

DIL - A Proof of Concept Study to Show the Efficacy of Conversational Agents for Heart Failure Patients

Sanjoy Moulik
University of CA

Samir Chatterjee
Claremont Graduate University

There is an exceptionally high rate of readmissions and rehospitalizations for patients suffering from Heart Failure. Best efforts to address this alarming problem from the Caregiver community have fallen short due to a shortage of trained clinical staff, failure to perform necessary self-management, and money. Using a Design Science Research framework, this work designed and evaluated "DIL" (Sanskrit word for Heart), a Conversational Agent that complements the work of clinicians in achieving the desired behavioral and clinical outcomes. The aim is to provide the hospital with an information system that could bridge the current gap in care that occurs when the patient transitions from the hospital to the home environment. In a pilot study, we show that DIL was able to demonstrate the efficacy and utility as a tool to assist patients with heart failure in improving their self-care.

Keywords: heart failure, MHealth, mobile health, Design Science Research (DSR), chatbot, self-care management, hospital readmission, conversational agent

INTRODUCTION

Motivation

Chronic conditions are increasingly impacting life expectancy and the cost of healthcare all over the world. Among these conditions, Cardiovascular Disease (CVD) stands out because its prevalence continues to rise. According to American Heart Association, Heart Disease and Stroke Statistics 2018 (Benjamin, et.al., 2018) CVD, listed as the underlying cause of death, accounts for nearly 836,546 deaths in the US (Benjamin, et.al., 2018). That's about 1 of every 3 deaths in the US. CVD is the leading global cause of death, accounting for more than 17.9 million deaths per year in 2015, a number that is expected to grow to more than 23.6 million by 2030 (Benjamin, et.al., 2018).

About 92.1 million American adults are living with some form of CVD or the after-effects of stroke (Benjamin, et.al., 2018). Direct and indirect costs of total CVDs and stroke are estimated to total more than \$329.7 billion (Benjamin, et.al., 2018); that includes both health expenditures and lost productivity. CVD and stroke accounted for 14% of total health expenditures in 2013-2014 (Benjamin, et.al., 2018). This is more than any major diagnostic group. Total direct medical costs of CVD are projected to increase to \$749 billion in 2035 (Benjamin, et.al., 2018).

In 2012, the American Heart Association reported that 1 out of 5 of individuals with heart failure dies within a year of diagnosis (Rojer, et.al., 2012). Although CVD is not curable, evidence shows that the quality of life and life expectancy of individuals with heart failure could be improved if the condition is managed by adhering to medications, monitoring symptoms, and low-salt diet intake, etc. (Benjamin, et.al., 2018). Hospital readmissions remain a continued challenge in the care of the heart failure patients. Although small gains have been made over the past 5 years, still more than 20% of patients are readmitted within 30 days and up to 50% by 6 months (O'Connor, 2017). The cost of these readmissions exceeded \$15 billion per year for Congestive Heart Failure (CHF) patients (MedCAP) (Benjamin, et.al., 2018). As readmissions continued, the Center for Medicare and Medicaid Services (CMS) decided to penalize hospitals with high heart failure readmission rates by not reimbursing them for the services they provide since prior studies show that heart failure hospital readmissions could be reduced with quality outpatient services and adherence to care.

Hospitalizations for this chronic progressive disease coupled with high incidences of rehospitalizations within 30 days of initial discharge, averaging 23-24%, are bankrupting the healthcare system with skyrocketing costs (Benjamin, et.al., 2018). Review of current literature clearly identifies that patient non-compliance with medication and lifestyle regimens remains the driving force confounding efforts to stem the rising costs of managing heart failure. The ability of health care providers to impact such patient behaviors is currently handicapped by a break down in patient/provider connection once the patient leaves the walls of either the hospital or the clinic.

Previous studies have tried to address self-care of CHF patients. The MyHeart system (Alnosayan, N., 2015) a Remote Patient Monitoring system was designed, developed and evaluated for the heart failure rehabilitation center at Loma Linda University Medical Center. The project tried to address the current gap in care that occurs when the patient transitions from the hospital environment to the home environment. One aspect of this study that stood out and is relevant to this study is the fact that Human support cannot scale, and is constrained by a number of factors. Even though data was remotely collected from homes, and the system could automatically classify the risks of patients, human health workers were needed to intervene and talk to patients for guidance.

The MyHeart system provided the background, and motivation for the current research to develop an automated intervention system that can be triggered based on patient's risk profile; a software artifact we call DIL that aimed to address the scalability of Human support problem.

There is a tremendous need for effective approaches to manage heart failure and CHF in such a way that deteriorating cases could be detected and treated before they require hospitalization and more specifically rehospitalization. Different strategies like nurse interventions have been tried out, and met with some success, but in practice their implementation remains low, and many patients do not receive self-care support once they are discharged.

The challenge with implementing such interventions could be related to the limited number of caregivers (clinicians, nurses) and the growing number of patients. Therefore, there is a need to identify strategies that have the potential to identify patients at risk and target clinical resources towards them; in particular some novel ICT-enabled strategies could be very beneficial.

Existing health literature articulates the need for research pertaining to heart failure self-care promotion especially a technology-based intervention.

A technology solution that can aid in the process of self-monitoring of vital health data, and provide timely interventions like reminders and motivations for the patients to follow clinical guidelines can go a long way to ease the caregiver burden. Previous studies with Conversational Agents or Chatbots have shown positive outcomes in self-efficacy and attitude change (Baylor, A, 2009).

The purpose of this study was to design, build and evaluate a novel artifact, DIL, a conversational agent for heart-failure patients. A question driving this research was, "Can DIL elicit behavior change, and better adherence to clinical guidelines for the patients?"

Research Objectives

The objective of the research was to design a CA based health-worker that can chat multiple times a week and not only carry on a conversation in an empathetic way, but also document the reasons the patients fail to adhere to clinical and behavioral guidelines. Anonymous recordings of Patient Conversations will allow the Healthcare community develop better adherence strategies in the future. The agent should act as a “buddy” and provide continuous encouragement and motivation to reach desired health goals.

The final objective is of course to evaluate the efficacy and utility of such a CA.

BACKGROUND – LITERATURE REVIEW

A conversational agent (CA) is a software program, which interprets and responds to statements made by users in ordinary natural language. It integrates computational linguistics techniques with communication over the internet. The term CA is primarily used in an academic context by scientists and engineers working on these systems to ensure satisfactory and relevant interaction with the user. They focus on developing interactive embodied conversational agents (ECA) and improving verbal and nonverbal human-computer interaction (CA). A CA can also be considered a dialogue system. It is in essence a computer system intended to converse with a human with a coherent structure. Dialogue systems have employed text, speech, graphics, haptics, gestures, and other modes for communication on both the input and output channel (Dialogue Systems).

The idea of a CA based intervention for chronic disease management is not new. Previous studies have shown large success in achieving behavioral change with this kind of approach. Chatbots also represent opportunities for positive social impact. Chatbots can make needed services more accessible, available, and affordable. They can strengthen users' autonomy, competence, and (possibly counter-intuitively) social relatedness (Folstad, et.al., 2018). CAs can serve as powerful technological mediators to impact motivational outcomes such as self-efficacy and attitude change (Baylor, A, 2009). The CA based intervention has been tried in a variety of settings from Alcohol consumption (Amini, et.al., 2013), promoting health behavior change in heart attacks (Silverman, et.al., 2001), psychosocial intervention deployed on hand-held computers (Johnson, et.al., 2004), hospital discharge nurse that explains written hospital discharge instructions to patients with low health literacy (Bickmore, et.al., 2009), psychotherapy (Lisetti, C. 2008), and as health dialog systems (Lodico, et.al., 2006). Ferguson et. al. proposed CARDIAC: An Intelligent Conversational Assistant for Chronic Heart Failure Patient Health Monitoring, that can conduct regular “checkup” interviews with patients to collect information relevant to their condition (Ferguson, et.al., 2009). Lisetti et. al. proposed an architecture for building an On-Demand Avatar-Based Health Intervention for Behavior Change (Lisetti, et.al., 2012).

Behavior change is a multi-step process. It begins with raising awareness by inspiring the curiosity, engaging the interest, and identifying the motivations of a target audience (Chatbots are Changing...). Through nudges like timed reminders, prompts, and challenges, chatbots can potentially inspire patients to adopt a more positive outlook towards and behavior.

Preliminary Results of a Randomized Controlled Trial on Childhood Obesity using a Text-based Healthcare Chatbot (THCB), (Kowatsch, et.al., 2017) with 15 patients indicate promising results with respect to intervention adherence (13.000 conversational turns over the course of 4 months or 8 per day and patient), scalability of the THCB approach (99.5% of all conversational turns were THCB-driven) and over-average scores on perceived enjoyment and attachment bond between patient and THCB.

According to a survey of 100 practicing physicians all across the United States (Palanica, et.al., 2019), many physicians believed in both costs and benefits associated with chatbots, depending on the logistics and specific roles of the technology. Chatbots may have a beneficial role to play in health care to support, motivate, and coach patients as well as for streamlining organizational tasks; in essence, chatbots could become a surrogate for non-medical caregivers. However, a lot of skepticism remain among the physicians on the efficacy of using Chatbots in self-care management. Based on the findings of the survey results (Palanica, et.al., 2019), many physicians believed that chatbots cannot effectively care for all of the patients' needs (76%, 76/100), cannot display human emotion (72%, 72/100), and cannot provide detailed

diagnosis and treatment because of not knowing all of the personal factors associated with the patient (71%, 71/100). Many physicians also stated that health care chatbots could be a risk to patients if they self-diagnose too often (74%, 74/100) and do not accurately understand the diagnoses (74%, 74/100), (Palanica, et.al., 2019).

Previous studies and literature especially in chronic disease management have clearly outlined a tremendous need for technological interventions especially in the areas of self-care and self-management of chronic diseases to ease the caregiver burden. A recent study showed ninety million Americans have inadequate health literacy, resulting in a reduced ability to read and follow directions in the healthcare environment (Bickmore, et.al., 2009). A large section of these patients fall through the crack, cannot follow through the behavioral guidelines as prescribed because of a lack of understanding, education, social isolation, or simply motivation to do better, and resource constraints of the hospitals and clinics. It is this gap that a CA can address.

In this research, the researcher took inspiration and ideas from the work in various fields as mentioned above and in subsequent pages, and designed and implemented DIL – a CA, in the specific domain of Heart Failure Patients to address the problems and issues discussed above. In particular, this is an agent that worked in conjunction with the 'MyHeart' system (Alnosayan, N., 2015). DIL would review Home monitoring dashboard results, and have conversations with patients. Techniques like Motivational Interviewing (MI) to persuade patients to adhere to clinical guidelines and better behavioral outcomes by using a combination of persuasive messages and pictures, inspirational videos, and healthcare articles.

STUDY METHODOLOGY

Research Questions

The research questions that guided this study was broadly classified under two different areas of interest:

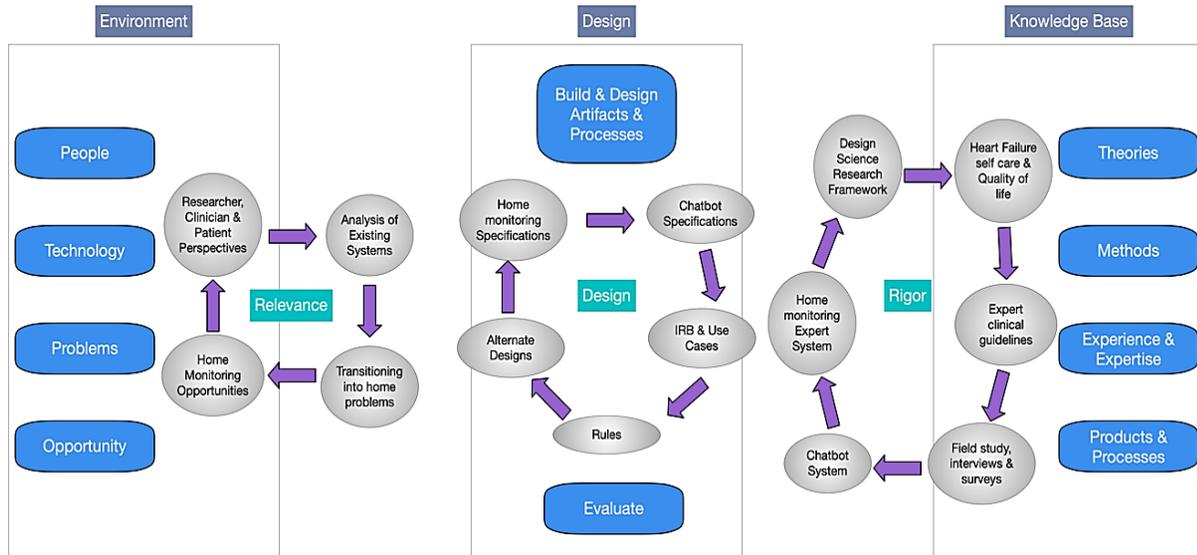
1. Human Computer Interaction:
Research Question 1: How much engagement can DIL elicit? – The system will track and measure the following:
 - i) How many times the patients responded to chat messages from DIL?
 - ii) What was the typical duration of a chat session?
 - iii) A post survey to get qualitative feedback from the patients.
2. Complementing Human Nurses and Case Workers
 - a. *Research Question 2:* Can DIL be a persuasive communication tool? Will the patients be as receptive chatting with DIL as they will be talking with humans? The research compared results from before and after the implementation of the proposed system, and measured improvements in compliance score, to determine the effectiveness of the DIL as an effective tool for persuasion, and elicit compliance to clinical and behavioral guidelines.
 - b. *Research Question 3:* Can an Assistive Technology like DIL lower the Caregiver burden in scaling with a large population? – An exit-survey of the patients to determine the Perceived Helpfulness of DIL was used to gauge how the users felt interacting with DIL.

Research Approach

This study used a Design Science Research (DSR) approach to address the research questions. Hevner and Chatterjee (2010) (Hevner, et.al., 2010), explain that the DSR approach is an iterative process that includes three cycles, namely relevance, design and rigor.

Design Cycle is where building and evaluating the design artifacts happens. Starting from the application context, the relevance cycle starts when the research requirements (opportunity and problem) and the acceptance criteria for evaluation are identified. These requirements and criteria continuously changed as the artifact "DIL" was being built and evaluated. Figure 1 illustrates the three cycles.

**FIGURE 1
DESIGN SCIENCE RESEARCH FRAMEWORK**



Source: (Gregor, Et.Al., 2013)

Design and Build

The Design Artifact, DIL was designed to specifically target Heart Failure patients. It builds on the "MyHeart" system, a collaboration with the Loma Linda University Hospital.

The MyHeart Home Monitoring system (Alnosayan, N., 2015) was geared towards Home Monitoring to promote self-care adherence. Every CHF patient participating in trial when discharged from Loma Linda Hospital were given a set of instructions that may include specific guidelines on diet, exercise and medications, as they transition from Hospital care to Home care. The Home monitoring system deployed a set of tools and took advantage of an Expert System that monitored remotely the patient’s adherence to the prescribed instructions. This tele-monitoring system captured various vital data viz. weight, BP, Pulse rate etc. and updated on a patient dashboard that was then used by the hospital staff, clinicians and nurses to follow up on the progress of these discharged patients.

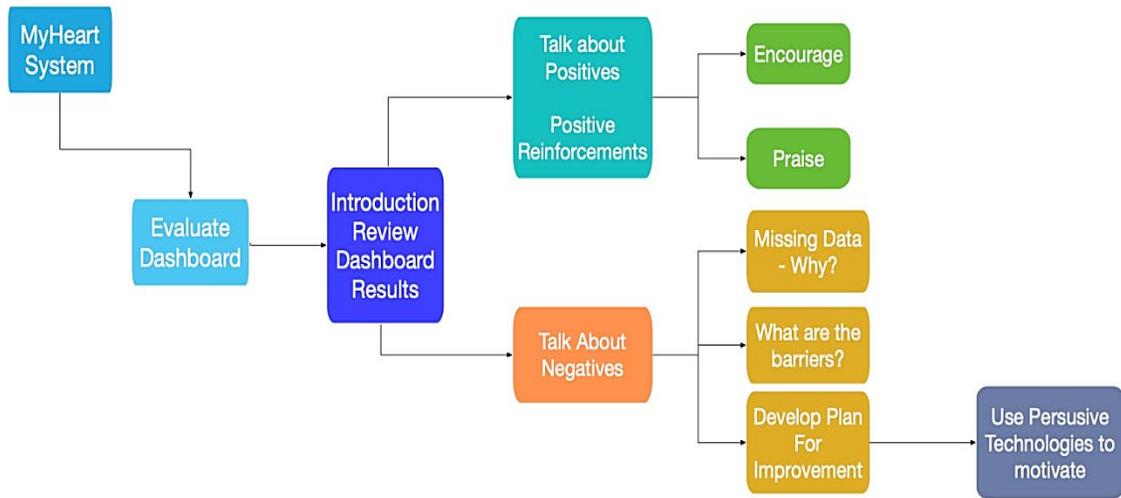
DIL employed the MyHeart system to collect patient health data on a regular basis. The patients were all given a set of home-monitoring devices like a Bluetooth enabled weigh scale and a blood pressure monitor. An app called MyHeart were used daily to provide their symptoms. The CA DIL evaluated the dashboard of patient health data, and used that data for conversation with patients. Figure 2 shows the Home-Monitoring Dashboard.

**FIGURE 2
HOME MONITORING DASHBOARD**

Patient Information													
Patient	Age	Gender	Weight	Systolic	Diastolic	BG	HR	Chest Pain	SOB at Night	Short Breath	Feet Swollen	Fatigue	Risk
56	M		Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Low
61	F		145.80	102	74	Missing	46	0	No	0	0	0	High
56	F		144.20	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Low
71	M		187.40	105	67	142	72	0	No	3	3	2	High
62	M		Missing	95	58	Missing	68	Missing	Missing	Missing	Missing	Missing	Medium
79	F		Missing	Missing	Missing	334	Missing	Missing	Missing	Missing	Missing	Missing	High
57	M		292.20	114	76	227	72	0	No	0	0	2	High
50	M		167.40	98	74	0	92	0	No	0	0	0	High

A review of this dashboard data shows that sometimes patients fail to live up to the required guidelines (as shown by the missing values, and many Medium and High-Risk categories marked in yellow and red) for a variety of factors outlined earlier. The hospital staff want to know the reasons for non-adherence, but struggle to routinely follow up with these large group of patients, for a number of reasons, more specifically lack of resources and money. It is this gap, that DIL hopes to fill. Figure 3 demonstrates DIL Conversation and Process Flow.

**FIGURE 3
DIL CONVERSATION AND PROCESS FLOW**



A conversation between DIL and the patients happened multiple times a week. DIL has the intelligence to carry on a conversation that was tailored to each patient's specific needs. A typical conversation started with a "Welcome" message and a review of the Dashboard results. The conversation then transitioned to the positive aspects of the dashboard results, congratulate the patient on a job well done with positive reinforcements. The next step was a review of the negatives, find out why the patient was unable to fulfill the requirements, and enquire about the barriers. At every step, lessons learned from previous research on

Persuasive Technology, and CAs were used to motivate the patient to do better, and achieve the desired outcome of full compliance to the prescribed instructions at the time of discharge.

The conversations are recorded for further review by a nurse and clinician for two main reasons:

1. To remove any errors in interpretation by the system.
2. To improve the system's future conversational capabilities.

System Architecture

DIL was not envisioned to replace a human intervention, but rather to complement any human intervention. It is one extra tool to address the limited resource problem that is very apparent in current scenario.

The real-world design of DIL incorporated a combined strategy design that started with a flexible design and used exploratory interviews and empirical data to understand the processes and people and is followed by a fixed design phase that incorporates a focused experiment (Robson, C., 2002).

First, it defined the problem, giving due consideration to barriers to self-management and the burden of Heart failure on patients, their families, and society at large. It then demonstrated, through a review of behavioral science theories relevant to the problem of motivating sustained Heart Failure self-management activities as well as existing IS research in the problem domain, the inadequacy of existing solutions to the problem. Thereafter, it developed and presented a purposeful IT artifact that adequately addressed the knowledge gap and solved the problem. The artifact is then evaluated through an experimental study.

Like all Design Science projects, the design and development of DIL also went through multiple iterations, trials and errors. Lessons learned at every stage were incorporated in the design and implementation of the next iteration and release. Valuable feedback on usability and functionality was received from alpha and beta testers who took on various pseudo-patient profiles to make the test scenario as realistic as possible.

The current iteration of DIL had three primary design objectives:

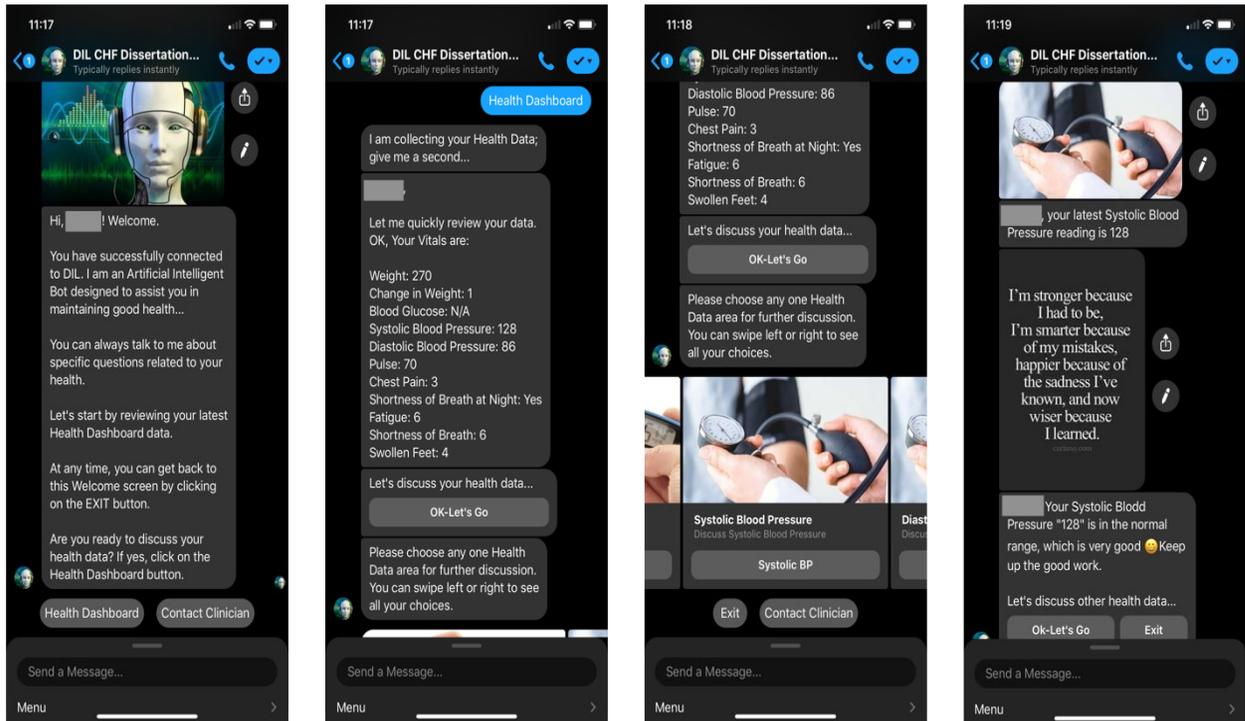
- It must be easy to use
- It can be deployed in the shortest time possible
- It must be easily deployable on a platform that most users are familiar with.

With these design objectives in mind, the researcher chose the Chatfuel development platform, and Facebook Messenger as the deployment platform. Chatfuel is the world's leading chatbot platform for Facebook Messenger; the development environment is mostly drag and drop building blocks that can be build and deployed in really quick time. With more than 2.4 billion active monthly users, Facebook and Facebook Messenger is widely popular and most users are extremely familiar with the platform. This minimizes the learning curve for the users to use such a system. Facebook Messenger also has cross-platform support - it works seamlessly on both iOS and Android based mobile phones and on desktop. Chatfuel offers basic artificial intelligence support, but keeping future needs and visions for the system in mind, provisions were made for Natural Language Processing (NLP) and Artificial Intelligence (AI) module by integrating with Google's AI.API.

So, DIL is in essence a "Chat-Bot" system, that uses the Facebook Messenger platform to communicate with the patient. The fundamental architecture of DIL draws inspiration from [29, 24], and uses Motivational Interviewing (MI), (Miller, et.al., 2002).

According to Nir Eyal in an article in Chatbots Magazine titled, "Bots and A.I. In Behavioral Concept with Nir Eyal", "Of all the ways humans communicate, texting might be the most direct. Text carries less superfluous information than other ways of sending information. With text, there are no voice intonations to decipher or accents to understand, no facial gestures to interpret, and no body language to translate. Text is something computers can understand and process quickly and it's why messaging is a great place for humans and A.I. to work together to serve customer needs." (Eyal, N, 2017). Figure 4 shows screenshots of DIL.

**FIGURE 4
DIL SCREENSHOTS**



Research Design

In this study, the researcher conducted an investigation to test the null hypothesis that there is no difference in health and behavioral outcomes due to intervention of a CA based IT artifact like DIL. Necessary Institutional Review Board (IRB) approvals were obtained from the University.

Participants in the trial were all heart-failure patients at various stages of recovery. They were monitored for two weeks prior to DIL's intervention. These patients all used the Home Monitoring system (Alnosayan, N., 2015). They used the MyHeart (Alnosayan, N., 2015) application in conjunction with Bluetooth enabled self-monitoring devices like a weight scale, a blood pressure monitor and uploaded the data using a tablet to a health dashboard (figure 2). The patients all answered 21 questions of the "Minnesota Living with Heart Failure" questionnaire at the beginning of the trial to gauge an understanding of the quality of life.

DIL intervened at the end of two weeks, and conversed with patients multiple times a week for the next 4 weeks. During this time the patients continued to use the Home Monitoring system, measured and uploaded their health data to the Health dashboard.

At the end of the trial, the Patients once again answered the "Minnesota Living with Heart Failure" questionnaire. They also participated in two surveys - a "DIL User Satisfaction & System Usability" survey and a second survey on the "Perceived Helpfulness of DIL." The Researcher compared the Health dashboard data before DIL's intervention to the dashboard data during intervention.

Quantitative data was collected by comparing Health dashboard data, from the "Minnesota Living with Heart Failure" questionnaire, and from the "DIL Usability survey." Qualitative data was obtained from the "Perceived Helpfulness of DIL" survey. Survey items included patient's feedback on the design and usability of the IT artifact DIL, and questions about patient perceptions and experiences.

To determine if there is a difference in health and behavioral outcomes, a study comparing Health data before and after DIL's intervention was pursued; therefore, the most appropriate research strategy was

mixed-method design. According to Creswell, triangulation mixed-methods design enables the concurrent collection of quantitative and qualitative data and comparison (Creswell, et.al., 2005).

Study Participants

The researcher aimed to show a proof-of-concept at this stage. So, five heart-failure patients were recruited in the study. They were recruited by reaching out to known physicians, friends and family. Each patient completed the entire duration of the trial, which included measurement of weight, blood pressure and pulse readings using Bluetooth enabled weight scale and blood pressure monitors, and answering health symptom questions via the app. The patients used an Android powered tablet and the MyHeart app to collect and upload these data to the Home monitoring dashboard (figure 2) (Alnosayan, N., 2015).

The patients also completed the "Minnesota Living with Heart Failure" questionnaire before and after the trial. At the conclusion of the trial, they participated in two surveys - the "DIL Usability" survey, and the "Perceived Helpfulness of DIL" survey.

Participant Demographics

There were five participants in the trial. The participants ranged in age; they were all male. All the patients belonged to the Asian American community, and lived in Southern California. Two participants were 50+, and three were in the range 45-50. A patient population of only 5 might seem small but according to Jeff Sauro, a small sample size between 5 and 30 is very common for usability studies (Sauro, J, 2010).

**TABLE 1
PATIENT DEMOGRAPHY**

<i>Patient</i>	<i>Gender</i>	<i>Ethnicity</i>	<i>Age</i>
Patient 1	M	Asian-American	50
Patient 2	M	Asian-American	55
Patient 3	M	Asian-American	76
Patient 4	M	Asian-American	50
Patient 5	M	Asian-American	46
			<i>Mean = 55.4</i>

RESULTS

The conversation between DIL and the patients were recorded for data analysis purposes and for review by the healthcare workers to design better conversational capabilities in the future. A broadcast message from DIL was sent to patients every day at 8:00 pm. The chat session duration varied slightly from person to person with the mean time spent at 18 minutes per session.

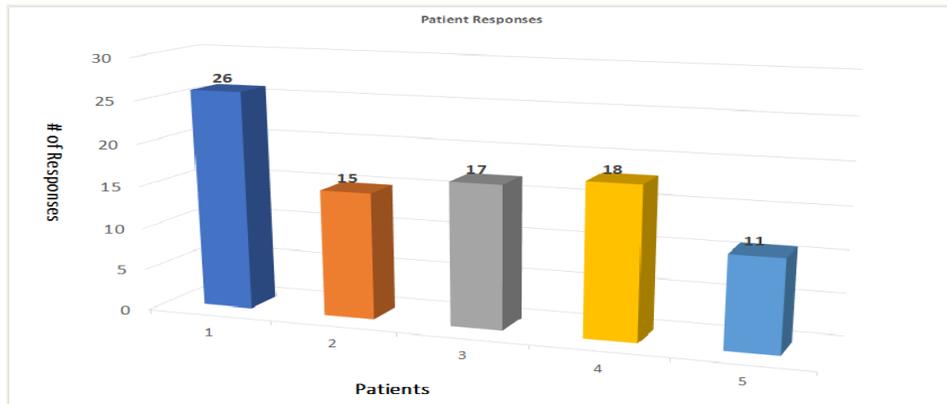
A post survey on User's satisfaction of DIL was also used to gauge Patients' experience with DIL.

Engagement

The patients typically chatted with DIL about 3 times a week. Each chat session reviewed the latest health dashboard data.

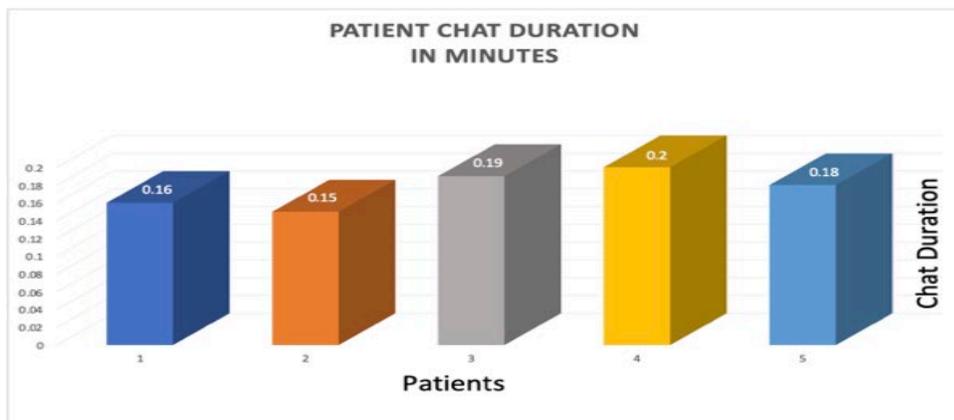
Figure 5 shows Patient Responses. The patients conversed on a variety of topics ranging from fluctuation in weight, Systolic and Diastolic Blood pressure, diet, exercise, and medicine. Over a four-week period during DIL's intervention, DIL sent messages to the patients every day for four weeks. Every patient's response rate was different, which was expected. Patient 1 responded 26 times, while patient 5 responded 11 times.

**FIGURE 5
PATIENT RESPONSES**



To accurately measure Patient Engagement, the Patient response rate is not enough. It is important to see how much time the patients spent with DIL. As mentioned before, DIL sent a message to the patients every day at 8:00 PM. The average time spent conversing with DIL varied slightly from patient to patient with 20 minutes at the high end for patient 4 to 15 minutes at the low end for patient 2. The average chat duration with DIL is 18 minutes. Figure 6 shows the Chat Duration by patients.

**FIGURE 6
PATIENT CHAT DURATION**



Satisfaction

The quantitative data as measured by the User Satisfaction and System Usability survey (Thomas, N, "How To Use The System Usability Scale...") indicated an average score of 91 (see table 2). This means that the patients were highly satisfied with the functionality of the system, and would not only use the system again, but also recommend to friends and family.

TABLE 2
PATIENT USER SATISFACTION SCORE

<i>Patient</i>	<i>User Satisfaction Score</i>
Patient 1	0.89
Patient 2	0.89
Patient 3	0.94
Patient 4	0.90
Patient 5	0.93
	<i>Mean = 0.91</i>

The researcher wanted to find out if the patients will be as receptive chatting with DIL as they will be talking with humans. The research compared results from before and after the intervention of DIL, and measured improvements in health and behavioral outcomes, if any, to determine the effectiveness of the DIL as a tool for persuasion, and elicit compliance to clinical and behavioral guidelines.

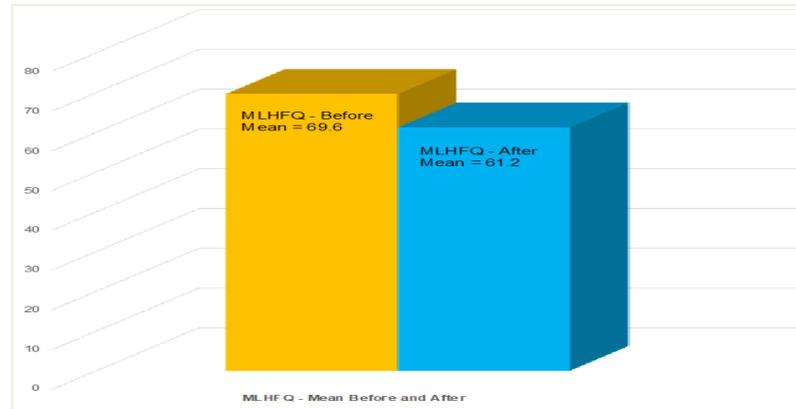
First, the Health dashboard measurements were compared from two weeks prior to DIL's intervention to four weeks during DIL's intervention. In particular, vital health parameters like fluctuations in weight, systolic and diastolic blood pressure, and pulse readings were compared. The mean scores for the five patients from before and after DIL's intervention were compared, and a paired t-test within the group was performed for a p-value of 0.05. A paired T-test is the ideal statistical measure to compare before and after scores within the same group as in this case. Table 3 summarizes the findings.

TABLE 3
COMPARISON OF HEALTH DASHBOARD RESULTS AND STATISTICAL SIGNIFICANCE

	<i>T-Value</i>	<i>T-Critical</i>	<i>P-Value</i>	<i>Statistical Significance</i>
Weight	1.63	2.776	0.05	Not significant
Systolic Blood Pressure	0.15	2.776	0.05	Not significant
Diastolic Blood Pressure	0.56	2.776	0.05	Not significant
Pulse	1.02	2.776	0.05	Not significant
<i>N=5</i>	<i>df=4</i>			

The second instrument used to answer RQ2, was the qualitative survey - Minnesota living with Heart Failure questionnaire (MLHFQ). The patients took the survey at the beginning of the trial, and again at the conclusion of the trial, and the results are extremely encouraging (see figure 8). There are twenty-one questions on this survey that can be divided up into two broad categories - physical and emotional domain. Questions 1-7, 12 and 13 ask questions about physical limitations, and fall under the Physical domain. Questions 17-21 ask questions mainly about the patients' emotional state, and fall under the emotional domain (Bilbao, et.al., 2016). The patients respond to each question on a Likert scale of 0-5, and the total score is calculated by adding up the response to each question. The mean score of each of the patients' total scores from before the trial was compared to the mean score from after the trial. As shown in Figure 7, the mean scores showed significant improvement (Mean before = 69.6, and Mean after = 61.2).

**FIGURE 7
MINNESOTA LIVING WITH HEART FAILURE QUESTIONNAIRE –
BEFORE AND AFTER SCORES**



The researcher also wanted to determine if DIL can fill the gap of scalability of human support, and its effectiveness. To that extent it was important to find out what the patients felt about interacting with DIL, and their experience. Did they find DIL helpful in achieving better behavioral outcomes was extremely important to know because if DIL can complement the efforts of the caregiver community, and motivate people to do better, then the fundamental problem of resource constraint is addressed to a large extent. The researcher used qualitative data – a post-trial survey, “Perceived Helpfulness of DIL” to gather patients’ feedback.

DISCUSSION

Overall conclusion to RQ1 is quite positive; it can be safely said that DIL scored high on User engagement and satisfaction, which is also confirmed by the General Feedback and Comments. Comments like *"The app is very well designed. It prompts me to monitor my health on regular basis"*, *"It reminds on timely basis & keep consistency"* and *"Great product to help track health goals"* only reinforces the positive user experience. This is probably also the reason for high engagement between DIL and the patients because when users' experiences with a technology artifact is positive, they are more likely to use the system.

**TABLE 4
PATIENT COMMENTS ABOUT DIL**

	<i>Patient Comments</i>
1	The app is very well designed. It prompts me to monitor my health on regular basis.
2	It reminds on timely basis & keep consistency. Because of daily reminders from DIL, I make it a point to walk for at least 20 minutes every day.
3	It would have been better if I could talk to the system.
4	Motivational messages in video format would have been more helpful.
5	Great product to help track health goals.

None of the health parameters that were measured showed any significant statistical change due to DIL's intervention. While at first, this might look discouraging for this research, but it is actually not a bad thing. The study duration was short - a total of six weeks, two weeks before DIL's intervention, and four weeks during DIL's intervention. A significant fluctuation is not desirable during such a short duration as any such change will be a cause of concern and would be considered a high-risk. All the five patients had

good family support, are well educated, and understand how to manage a chronic disease like heart failure better than the average population to begin with. In essence they were adhering to clinical guidelines like taking medications in a timely manner, controlling their diet and exercising on a regular basis. Given this scenario, the slight improvements in the mean values for all the four health parameters that were measured was actually an excellent health and behavioral outcome.

The findings indicate an extremely positive feedback from the patients on DIL's helpfulness on their quality of life and achieving a better health and behavioral outcome. Each of the five patients gave an "extremely positive" to "mostly positive" response to this five-question survey.

For RQ2 measured by the qualitative survey MLHFQ, the results were positive. A change of 4 in the Mean score is considered significant.

Overall conclusion to RQ3 is positive; given the extremely positive experience of the patients, there is definitely room for such an IT artifact in supporting patients as they make the transition from hospital to the home setting. Although it can be argued that the patient population in the trial was small and did not adequately represent the patient population with Heart Failure at large, and further exploration is definitely required, the early results indicate that DIL can play a role in alleviating the caregiver burden.

CONCLUSIONS

The research used Design Science Research methodology. The primary contribution is the artifact: DIL - a CA for individuals with Heart Failure; a chatbot that combines expert opinions with theories from healthcare, psychology, and technology. Although CAs are not new, the specific application of such a system to work with heart failure patients is new. This type of system was not found to exist at the beginning of this research project.

Contribution to Practice

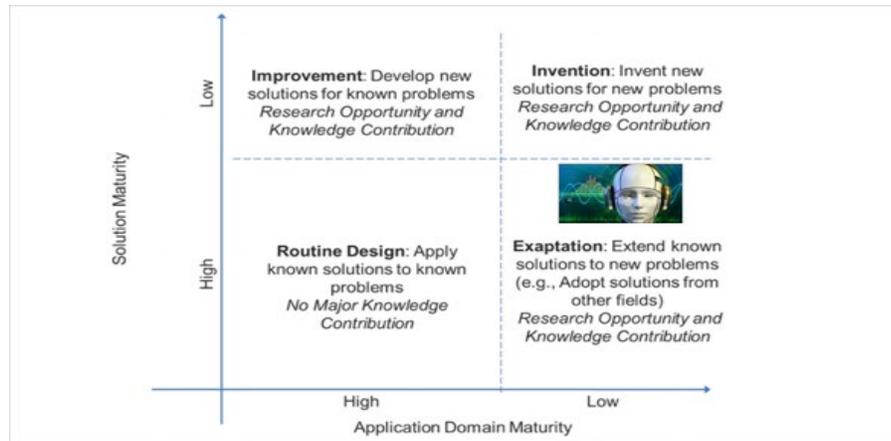
This research will hopefully empower and motivate patients and facilitate better communication among patients and clinicians. Communication is key to improving health outcomes and reducing readmissions. The artifact DIL is a direct attempt to address the Resource Problem, specifically overloading and shortage of staff issue. All conversations between DIL and an individual patient is recorded for further review by a knowledgeable healthcare worker like a Doctor, nurse or a clinician. These recordings can provide valuable insight about a patient's health, her adherence to clinical guidelines, and eventually progress towards better health and lifestyle.

Contribution to Theory and Knowledge

The research used two design principles that in all likelihood led to successful working of the DIL artifact. First, since this was meant to be used by elderly patients, simplicity and ease of use was a core design principle. Additionally, the use of rich media content (images and videos) were used as persuasive tools, which are proven to be far more effective than pure text content. Any future artifact constructed for elderly patients must adhere to these design guidelines.

This research can also inform the design of future CA systems and bring awareness of strategies to develop better customized interventions. Because these strategies will be discovered in a natural setting, they will be feasible and could form a foundation to develop guidelines for efficient self- management. Figure 8 shows the DSR-Knowledge Contribution Framework (Gregor, et.al., 2013).

**FIGURE 8
OVERLAYING THE IT ARTIFACT DIL ON THE KNOWLEDGE
CONTRIBUTION FRAMEWORK**



Gregor, et.al., 2013

Gregor and Hevner (Gregor, et.al., 2013) developed a 2x2 contribution framework. The framework positions contribution based on the maturity level of the solution and application domain. As shown in Figure 8, there are four types of contributions: Routine Design, Exaptation, Improvement, and Invention.

DIL is a CA specifically designed for Heart Failure patients with the primary objective of complementing the caregiver community who are over-burdened and under-staffed; assisting patients in self-care and management of their chronic condition, motivating them to adopt better behavior and a healthy lifestyle, and eventually achieving better health outcomes.

A CA based intervention is not new. Research has shown large success in achieving behavioral change with this kind of approach, and has previously been tried in a variety of settings from Alcohol consumption (Amini, et.al., 2013), promoting health behavior change in heart attacks (Silverman, et.al., 2001), psychosocial intervention deployed on hand-held computers (Johnson, et.al., 2004), hospital discharge nurse that explains written hospital discharge instructions to patients with low health literacy (Bickmore, et.al., 2009), psychotherapy (Lisetti, C. 2008), and as health dialog systems (Lisetti, et.al., 2008).

DIL borrows ideas and best practices from previously tried and tested applications in different settings, and aims to design and implement a CA in the specific domain of heart failure patients that has not been tried before. In this regard, the IT artifact DIL falls in the Exaptation quadrant of the Knowledge Contribution Framework (Gregor, et.al., 2013).

Limitations and Future Research

Although the study was well received by the patients participating in the trial, and the findings were encouraging, there is surely room for improvement.

One limitation of the study was the short duration of only six weeks. A future random control trial with a minimum of ten patients over three to six months will be ideal to really gauge the impact of DIL on behavior and health outcomes - an improvement on all health vitals like weight, blood pressure and other symptoms will really validate the efficacy of a CA like DIL. If possible, the trial should include two groups of patients - the first group will receive support from nurses and clinicians, and the second group will be supported by DIL. At the end of the trial health and behavioral outcomes can be compared to measure the effectiveness of DIL.

Further an integration with IBM Watson - a rule based expert system that can accurately predict the recurrence of heart failure by monitoring a patient's individual responses during conversation with DIL and his or her health dashboard results will go a long way in addressing the resource problem mentioned in earlier sections.

Finally, the current artifact was limited to home monitoring dashboard results. In an ideal scenario, DIL will not only review these dashboard results, but further integrate with the patients' electronic health records residing in the primary care physician's office, and take advantage of the rich repository of patient health information. DIL can then mine that data, and provide a more comprehensive picture of the patients' health, and develop the capability to truly converse with patients using Natural Language Processing (NLP) in an intelligent way. DIL can also proactively notify the doctors, nurses and members of the caregiver community when it identifies patients with high risk, so that the doctors and the nurses can follow up with patients before something goes wrong.

REFERENCES

- Alnosayan, N. (2015). *Development of a Home Telemonitoring System to Support Self-Care in the Context of Congestive Heart Failure: A Rule Based-Expert System Approach*. Dissertation Thesis.
- Amini, R., Lisetti, C., Yasavur, U., & Rishe, N. (2013). On-demand virtual health counselor for delivering behavior-change health interventions. *Proceedings - 2013 IEEE International Conference on Healthcare Informatics, ICHI*.
- Baylor, A. (2009). *Promoting motivation with virtual agents and avatars: Role of visual presence and appearance*. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*.
- Benjamin, E.J., Muntner, P., Alonso, A., Bittencourt, M.S., Callaway, C.W., Carson, A.P., . . . on behalf of the American Heart Association Council on Epidemiology and Prevention Statistics Committee and Strokes Subcommittee. (2018). *Heart failure fact sheet*. Center for Disease Control and Prevention. Retrieved from https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_heart_disease.htm
- Benjamin, E.J., Virani, S.S., Callaway, C.W., Chang, A.R., Cheng, S., Chiuve, S.E., . . . on behalf of the American Heart Association Council on Epidemiology and Prevention Statistics Committee and Strokes Subcommittee. (2018). *Heart disease and stroke Statistics— 2018 update: A report from the American heart association statistics committee and stroke statistics subcommittee*. Retrieved from https://www.heart.org/-/media/data-import/downloadables/heart-disease-and-stroke-statistics-2018---at-a-glance-ucm_498848.pdf
- Bickmore, T.W., Pfeifer, L.M., & Jack, B.W. (2009). Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. *Proceedings of the 27th international conference on Human factors in computing systems - CHI*.
- Bilbao, A., Escobar, A., García-Perez, L., Navarro, G., & Quirós, R. (2016). The Minnesota Living with Heart Failure Questionnaire: Comparison of Different Factor Structures. *Health and Quality of Life Outcomes, 14*, 23. <https://doi.org/10.1186/s12955-016-0425-7>
- Chatbots are Changing. Can They Change Behavior, Too?* (n.d.). Retrieved from <http://www.carrotnewyork.com/blog/2018/6/19/0o8opfkx8jx5jysjzkzybgm9lbcqud>
- Conversational Agent. (n.d.). Retrieved from https://www.chatbots.org/conversational_agent/
- Creswell, J.W., Hanson, W.E., Plano Clark, V.L., Petska, K.S., & Creswell, J.D. (2005). *Mixed Methods Research Designs in Counseling Psychology*. p.373. Faculty Publications, Department of Psychology.
- Dialog Systems. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Dialogue_system
- Eyal, N. (2017). *Bots and A.I. In Behavioral Concept with Nir Eyal*. Retrieved from <https://chatbotsmagazine.com/bots-and-a-i-in-behavioral-concept-with-nir-eyal-b13922c8dd6>
- Ferguson, G., Allen, J., Galescu, L., Quinn, J., & Swift, M. (2009). *CARDIAC: An Intelligent Conversational Assistant for Chronic Heart Failure Patient Health Monitoring*.
- Folstad, A., Brandtzaeg, P., Feltwell, T., Law, E., Tscheligi, M., & Luger, E. (2018, April). Chatbots for social good. *CHI Conference on Human Factors in Computing Systems CHI EA 2018 2018 04 21*

- 2018 04 26. *Conference on Human Factors in Computing Systems - Proceedings*.
doi:10.1145/3170427.3185372
- Gregor, S., & Hevner A.R. (2013, June). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2).
- Hevner, A.R., & Chatterjee, S. (2010). *Design Research in Information Systems, Theory and Practice*. Springer Publications.
- Johnson, W., LaBore, C., & Chiu, Y. (2004). A pedagogical agent for psychosocial intervention on a handheld computer. In *AAAI Fall Symposium on Dialogue Systems for Health Communication* (pp. 22–24).
- Kowatsch, T., Nißen, M., Shih, C-H.I., Rügger, D., Volland, D., Filler, A., . . . l'Allemand, D. (2017). *Text-based Healthcare Chatbots Supporting Patient and Health Professional Teams: Preliminary Results of a Randomized Controlled Trial on Childhood Obesity*. <https://doi.org/10.3929/ethz-b-000218776>
- Lisetti, C. (2008). Embodied Conversational Agents for Psychotherapy. In *Proceedings of the CHI 2008 Conference Workshop on Technology in Mental Health*.
- Lisetti, C., & Wagner, E. (2008). Mental health pro- motion with animated characters: Exploring issues and potential. In *Proceedings of the AAAI Spring Symposium on Emotion, Personality and Social Behavior*.
- Lisetti, C., Yasavur, U., & Visser, U. (2012). Building an On-Demand Avatar-Based Health Intervention for Behavior Change. *Proceedings of the Twenty-Fifth International Florida Intelligence Research Society Conference*.
- Lodico, M.G., Spaulding, D.T., & Voegtler, K.H. (2006). *Methods in educational research: From theory to practice*. Jossey-Bass: San Francisco.
- Miller, W.R., & Rollnick, S. (2002). *Motivational interviewing (2nd Ed.)*. *Preparing people to change addictive behavior*. New York: Guilford Press.
- O'Connor, C.M. (2017). *Heart Failure Readmission Rates*. Retrieved from <http://heartfailure.onlinejacc.org/content/5/5/393>
- Palanica, A., Flaschner, P., Thommandram, A., Li, M., & Fossat, Y. (2019). Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. *J. Med Internet Res.*, 21(4), e12887.
- Robson, C. (2002). *Real World Research - A Resource for Social Scientists and Practitioner-Researchers*. Malden: Blackwell Publishing.
- Sauro, J. (2010). "Why You need to test with five users...", *Best practices for using statistics on small sample size*. Retrieved from <https://measuringu.com/five-users/>
- Silverman, B., Holmes, J., Kimmel, S., Branas, C., Ivins, D., Weaver, R., & Chen, Y. (2001). Modeling emotion and behavior in animated personas to facilitate human behavior change: The case of the HEART-SENSE game. *Health Care Management Science*, 4(3), 213–228.
- Thomas, N. (n.d.). *How To Use The System Usability Scale (SUS) To Evaluate The Usability Of Your Website*. Retrieved from <https://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-website/>