re-expressed the content of aids in an object/action stylized language that places objects and actions at a consistent location. Peripheral steps are shown as a quick summary. When the user selects a step as the focus, it dynamically expands to present additional, context-relevant details. The secondary task simulates the doctor's multiple attentional demands. This dual-task methodology converts attentional load into errors [Martin 2007]. Consequently, the dynamic aids' lower error rate suggests that the step-at-a-glance pattern was effective in reducing attentional load.

Professional Acceptance: Dynamic aids feature a prominent digital display that integrates multiple resources. This prominent presentation appears to have been successful: participants estimated that Dynamic Procedure Aids improved their score by 15.3%; paper aids by 4.4%. This difference is significant ($t = -4.52$, $df = 56.0$, $p < .001$). Belief in an interface's efficacy encourages professional acceptance, and some prior work has been hamstrung by a lack of belief by doctors in this efficacy. Because this paper's study relied on volunteers, it will be important for future work to assess the perceived efficacy in the broader medical community.

Limited Attention: Crises have multifarious activities competing for scarce cognitive resources. Attentional overload more acutely affects people with less training and practice, because each task requires more conscious effort and attention [Ericsson 1996]. Consequently, a change in the novice/expert performance spread often indicates a change in the attentional bandwidth required. Improving newcomers performance is especially important because they commit more errors [Philipps 2010]. In this study, unsurprisingly, doctors had a higher accuracy rate than medical students (74.5% v. 67.0%, see Figure 8). However, medical students performance increased far more with the dynamic aid (21% for students, 7.5% for doctors). This suggests that dynamic aids are more attentionally efficient, providing users with more headroom for the intrinsic demands of the tasks.

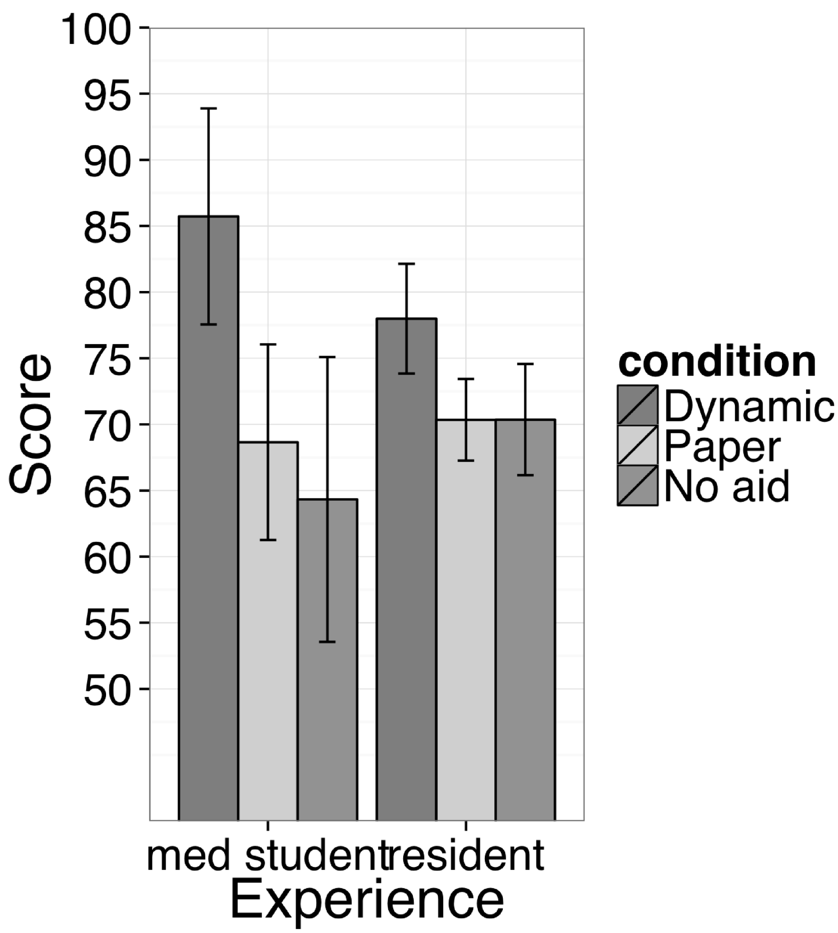


Figure 9: While overall residents outperformed medical students, students received significantly larger benefit from using dynamic aids.

3.3.2. (When) do paper aids help?

Notably, this study found no significant advantage of paper aids compared to no aid. By contrast, several prior studies *have* found increases in team performance metrics and protocol adherence [Harrison 2006; Ziewacz 2011; Arriaga 2013]. We see three likely factors for this difference: usage, teams, and training.

First, this study did not force participants to use any aids, paper or dynamic, since the study was specifically interested in measuring voluntary use. In the dynamic condition, everyone referred to the aid at least ten times. By contrast, in the paper condition, five participants used the aid fewer than ten times. When participants elected not to use the aid, they were essentially placing themselves in a no-aid scenario. A post-hoc t-test comparing infrequent (< 10 uses) and frequent (10 or more) paper aid usage did not find a significant impact of aid usage on score. Future work should explicitly assess the impact of mandated versus voluntary usage.

The second difference is that prior studies have evaluated the impact of cognitive aids on medical teams, which includes the coordination benefits that aids may provide. In this study, we measured individual performance—when by definition there is no team coordination to be done.

The third difference is that prior studies probably offered more training with the particular cognitive aids studied. In this study, participants received about two minutes of training with each aid style. While previous work makes no mention of training time or familiarity, personal communication reports tens of hours of training for [Harrison 2006]. In the debriefing, participants often reported that the factor most inhibiting their use of paper aids was lack of familiarity. Many had different aids that they had practice with and therefore preferred. Given these constraints, it is particularly striking that participants were able to use the digital aid well with such minimal training. Aggregating the results of this study and the prior literature suggests that paper aids are valuable when used, but that underuse may minimize their practical impact, and that digital aids may provide a smoother and more effective adoption path.

3.3.3 Advantages and Limitations of Narrative Simulation

Narrative simulation is a new style of evaluation for cognitive aids and medical skills and knowledge. Other potential evaluation domains used in the medicine are real surgery and high fidelity simulations. For our purposes real surgery is not an option. Real crisis situations are relatively rare (2 - 4 a month in our moderately sized research hospital) so they are hard to be present for. In addition, and perhaps more importantly, we are unwilling to test a novel interface for the first time when a life is on the line. Our discussion here will therefore focus on narrative simulation and high-fidelity simulation.

In contrast to high-fidelity simulation, the goals for narrative simulation are not high realism or perfect understanding, but rather the narrative simulations allow for rapid evaluations of novel tools to get early understanding of strengths and weaknesses as well as to enable rapid design iterations. Thus, one of the biggest strengths of narrative simulation is it's speed. Participants can be run by a single moderator on multiple scenarios in an hour and use of the tool can be observed. In contrast, high fidelity simulations require eight to twelve supporting doctors and staff in order to test two people at a time and require at least an hour for each scenario and debrief.

Another benefit of Narrative simulation is that it allows easy comparison between participants. All participants are asked exactly the same questions at exactly the same time in the simulation after seeing exactly the same information. In contrast to this, high-fidelity simulations are an intricate dance between doctors behind the scenes controlling patient vitals and giving instructions to the doctors and nurses in the room who are playing supporting roles, while the doctors in the hot seat are reacting and making treatment decisions. Much like real life, no two high-fidelity simulations are exactly the same after ten minutes.

Narrative simulations also have some clear limitations in comparison to high-fidelity simulation. Evaluations are on an individual on their role within a team environment, which inherently means that issues around coordination and communication can't be evaluated. High-fidelity simulation is used to study coordination and communication issues in dyads [Manser 2009] and at our research hospital it is used for training and maintenance for larger teams.

The second major limitation of narrative simulations is that they have not been rigorously linked to medical outcomes. In particular, it is not known how performance in narrative simulations relates to performance in high-fidelity simulations. Performance of participants in High-fidelity simulations is often scored by actions that relate to improved patient outcome [Manser 2009, Harrison 2006, Ziewacz 2011]. Performance in Narrative Simulations is scored by answering questions relating to the sequence of treatment steps described by best practice.

While we believe that this study using Narrative Simulation shows significant evidence of the benefits of Dynamic Procedure aids, additional work will be required to definitively show that these aids improve patient outcome and that Narrative Simulations are a consistent and sensitive measurement tool.

4. DYNAMIC PROCEDURE AIDS

This study found that the visual and information design of checklists can influence effectiveness. Showing the right information at the right time is the most effective strategy. It highlights an important limitation of paper checklists, and an opportunity for well-designed software interfaces.

4.1. Benefits of Dynamic Aids

From our study, we note examples of ways that digital aids help.

Digital aids can track changes in best practices and protocols. Medical best practices change frequently, so even a doctor who perfectly recalls their medical-school knowledge may not have an up-to-date response. Take, for instance, cardiac arrest. Here is a setup and question from our study:

> You are 10 minutes into treating a cardiac arrest. The patient’s heart is in ventricular fibrillation, a heart rhythm that can be fixed by defibrillation. In the scenario your team has just shocked the patient and it looks like their heart rhythm has returned to normal. What do you do next?

Two answers are generally given here. The first is to check the patient for a pulse. This answer is given because although the patient’s heart monitor shows a normal (sinus) rhythm, they may not yet have enough blood pressure for a pulse. If the patient has a pulse, he or she is doing well. If not, the patient is still in cardiac arrest. Prior to 2010, best practice was to check for pulse and rhythm changes immediately after shock, but later research showed this was not the best treatment. It is better to immediately perform 2 more minutes of post-shock CPR for *all* patients that have been in cardiac arrest, even if they have a pulse [Link 2010]. This led to a protocol change in 2010. The current protocol answer is to always perform CPR after shocking the patient, with a possible exception when the patient has been in cardiac arrest for less than one minute.

Performing CPR before checking for a pulse (the hoped-for outcome of the shock) is both counter-intuitive and counter to previous training for many participants. And the 24 participants trained in ACLS before 2010 initially learned a protocol that is no longer current. Consequently, it is likely to be performed wrong without a reminder. The results reflect this: 9 of the 11 participants who saw this in the dynamic condition responded correctly; while 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 can provide access to more information. Participants often forgot specifics of the protocol such as dosing, timing, joules, and appropriate ordering. A paper aid has to fit and display all the specifics all the time. A dynamic aid can track the changing scenario and provide appropriate detail in real-time, without the clutter of unnecessary details.

Digital aids can reduce costs and variability of information access. Paper aids can be tough to find, easy to lose, and inconvenient to hold. Two different participants dropped the paper aids on the floor while trying to use them, multiple participants missed questions while trying to look for information in the paper aids, and some participants 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" [Harrison 2012]. 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 the narrative simulation approach for evaluating crisis response. Three attributes led us to this approach. First, the consistent structure of the scenario-response approach enables us to elicit situated medical responses and compare them across participants. Second, the enforced pacing maintains an element of realism in terms of timing, and helps assess and support people's performance under tight time demands. Third, narrative simulation is a relatively fast and cheap technique for training and evaluation. Clearly, higher-fidelity approaches also have value by helping doctors practice motor skills in a physically authentic venue. Our experience has been that simulations provide an excellent venue for introducing and evaluating digital aids. Using aids and simulation together helps both training and research. This insight builds on several decades of research into simulators for crisis response [Degani 1993; Gaba 2001], and we hope future researchers will find it valuable to build on the strategies introduced here.

Since we have discussed the training benefits of dynamic aids, we should address a related worry: will checklists and other aids de-skill experts? People as far back as Socrates have worried that knowledge recorded on paper and other media will become a crutch that de-skills memory [Plato 1961] (though it is only through recorded media do we know this view). However, with checklists as with books, this isn't a zero-sum game. Yes, people "delegate" the memory of some knowledge to recorded media (when they believe they can access it later) [Sparrow 2011]. Given the fragile nature of memory, this is often a wise choice. Concurrently, people strengthen their information search, assessment, and integration skills—improving the quality of diagnosis and treatment.

4.2. Generalizing Dynamic Aids

This paper introduced Dynamic Procedural Aids. Shared displays give procedures a quickly findable location and facilitate communications and coordination for the team. Steps at a Glance allows for rapid assimilation at minimal load of procedure steps while multitasking with the main task. Resources at a Glance allows for rapid access to resources while multitasking. Attention Triage provides support for the allocation of attention.

This interface paradigm responds to the characteristics of complex perilous procedures. It is focused on aiding the execution of activities organized into formal procedures and incorporates the notion of a checklist. By displaying and checking off steps, it should reduce errors. It is team focused and multitasking focused by being glanceable. Instantiating this paradigm for Dynamic Procedure Aids for the OR or Code Blue is roughly

a. **Shared Displays** Mirrored stadium displays using crash cart
b. **Steps at a Glance** Read checklist step in a glance, simplify display, focus on current context, object/action checklist language
c. **Resources at a Glance** Access resource unit in a glance, OR team names, supply stocks, lab orders
d. **Attention AIds** Drug timers

We should emphasize that the Dynamic Aids paradigm can be applied in many settings beyond crisis response. This paper explores and crystallizes the paradigm's components so that their use can be seen in other settings. Perhaps the most frequent paced, perilous task is driving. Using the Dynamic Aids frame to analyze a GPS display enables us to see how these same components coordinate to mitigate drivers' attentional burden. With GPS displays, the components are as follows:

1. **Shared Display** Car GPS display
2. **Steps at a Glance** Turn by turn instructions
3. **Resources at a Glance** Road names, estimated arival time, coffee shop locations
4. **Attention Aids** Display update, spoken turn-by-turn

GPS navigation, unlike paper maps, provides a quickly findable display that can usually be seen by both drivers and passengers. Like dpAid, input is often best delegated to the person in the support role (a nurse or passenger). Turn-by-turn navigation reveals directions with step-at-a-glance. Information readouts provide resources at a glance, like estimated arrival time, current gas mileage, and potential locations to stop (for *e.g.*, gas, money, or coffee). The car's current location is displayed with a large, easily-found marker, helping to triage attention. While driving a car is not nearly as complex as surgery, mistakes often result in death or injury, and a major current concern is with how the use of misconsidered electronic devices, such as GPS systems or smartphones causes driver distraction and leads to accidents. Driving is what we might call a routine perilous procedure. There are many potential designs for reducing attentional load and other benefits. As we have done in this paper, narrative simulation could be used to quickly compare such designs to find the best improvements.

5. CONCLUSIONS AND LOOKING FORWARD

As adoption of smartphones, tablets, and heads-up displays increases, medical practice during emergent events will also continue to evolve. Smartphones can provide doctors with critical information about a patient, serve as a communication channel, and also provide cognitive aids tailored to the situation. Similarly, heads-up displays may one day replace the 20th century pager, and serve as a delivery mechanism for private use of cognitive aids. On the other end of the visibility spectrum, wall-sized displays and pixels everywhere—from digital drapes to wearable computing—provide ways to increase shared understanding and visibility of important information. While our focus has been on medical aids, the Dynamic Aid user interface paradigm was designed to be broadly useful for designing real-time assistive user-interfaces.

Deploying checklists and other cognitive aids through software has broader benefits for authoring, sharing and distributing best practices. One of the major challenges of creating excellent checklists is that someone who is an expert in both medicine and graphic design must individually craft them. This limitation prevents site-specific checklists, impedes their broader creation, and slows their revision as medical knowledge evolves. Encoding best layout practices in software enables more medical experts to participate in checklist creation and revision. And digital distribution can speed their adoption around the world. Digital aids also support automatic recording of medical procedures. Looking further into the future, Dynamic Procedure Aids may help reveal new correlations between treatment and outcome. This additional information could help medical professionals make the best situation-specific decisions.

Designing tools to support crisis response can be a challenge given the paced, high-risk, multitasking and team-reliance of the medical domain. Digital aids offer the ability to reduce the impedance between a doctor's needs and the information shown, to improve adoption, and increase awareness.

ACKNOWLEDGMENTS

We thank Dr. David Gaba, as well as Anna Clemenson and the other members of the Stanford Anesthesia Informatics and Media (AIM) lab, for their guidance and assistance. The photographs are courtesy of (and cc-by) Stanford EdTech. We thank Wendy Mackay, Sara Goldhaber-Fiebert, and Pat Hanrahan for their feedback along the way. Undergraduate researchers Kyle M. Barrett, Katherine Chen, Justin Lee, and Tonya Yu helped build many of the prototypes and run experimental studies. Funding for this project was supported by the Stanford DARE fellowship, the NSF POMI project, and NSF CAREER.

REFERENCES

Ariadne Labs (2013). Operating Room Crisis Checklists. Project Checklist. http://www.projectcheck.org/crisis-checklist-templates.html Downloaded 2013 April 10.

Arriaga, A. F., Bader, A. M., Wong, J. M., Lipsitz, S. R., Berry, W. R., Ziewacz, J. E., ... & Gawande, A. A. (2013). Simulation-Based Trial of Surgical-Crisis Checklists. New England Journal of Medicine, 368(3), 246-253.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of memory and language, 59(4), 390-412.

Bardram, J., Gueddana, S., Houben, S., and Nielsen, S. ReticularSpaces (2012). Proceedings of the ACM annual conference on Human Factors in Computing Systems - CHI '12, ACM Press, 2845.

Bardram, J.E. and Bossen, C. Mobility Work: The Spatial Dimension of Collaboration at a Hospital. (2005). Computer Supported Cooperative Work (CSCW) 14, 2, 131–160.

Bardram, J.E. and Hansen, T.R. Why the plan doesn't hold. (2010). Proceedings of the ACM conference on Computer Supported Cooperative Work (CSCW), ACM Press, 331.

Bickmore, T.W., Pfeifer, L.M., and Jack, B.W. Taking the time to care. (2009). Proceedings of the 27th international conference on Human factors in computing systems - CHI, ACM Press, 1265.

Birnholtz, Jeremy, Ranjan, Abhishek, & Balakrishnam, Ravin (2010). Providing dynamic visual information for collaborative tasks: Experiments with automatic camera control. Human-Computer Interaction 25:261-287.

Boorman, D. Safety benefits of electronic checklists: An analysis of commercial transport accidents. (2001). Proceedings 11th International Symposium on Aviation Psychology, 5–8.

Brennan, Troyen A.; Leape, Lucian L.; Laird, Nan M., et al. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study I. N Engl J Med. 324:370–376, 1991.

Burden, A.R., Carr, Z.J., Staman, G.W., Littman, J.J., and Torjman, M.C. Does every code need a "reader?" improvement of rare event management with a cognitive aid "reader" during a simulated emergency: a pilot study. (2012). Simulation in healthcare : journal of the Society for Simulation in Healthcare 7, 1, 1–9.

Burian, B.K., Barshi, I., and Dismukes, K. The Challenge of Aviation Emergency and Abnormal Situations. (2005). Moffett Field, California.

Burian, Barbara K. (2006a). "Design guidance for emergency and abnormal checklists in aviation." In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, no. 1, pp. 106-110. Sage Publications.

Burian, B. K. (October 2006b). Aeronautical emergency and abnormal checklists: Expectations and realities. In the Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, San Francisco.

Card, S. K., Mackinlay, J. D., and Shneiderman, B. (1999). Information Visualization: Using Vision to Think. San Francisco: Morgan Kaufman.

Card, S. K., Moran, T., & Newell, A. (1983). The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chu, L., Fuller, A. Appendix C: Crisis Management Cognitive Aids in Manual of Clinical Anesthesia, Lippincott, 2011.

Clark, H. H. (1996). Using Language. Cambridge: Cambridge University Press.

Clark, H. H. & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, R. M. Levine, & S. D. Teasley (eds.). Perspectives on socially shared cognition, 127-149. Washington, DC: APA.

Degani, A. and Wiener, E. Human factors of flight-deck checklists: The normal checklist. 1990.

Degani, A. and Wiener, E.L. Cockpit Checklists: Concepts, Design, and Use. Human Factors: The Journal of the Human Factors and Ergonomics Society 35, 2 (1993), 345–359.

De Vries EN, Prins HA, Crolla RM, et al. (2010). SURPASS Collaborative Group Effect of a comprehensive surgical safety system on patient outcomes. N Engl J Med 2010;363:1963–5.

De Vries EN, Eikens-Jansen MP, Hamersma AM, et al. (2011). Prevention of surgical malpractice claims by use of a surgical safety checklist. Ann Surg 2011;253:624–8.

Ericsson, K. A. and Lehman, A. C. (1996). Expert and exceptional performance: Evidence of Maximal adaptation to task constraints. *Annual Review of Psychology* 47:273-305.

Fogarty, J., Ko, A.J., Aung, H.H., Golden, E., Tang, K.P., and Hudson, S.E. (2005). Examining task engagement in sensor-based statistical models of human interruptibility. ACM Press, New York, New York, USA, 2005.

Fourcade, Aude, et al. (2012). Barriers to staff adoption of a surgical safety checklist. BMJ Qual Saf. 21(3):191-197.

Gaba, D.M., Fish, K.J., and Howard, S.K. Crisis Management in Anesthesiology. Churchill Livingstone, New York, 1994.

Gaba, David M., *et al.* (2001). Simulation-Based Training in Anesthesia Crisis Resource Management (ACRM): A Decade of Experience. Simulation Gaming 2001 32: 175

Gaba, David (2007). Out of this nettle, danger, we pluck this flower, safety: Healthcare vs. aviation

and other high-hazard industries. *Simulation in Healthcare* 4:213-217.
Gawande, Atul (2009). The checklist manifesto: How to get things right. New York: Metropolitan Books.
Gergle, Darren, Kraut, Robert E., & Fussell, Susan R., (2013). Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction* 28:1-29.
Harrison, T.K., Manser, T., Howard, S.K., and Gaba, D.M. Use of cognitive aids in a simulated anesthetic crisis. Anesthesia and analgesia 103, 3 (2006), 551–6
Harrison, T.K. (2012) Personal communication.
Haynes AB, Weiser TG, Berry WR, et al. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. N Engl J Med 2009;360:491–9.
Haynes AB, Weiser TG, Berry WR, et al. (2011). Safe Surgery Saves Lives Study Group; Changes in safety attitude and relationship to decreased postoperative morbidity and mortality following implementation of a checklist-based surgical safety intervention. Qual Saf Health Care 2011;20:102–7.
Hiltz, S.R., Diaz, P., and Mark, G. Introduction: Social media and collaborative systems for crisis management. ACM s. on Computer-Human Interaction 18, 4 (2011), 1–6.
Horn, R. E. (1988). Mapping Hypertext: Analysis, Linkage, and Display of Knowledge for the Next Generation of On-Line Text and Graphics, The Lexington Institute, 1989 (Japanese translation published by Nikkei Bu, David W. (2007) Doing Psychology Experiments, Wadsworth Publishing. Chapter 7, page 146.
Hutchins, Edwin. "How a cockpit remembers its speeds." Cognitive science 19.3 (1995): 265-288.
Hunziker, S., Johansson, A.C., Tschan, F., et al. Teamwork and leadership in cardiopulmonary resuscitation. Journal of the American College of Cardiology 57, 24 (2011), 2381–8.
Iqbal, S.T. and Bailey, B.P. Oasis. ACM Transactions on Computer-Human Interaction 17, 4 (2010), 1–28.
Kendell, J., and Barthram,C. (1998). Revised checklist for anaesthetic machines. *Anaesthesia* 53:887-890.
Leape, Lucian L.; Brennan, Troyen A.; Laird, Nan M., et al. The Nature of Adverse Events in Hospitalized Patients: Results of the Harvard Medical Practice Study II. N Engl J Med. 324(6):377–384, 1991.
Link, Mark S., et al. Part 6: Electrical Therapies Automated External Defibrillators, Defibrillation, Cardioversion, and Pacing. 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 122.18 suppl 3 (2010): S706-S719.
Low, D., Clark, N., Soar, J., et al. A randomised control trial to determine if use of the iResus© application on a t phone improves the performance of an advanced life support provider in a simulated medical emergency. Anaesthesia 4 (2011), 255–62.
Mackay, W.E. Is paper safer? The role of paper flight strips in air traffic control. ACM Transactions on Computer-Human Interaction 6, 4 (1999), 311–340.
Makary, M.A., Holzmueller, C.G., Thompson, D., et al. Operating room briefings: working on the same page. Joint Common journal on quality and patient safety / Joint Commission Resources 32, 6 (2006), 351–5.
Manser, T., Harrison, T.K., Gaba, D.M., and Howard, S.K. Coordination patterns related to high clinical performance simulated anesthetic crisis. Anesthesia and analgesia 108, 2009, 1606–15. http://www.ncbi.nlm.nih.gov/pubmed/19372324.
Mark, G., Gudith, D., and Klocke, U. The cost of interrupted work. Proceeding of the twenty-sixth annual Conference on Human factors in computing systems - CHI '08, ACM Press (2008), 107.

McConnell, D.J., Fargen, K. M., & Mocco, J. (2012). Surgical checklists: A detailed review of their emergence, development, and relevance to neurosurgical practice. Surgical Neurologist International 3:2.

MegaCode [2012]. http://acls-algorithms.com/acls-megacode-simulator

Mentis, H.M., O'Hara, K., Sellen, A., and Trivedi, R. Interaction proxemics and image use in neurosurgery. Proceedings the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12, ACM Press (2012), 927.

Neumar RW, Otto CW, Link MS, Kronick SL, Shuster M, Callaway CW, Kudenchuk PJ, Ornato JP, McNally B, Silvers SM, Passman RS, White RD, Hess EP, Tang W, Davis D, Sinz E, Morrison LJ. Part 8: adult advanced cardiovascular life support: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2010;122(suppl 3):S729 –S767.

Oulasvirta, A., Tamminen, S., Roto, V. and Kuorelahti, J. 2005. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05). ACM, New York, NY, USA, 919-928.

LaPorte, Todd R. and Consolini, Paula M. (1991). Working in practice but not in theory: Theoretical challenges of "High Reliability Organizations". Journal of Public Administration Research and theory: J-PART 1(1):19-48.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review 95:492-527.

Nee, Claire & Meenaghan, Amy (2006). Expert decision making in burglars. *Brit. J. Criminol* 46: 939-945.

Patient Safety Strategies - RapidCode. http://www.patientsafetystrategies.com.

Patterson, E.S. Communication Strategies From High-reliability Translation is Hard Work. Annals of Surgery 245, 2 (2, 170–172.

Phillips, D., Barker, G. (2010). A July spike in fatal medication errors: a possible effect of new medical residents. Journal of general internal medicine. Vol. 25(8) p. 774-9.

Plato. The Collected Dialogues: Phaedrus. (1961). Princeton University Press.

Pfeifer Vardoulakis, L., Karlson, A., Morris, D., Smith, G., Gatewood, J., and Tan, D. Using mobile phones to present medical information to hospital patients. Proceedings of the 2012 ACM annual conference on Human Factors in Computing ems - CHI '12, ACM Press (2012), 1411.

Pronovost, P., Needham, D., Berenholtz, S., et al. An intervention to decrease catheter-related bloodstream infection the ICU. (2006). The New England journal of medicine 355, 26 (2006), 2725–32.

Ramachandran, D., Canny, J., Das, P.D., and Cutrell, E. Mobile-izing health workers in rural India. Proc. of the 28th International conference on Human factors in computing systems - CHI '10, ACM Press (2010), 1889.

Rehmann, Jacqueline T., Stein, Earl S., and Rosenberg, Bruce L. (1983). Subjective pilot workload assessment. Human Factors, 24(3), 297-307.

Rochlin, G.I., LaPorte, T.R., and Roberts, K.H. The Self-Designing High-Reliability Organization : Aircraft Carrier ht Operations at Sea. Naval War College Review 40, 4 (2005), 75–90.

Rouse, S. H., Rouse, W. B., and Hammer, J. M. (1982). Design and evaluation of an on-board computer based information system for aircraft. *IEEE Transactions on Systems, Man, and Cybernetics,* 12(4), 451-463, quoted in Degani & Weiner Human Factors of Flight-Deck Checklists: The Normal Checklist. NASA Contractor Report 177549 .

Salvucci, Dario D. and Taatgen, Niels A. (2011). *The Multitasking Mind*. Oxford: The Oxford University Press.

Sarcevic, A., Marsic, I., and Burd, R.S. Does size and location of the vital signs monitor matter? A study of two tra centers. AMIA Annual Symposium proceedings / AMIA Symposium (2010), 707–11.

Sarcevic, A., Palen, L.A., and Burd, R.S. Coordinating time-critical work with role-tagging. Proceedings of the ACM conference on Computer Supported Cooperative Work - CSCW '11, ACM Press (2011), 465.

Sellen, A.J. and Harper, R.H.R. The Myth of the Paperless Office. (2003).

Sparrow, Betsy, Jenny Liu, and Daniel M. Wegner. "Google effects on memory: Cognitive consequences of having information at our fingertips." science 333.6043 (2011): 776-778.

Thomassen, Ø. Et al (2010). Checklists in the operating room: Help or hurdle? A qualitative study on health workers' experiences. BMC Health Services Research 10:342.

Tufte, E. R. (1990). *Envisioning Information*. Cheshire, CT: Graphics Press.

Verdaasdonk, E. G. G., Stassen, L. P. S., Widhiasmara, P. P., Dankelman, J. (2009). Requirements for the design and implementation of checklists for surgical processes. *Surg. Endosc* 23:715-726.

Winters, B. D., Gurses, A. P., Lehmann, H., Sexton, J. B., Rampersad, C. J., & Pronovost, P. J. (2009). Clinical review checklists-translating evidence into practice. Crit Care, 13(6), 210.

Ziewacz, John E. et al. (2011). Checklists for the operating room: Development and pilot testing. J. American College of Surgeons, 213(2): 212-217e10.