Instruments for Exploring the Training and Aesthetic Dimensions of Edutainment: Case of the Heart Sense Game

Technical Report

by

Barry G. Silverman[1], William Holmes[4], Melanie Green[3], John Holmes[4], Stephen Kimmel[4], Joshua Mosley[2], John Pourdehnad[1], Iraj Zandi[1], Gnana Bharathy[1], Michael Johns[1], Beth Ohr[4], Ransom Weaver[1], Jesse Berlin[4], Lynne Taylor[4]

1 - Ackoff Center for Advancement of Systems Approaches (ACASA), Engineering School
2 - Digital Media Design Program, Graduate School of Fine Arts
3 - Psychology Dept., School of Arts and Sciences
4 - Center for Clinical Epidemiology and Biostatistics, School of Medicine
University of Pennsylvania, Philadelphia, PA 19104-6315
barryg@seas.upenn.edu

Sponsored by

National Library of Medicine (Contract 467-MZ-902060)
An The National Heart Attack Alert Program of the
National Heart, Lung, and Blood Institute
National Institutes of Health

October 2003

ABSTRACT

This report presents instruments for measuring the training and aesthetic dimensions of edutainment with particular focus upon interactive drama games that serve as health behavior change interventions. After overviewing the specific Heart Sense game that we created (Sect. 2), we explain the instruments (Sect. 3) and our plan for both validating them (Sect. 3.4) and using them in a clinical trial (Sect. 4). Overall, in addition to a demographics instrument we have developed instruments for metrics in both training dimensions (Knowledge, Stated Intent, Willingness to Pay) as well as in aesthetic dimensions (Narrative Engagement, Game Entertainment, Usefulness, and Usability). We also make use of two previously developed instruments on decisional conflict and need for cognition. The draft instruments are provided in the Appendices.

TABLE OF CONTENTS

1) Introduction

This effort is aimed at constructing and evaluating a game-based intervention for health behavior change. The many design issues we have been struggling with (e.g., constructivist vs. instructionist pedagogy, stealth learning, game genre, and entertainment, interactive drama, and narrative engagement, etc.) raise the prospect that it is vital to evaluate such interventions. The state of the art today for evaluating edutainment is not robust, and we found we had to develop our own instruments to support exploring the training and aesthetic dimensions of various game-based interventions. The purpose of this report is to explain the instruments we developed and are now using in a clinical trial.

We believe these instruments will be helpful to others attempting to evaluate the efficacy of edutainment interventions, in general, and of story-based interventions, in particular. To further their utility, we hope to explore the validity of these instruments in our upcoming trial. Until that is completed, the reader is cautioned that the current version of the instruments, as contained in this report, is prior to our validity assessment efforts.

In this report, Section 2 provides some detail on the nature of the HEART SENSE game-based intervention we have developed, Section 3 overviews the set of instruments we created and plans for exploring their psychometric properties, and Section 4 presents the design of the trial we are now undertaking for the next 3 months in order to explore the validity of these instruments and to assess the efficacy of the drama game genre for this type of intervention. The Appendices provide the text of each of the seven instruments we have created, and the adaptation of two others. We omit showing the web screen views of the instruments.

Rationale for a Game -- Delay in seeking care for acute myocardial infarction (AMI) is known to be a potent predictor of poor outcome. This is especially the case where pre-hospital delay exceeds the 12-hour window allowed for the initiation of thrombolytic therapy and/or angioplasty. Yet, it is known that people often delay much longer than this. In one study [1], AMI patients in Los Angeles were found to delay as much as 92 hours after the onset of symptoms. Delay in seeking care has implications beyond thrombolytic therapy, however. For example, patients delaying in presenting to the emergency department may telegraph to their physicians that their clinical situation is not important, and may not be admitted to the hospital for an AMI as a result [2]. Some of the reasons for delay include a bias toward an optimistic outcome [3]; lack of cardiovascular knowledge [4]; differences in ethnic background and race [5]; gender and age [6]; self-perception of symptoms [7], and prior medical history [8].

Attempts to reduce delay have focused on pedagogical materials, such as pamphlets that aim to instruct the user by forcing memorization of concepts. Given the continuing high rate of delay in seeking care, it is clear that these methods are only partially effective. Other approaches, including interventions such as the REACT trial [9] have enjoyed more success, if only in limited areas. Finally, while some studies focus on symptom recognition as a strategy for reducing pre-hospital delay, it is not clear that they address the larger behavioral etiology. An alternative approach is needed that addresses education, but also aims to change beliefs and attitudes about AMI, risk, and delay.

One alternative approach relies on a constructivist pedagogy that minimizes rote learning by incorporating the trainee into a realistic, non-threatening scenario in which she discovers for herself the key elements of the material to be learned. To that end we have finished developing an interactive drama game, called HEART SENSE [10,11,12] (about 50 minutes of continuous play for once through playing), in which the player is engaged in the drama. In working through the drama, the player encounters the need to apply and rehearse knowledge about heart attack symptoms and delay issues. This is stealth learning since the requisite knowledge is obtained from the reactions of the heart attack victims to the player's actions, from the cues of other actors in the drama, and from seeing diverse outcomes while pursuing one's goals in the virtual world. The idea of an interactive drama as the game genre stems from the concept that players in the target age group are less likely to be sophisticated gamers, and are likely to respond to games that have a high story content much as television shows and movies do.

2) The HEART SENSE Game

HEART SENSE is a computerized role-playing game that was developed by a large, highly multidisciplinary team including game designers (BGS, MJ), an animator (JM), multimedia developers (RW), scriptwriters, graphics artists, voice-over actors, song writers, a cardiologist (SK), educators, and clinicians. Most of the authors of the current report served on the evaluation team, however, they participated throughout the intervention design process to contribute feedback on content, story, choreography, and so on as each version was prototyped.

The goal of HEART SENSE is to help the player to overcome his or her own symptom recognition and resistive behavior issues before he or she has a first heart attack. The underlying assumption of this intervention is that learning about these issues in the context of a story world with believable plot, story line, characterizations, animation, and set design will help to reduce delay in seeking care if one ever encounters a heart attack in the real world.

Figure 1 shows the cast of characters in the game; the chap in the black shirt (center), is Jack Janus, the game's hero, and is used as the avatar for the player. Thus, the player uses Janus as his or her surrogate in the game. There are three heart attack scenarios in HEART SENSE, with the victims played by Joe (the African-American chef); Danny, the Latino man in the baseball cap; and Catherine, an elderly Caucasian woman, at the far left of the figure. The other characters play supporting roles in each of the three heart attack scenarios.

Figure 1. The HEART SENSE Cast.

(Figures drawn and animated under the supervision of Joshua Mosley, MFA, University of Pennsylvania).

In the game, the user helps Jack try to resolve a town dispute, and simultaneously rescue his career and win back the woman who is the object of his affections (Wanda, the blonde woman to the far right in Figure 1). Jack is a newly graduated physician who is down on his luck after giving up medicine out of guilt over failing to save his mother when she had a heart attack. He shows up at the office of his former mentor (Dr. Addams, the character in the white coat in Figure 1) for a chat, and discovers that he has been kidnapped. The hero's brother, Charlie Janus (not shown), is a businessman who wants to change the town by bringing in a casino. Most of the townsfolk resist this plan, in part because it will be built on the town's central park, gardens, and birdlife sanctuary. Dr. Addams is also the town's mayor and a leading voice against the casino. There is to be a town meeting at 7 PM that evening to vote on the casino project, and with Addams missing, the vote might very well go in favor of the casino.

On his journey to find the kidnapped doctor and mayor, the hero is tempted to throw in with his brother. It is up to the player to decide which side of the town development cause to side with. The player is taken around the town to discover how fellow townsfolk feel about the issues, and to find clues to help either the brother or the solving of the kidnapping.

In moving through the town, the player encounters three heart attack scenarios in various settings, each of which is very different in character. The first is a "typical" event, occurring in an overweight African American man (Joe, the Pizza shop owner). His signs and symptoms include crushing retrosternal chest pain, diaphoresis, and eventually, collapse. Joe has never missed a day of work and is reluctant to abandon his shop just before the lunch rush. The second is an "atypical" event, in that the victim is an elderly woman (Catherine), presenting with signs and symptoms that deviate from the expectations held by the lay public. Catherine has raised two children who now live in the suburbs and she does not want to be a burden to them for any reason. The third is a "hybrid" event, in that it occurs in the context of a softball game to an otherwise healthy-appearing younger man (Danny), where muscle strain and pain are common, and could be confused with a heart attack. Danny is the team leader and does not want to appear weak in front of his mates.

In each scenario, the hero's job is to help the victim to get past the various barriers he or she erects to avoid seeking urgent care. However, this task is non-trivial, even within the context of a computer game. For a variety of social, attitudinal, or behavioral reasons victims often don't believe they are having a heart attack or don't want to seek urgent care. Thus, there are significant obstacles that the hero must face in helping these characters help themselves. This is especially true among those having a first heart attack. Statements such as "I can't go to the hospital, because my grandchildren are coming over here after school" or "This kind of thing doesn't happen to people like me - I'm big and strong, and Latino people don't have heart attacks like other people do" would be expected in real life, and they are incorporated into the dialogue of HEART SENSE as well.

The user, playing as the hero Jack, can decide to help each victim to accept the need for care, move into a state of self-efficacy, and call 911. Alternately, Jack can decide that others in the scenarios should tend to the victims since there is little time and he must convince his brother to make him a partner in the casino deal. At the end of the game, Jack confronts his brother and has the opportunity to rescue his mentor, or join his brother in his town redevelopment initiative. While the clinical objective of the game is to help the characters pass through their different barriers to urgent care, it is possible that the user, by manipulating the hero's behavior, can harm the heart attack victims. That is instructive and reveals lessons on acute care, modeled by other players in the scenes. These players, such as the female Wanda, focus their ire upon Jack if they are thrust into this role. Thus Jack can gain a lucrative career in casinos at the expense of the affection of the townsfolk.

3) Overview and Derivation of Instruments

In general, there are three principal sets of instruments of interest to the evaluation of game-based interventions: demographics, training, and aesthetics. We first list these categories and the dimensions to be measured by instruments under each category. We then explain some of the background surrounding the design of each instrument. As a general strategy, our team's metrics expert (WH) has guided us to design each instrument as a dimension to be measured. Within each instrument the dimension is further decomposed into sub-dimensions that appear to contribute to the total dimension score for that instrument. We then created textual questions that we thought would capture answers of relevance to each sub-dimension. This discussion covers the dimensions and sub-dimensions. The full instruments showing the actual textual questions for each sub-dimension are included in Section 5. The categories of dimensions and the primary dimensions of our instruments are:

- Demographics (D)
  - Physical and Socioeconomic Attributes
  - Cognitive Attributes (need for cognition)
- Training (KID - WTP)
  - Knowledge (K)
  - Stated Intent-1 (I)
  - Decision Conflict (D)
  - Intent-2: Willingness to Pay (WTP)
- Aesthetics (EEUU)
  - Engagement (story) (ENG)
  - Entertainment (gameplay) (ENT)
  - Usefulness (UF)
  - Usability (UB)

Demographics include the obvious physical and socioeconomic traits that help to differentiate subjects such as age, gender, education, and the like. We use inclusion and exclusion criteria (see Section 4) prior to assignment of subjects to the trial, and these criteria are omitted from our demographic instrument, though they are implicit in its design. Demographics also can include psychological differences. In the case of an intervention, one could measure predisposition to learning styles or to message formats. Our team's psychologist (MG) has guided us to include one such dimension -- the "need for cognition" scale which measures High Need for Cognition (HNC) and Low Need for Cognition (LNC). There are literally hundreds of studies showing that need for cognition moderates persuasion effects. The most relevant research suggests that individuals who differ in need for cognition also differ in preferences for message formats. High need cogs aren't scared off by something like a pamphlet, whereas low need cogs prefer things to be simple and fun. So, one hypothesis is that low need for cognition individuals will more persuaded by the game, and high need for cognition individuals may be equally or more influenced by the pamphlet. There already is an instrument for measuring need for cognition and predisposition to message styles. MG has incorporated into our demographics instrument, the shortened form of a validated instrument on need for cognition (Cacioppo et al., 1984, 1996) [13, 14]. The question wording we're using is the one that has been validated with a national sample (see ftp://ftp.nes.isr.umich.edu/ftp/resourcs/psreport/98pilot/bizer98.pdf for details). We did change the response scale, though, to be more consistent with the format of the other questions we're asking and to be more user-friendly in the web version utilized by our instruments.

3.1) Edutainment Discipline Silos

Several disciplines must collaborate and overcome their boundaries to have a successful edutainment intervention. Specifically, [11] points out that three divergent guidelines affect different sub-disciplines in the edutainment field: training/content, active gameplay aesthetics, and passive-listening but engaging story and characters. One extracts ideas from the "discipline silos" of trainers, interactive gamers, and writers and synthesizes the ideas into a coherent edutainment piece. In doing so, many design tradeoffs are made, and the result seeks to balance contributions of each of the disciplines (silos). In the end, one wishes to measure how well the piece of edutainment influences subjects in the dimensions important to each discipline. To straddle the three silos of edutainment and assemble the instruments of this report required contributions from most of the authors. For example, WH was the prime on the K and I instruments, with help from MG, JH, and SK. BGS was the prime initiator of the WTP instrument with help from JP and IZ, and with much of the implementation done by MG, among others. The follow-up WTP instrument is the creation of JH and BO. The aesthetics or EEUU instruments are the creation of BGS with structuring and wording guidance from WH, as well as edits from MG, JH, and others. The exceptions to this are that: (1) the transport (T) sub-dimension of ENG is the creation of MG alone, and (2) MJ collaborated as creator on several of the sub-dimensions of ENT. All of the instruments were turned into webforms by GB, with significant visual feedback from JH and BO and programming help from RW among others. Paper-based versions of the instruments suitable for trial use were assembled by BO.

3.2) Training Dimensions

To assess the effects of the game, we will measure four training-oriented dependent variables:

Knowledge (K) measures will assess individuals' recognition of heart attack symptoms and their awareness of risk factors. If individuals cannot recognize symptoms, they may not know that they need to call 911.

Behavioral intent measures will assess individuals' intention to seek treatment in a variety of circumstances. Because we do not have the resources (sample size and time) needed to assess actual behavior (calling 911), behavioral intent measures are the best proxy measure available. We shall measure both ***stated intent (I)*** and underlying behavioral tendency as captured by a ***Willingness to Pay (WTP)*** metric.

Decision conflict **(D)** assesses individuals' certainty about their decision to call 911. This measure can be interpreted as a measure of self-efficacy. To measure this dimension we adapt an already validated decision conflict instrument as found in O'Connor et al. (2002) [15]. Our sole adaptations are to replace keywords that make it relevant to the heart attack context, and as such we omit discussion of this instrument here.

More explicitly, we are interested in these dimensions within the context of heart attack delay reduction. The current HEART-SENSE Game is a multi-media interactive, computer-based health promotion program designed to improve health knowledge and influence behavior among individuals thought to be at risk of a first acute myocardial infarction (heart attack).

One of the first parts of the model is Knowledge or symptom and risk recognition. Knowledge about heart attack risks and symptoms plays a role in an individual's final decision not to delay when faced with heart attack symptoms. In the case of heart attacks, a person has to recognize that certain physical signs can indicate a heart attack. The person also needs to recognize that he or she is at risk. If a person does not think he is having a heart attack, his intention to call 911 for heart attack treatment will not be acted upon. The game attempts to provide the necessary knowledge by providing vivid depictions of individuals experiencing shoulder pain and heartburn-type symptoms as well as the more familiar chest pain. Similarly, women, minorities, and individuals who exercise are shown experiencing heart attacks, thus illustrating the widespread nature of risk.

Completion of the "Knowledge" questionnaire will provide a score (K) that indicates each participant's awareness of the importance of calling 911 in the event s/he is having a heart attack. The total score is obtained from the combined scores of two tasks and four subscales (respectively): accurate generation of heart attack symptoms (AG); accurate recognition of heart attack symptoms (AR); knowledge of symptom pattern possibilities (S); knowledge about risk factors for heart attacks (RF); knowledge about issues surrounding treatment (T); and knowledge about one's plan for action in the event one were to have a heart attack (PA). (Not all elements of each hypothesized knowledge subscale nor all hypothesized knowledge subscales could be tested since inclusion of items in the questionnaire was dependent on what was covered in all planned intervention arms.) Overall, we hypothesize that $K = AG + AR + S + RF + T + PA$.

The second dimension or scale of interest is that of Intent (I), or precisely, of shifting intended behavior once the subject recognizes the threat to their health. In general, any given behavior includes an action (e.g., calling

9-1-1), a target (e.g., a telephone), and a context (e.g., experiencing chest pain). Behavior theorists contend that the most effective interventions will be those directed at a single behavior rather than at multiple behaviors or behavioral categories (Hardin & Reis, 1997) [16]. In this way, the HEART-SENSE game has targeted "delay in calling 9-1-1" as opposed to a more general goal such as "improved heart health". In this case, usage of a telephone is a common skill that most individuals are able to perform without difficulty. In particular, the 9-1-1 exchange has been promoted by the HEART-SENSE game because of its obvious simplicity relative to seven or 10-digit dialing schemes. Furthermore, the vast majority of the U.S. has been equipped with 9-1-1 technology and individuals have been made aware of its proper uses. Environmental hindrances will also be relatively minor when compared to barriers to intention. Lack of a telephone is probably the most prominent environmental hindrance. Nevertheless, more than 97% of all households in Pennsylvania and more than 99% of households in Philadelphia have telephones making this barrier somewhat negligible (Pennsylvania Data Center, 2000 [17]; Norman, 2000) [18].

In order to effectively change 9-1-1 usage behaviors related to acute myocardial infarction, we have borrowed concepts from several different behavioral models to tailor the HEART-SENSE game to the needs of individuals thought to be at risk of acute myocardial infarction (Ajzen & Fishbein, 1980 [19]; Bandura, 1986 [20], 1994 [21]; Fishbein, 2000 [22]; Fishbein et al., 1992 [23]; Becker, 1974 [24]; Rosenstock et al., 1994 [25]). This piecemeal strategy has been used in the past (Fishbein, 2000 [22]; Rhodes, Fishbein, & Reis, 1997 [26]). However, the primary theoretical model is Ajzen & Fishbein's theory of planned behavior (an extension of the theory of reasoned action). This model, which has been widely used in health contexts, states that the immediate determinant of behavior is an individual's intent to perform the behavior. Studies have shown moderate to high correlations between intention and actual behavior (e.g, McArdle, 1972 [27]; Fishbein & Coombs, 1974 [28]; see Shepard, Hartwick, & Warshaw, 1988 [29], for a meta-analysis of the intention-behavior link). The determinants of behavioral intent are discussed below.

Our game helps users understand these barriers on a pragmatic plane. The HEART-SENSE game includes information about the benefits of calling 911 (efficacy of prompt treatment; ability to survive a heart attack versus not) which should lead to a positive attitude toward seeking treatment quickly. The game also includes statements (made at various times by the village characters) to inform the user of what others think they should do, and what others have done, in a heart attack situation with a positive outcome. In this way, the user is motivated by social pressure that others think they should perform the behavior in question. The game also attempts to address social barriers to seeking treatment, such as embarrassment about looking weak by seeking medical treatment (baseball scene) or not wanting to trouble or inconvenience others (garden scene). The game addresses further barriers such as conflicting responsibilities (e.g., of a job in the pizza scene).

Completion of the "Intent" questionnaire will provide a score (I) that indicates each participant's stated intent to call 911 in the event s/he is having a heart attack. The total score is obtained from the scores of three subscales: recognition of symptoms that indicate a heart attack is likely enough for call to 911 (R); likelihood for being affected by barriers to calling 911 in the presence of heart attack symptoms (B); and perceived "immunity" from having a heart attack (IM). Overall, we hypothesize that $I = R + B + IM$.

There is a large literature on stated intent vs. actual behavior, and on the importance of attempting to measure the underlying intent [Argyris & Schön (1974) [30], Irwin & Baron (2001) [31], Kelly et al. (2003) [32]. Sometimes this literature is referred to as impression management. Argyris & Schön, for instance, refer to the differences as those between the "theory espoused" and the "theory in use." Irwin & Baron, in turn, refer to the different ways one might seek to measure the underlying intent, the theory in use that tends to drive behavior despite what subjects state they might do. One approach to capturing underlying intent is to follow the subjects from each arm of the trial around for several years and to record their delay behaviors when and if they actually have a first MI. This requires a stage two clinical trial, and we do not have funding for such an effort.

In lieu of a stage two trial, researchers often turn instead to surrogate metrics such as subjects' willingness to pay for items that are important if indeed they have had a shift in intent. Our approach to this is to ask our subjects what portion of their subject bonus fee they would be willing for us to utilize to send a 9-1-1 pager to a loved one. The WTP item is intended to capture a somewhat more behavioral aspect of intent. That is, are individuals willing to take some action to decrease future delay? Raising the cost of taking this action may produce differences between the groups, such that only the people most concerned about delay (and thus perhaps most willing to seek future treatment) might be willing to give up money in order to obtain a delay-reducing device.

Previous research suggests that WTP questions tend to tap into different concerns than attitude questions. Specifically, when people answer WTP questions, they're thinking about their budgets, individual situations, and the like, perhaps more so than their values. The WTP question thus provides different information about how valuable people perceive the intervention to be, and probably should be used as a supplement to, rather than as a replacement for, direct measures of intent.

3.3) Aesthetic Dimensions (EEUU)

In this section we introduce four more dimensions that are relevant to edutainment and that pertain to their aesthetics. The acronym for these dimensions is EEUU which refers to Entertainment (ENT), Engagement (ENG), Usefulness (UF) and Usability (UB). Each of these four dimensions is further defined below along with a discussion of the relevant sub-dimensions and their genesis. The order of these instruments when administered to subjects (players) follows that just listed, however, in the ensuing paragraphs we present them in a slightly different order for the sake of the discussion.

Although training requires players to progress through stories (pedagogically valuable scenarios), at its heart game play is not about interactive fiction though there are those who buy interactive fiction games. Interactive drama is all about storytelling from the author, while game-play is much more about story creation by the player – and these competing aesthetics need to be resolved if pedagogical games are to achieve their potential in general. In our instruments we set out to capture these two diverse aesthetics – narrative and gameplay – within the idea of two main dimensions for evaluating interactive, story-based games. Those two primary dimensions include engagement and entertainment, respectively. Before explaining these, some further background is warranted.

Many game designers try to confront their players with a tradeoff decision, a dilemma. Thus you can't win a Pac-man level unless you risk your lives and collect all the food/energizers that the ghosts more or less protect. Likewise in Space Invaders you score on offense, though you live longer by taking evasive tactics and going on defense. Making the tradeoff decisions (reluctantly answering the call), venturing forth to collect rewards (questing), and reaching the ultimate objective (conquering the enemy) – this sounds an awful lot like the "hero's journey" [33]. In fact, it is a hero's journey, one where the player is the hero. Without ever writing a single story, game designers are brilliant at inserting mechanics (dilemmas, actions, rewards) that allow players to create innumerable new stories. Because these stories are so different, yet exist about a common experience, the players readily share them with one another, compare them, brag about them, etc. And the players are encouraged to replay the game to live through new stories that they can further brag about.

Another way to think of these ideas is that there are several central aesthetics and mechanics that should emerge from game design. For one thing the game designer must try to preserve the overall entertainment score (ENT) which is a function of the (F) or "fun quotient" that derives heavily from four contributing sub-processes, at a minimum, including: (R) "Rules" of the game that are satisfying such as how to move around, interact, fight, persuade, etc.; (C) "Control" remains in the hands of the player in the sense of creating his own story and selecting tactics to deal with dilemmas along the way; (O) fairness of outcomes along the way and in the end (did player's choices have believable consequences) - social contract between designer and player; and (AT) accumulate-threaten aesthetic or reward-punishment - the game provides the player with an opportunity to collect things, threatens him with their loss (scare/thrill), allows him to try and protect them, etc. These four mechanics are central to almost any game, and they must be appealing if the game is to be fun for the player: overall, we might say that ENT = F + R + C + O + AT. However, they also must be tuned to the class of player (demographics) targeted for the specific game.

More than any other mechanic of game-play, narrative in game raises the idea of destroying the central aesthetic – that players create their own stories and that is what keeps them coming back. Other game-play mechanics more or less have a story inside them, in fact countless stories inside them. Further, many of these mechanics have built in skill training functions at the same time that they permit unconstrained play and inquiry. For example, a racing and chasing mechanic includes lessons about how to chase down bad guys and cut them off from escape. Likewise a combat game has built in weapon firing and target damage models, plus skill challenges such as room clearing, among others. If one invests in the realism of these mechanics, they provide useful training and transferable skills (Green, 2003) [34].

It's also a fact that students learn the most and retain it the longest when they must teach a topic to others. One always learns a subject better if one is confronted with being the teacher rather than the student. So a microworld could be quite a powerful training device if it affords teaching opportunities, or even better if it thrusts the player into roles where other characters will be vulnerable and dependent on the player to teach for a successful conclusion to be reached.

Why should the player care to become a teacher? What can drive them to reach this level of learning? People reflect this kind of passion for videogames and at the movies. When game mechanics work and when characters are likable, players (viewers) achieve enormous empathy for the characters and are willing to go to great lengths to save them and to help them work out their problems (e.g., as in 'God' games such as The SIMS, virtual Petz, and Tamagotchi), and to go on quests on their behalf or assist them in shifting their behaviors to more successful models such as in role playing games. In these milieus, players become invested in characters and reveal willingness to learn skills that will help these dependent characters to cope. Various authors point out how people are generally attracted to characters the more they see some of themselves in the character or in the experiences the character is going

through: e.g., see (Meretsky, 2001 [35]; Decker, 2002 [36]). Other authors point out how important the story structure (plot) is. In general, both are important as is place, and one must consider these 3 Ps in evaluating a storyworld: people, plot, and place.

In sum, we are concerned with the degree to which the user is caught up in or "engaged" in the edutainment. One can think of engagement (ENG) as the strength of the bond or how compelling is the experience of the artifact (can the user stop easily): the level to which one is personally involved in and affected by a story. In the literature on engagement, some researchers distinguish between things like immersion and presence or transport [e.g., see Whitton, 2003 [37] -- the former referring to what the technology delivers (3-D effects, sound quality, etc.), while the latter refers to the absorption of or impact upon mental state of the user (Green & Brock, 2004) [38]. We think this distinction is relevant, but there is another dimension that is equally important in story-worlds since it is not just the immersive technology that impacts presence, but its also the 3Ps of the story. Books have no immersive technology yet readers often find themselves transported, and compellingly so.

The result of all this is that there are several sub-dimensions that seem worth trying to measure with an engagement instrument. Overall, we might state that ENG = T + P1 + P2 + P3, where each of these include:

TRANSPORT, T, defines the degree of 'presence' at the top level. According to Green & Brock (2000) [39], transport is the phenomenological experience of being absorbed in a story.

PLOT (P1) - Degree to which one is interested in themes raised and conflicts resolved.

PEOPLE (P2) - a) Degree to which characters are "like me" and/or "like people close to me."
b) Degree to which the characters are interesting to me.
c) Degree to which character development occurs.

PLACE (P3) - a) Degree to which one can transfer the story experience to one's real life.
b) Degree to which actual sensate experience (i.e., sound, visuals, smells, terrain, sets, etc.) of story is "like real." This is what others measure as 'immersion'.

In addition to entertainment and engagement, there are a number of other aesthetics dimensions and sub-dimensions that affect players' sensibilities with respect to a piece of edutainment. From a pragmatic perspective, we are most interested in those that foster the piece of edutainment being usable (UB) by and useful (UF) to the player. With respect to a health intervention such as Heart Sense Game, we have targeted a middle aged and elderly population. This heightened our concern that the design avoid many of the traditional game mechanics (e.g., combat, racing, object manipulation, etc.) and their attendant playability issues. We wanted no challenge to playability or usability other than simply mouse or touchscreen selections to advance the story along the desired branches. This concern drove our usability instrument to a rather simplistic one with the concerns listed below.

Usability concerns the human factors or keystroke "ilities" (usually called human computer interface issues, but recently also concerned with playability topics). Usability (UB): one's functional ability to progress through a task/intervention; UB = U + N + L, where

Understandability (U)=ability to grasp what it is one is to do;

Navigability (N)=ability to move around within the realm of what it is one is to do;

Learnability (L)=ability to master controls/ideas that allows one to do what it is one is to do;

Rememberability, a vital usability topic, is not able to be assessed as part of this study protocol.

Edutainment that is highly usable does not ensure that it is also <u>useful</u> in the mind of the player. Nor for that matter does entertainment or engagement go to this issue. Hence another aesthetic of interest is to determine to what extent the edutainment is felt to be of use by the player. One might be tempted to equate usefulness to the training dimensions; however, usefulness is more of an aesthetic since it captures the player's perception of the value and credibility of the edutainment. In prior research, the lead author has identified tests that indicate if an artifact is useful (Silverman, 1992) [40]. We adapt those here as follows.

Usefulness (UF): The degree to which the content of an intervention is clearly applicable to the particular workings of one's life in ways that are valid, helpful, and workable; UF = CL + VC + H +W, where

Clear/Logical (CL)=clarity and logic of the content in presenting applicability of intervention to one's life;

Valid/Corresponding with reality (V)=the degree to which the presentation of content corresponds with what happens in real life;

Helpful (H)=shows one what to do;

Workable (W)=allows one to practice/rehearse what it is one's to do.

3.4) Data Analysis Plan for Assessment of Newly Developed Heart Sense Instruments

Seven new assessment measures ("instruments") have been developed for the HeartSense clinical trial (KI-WTP-EEUU; see above). Some have additional dimensionality hypothesized. All instruments/dimensions will undergo psychometric assessment prior to assessing study aims that depend on instrument and/or dimension data. We will proceed as described below (without any prior imputation of missing values):

1. Examine distribution of item responses. Concern about inclusion of individual items will exist if: sample proportion answering an extreme response option is =70% or skewness is =-1 or =+1; and/or the sample proportion with a missing value for an item is =10%. Determining whether to remove an item meeting either of these criteria will depend on the importance of the item in defining the underlying construct.
2. Examine distribution of instrument scores and any hypothesized dimensions of each instrument to determine whether ceiling/floor effects exist. If =20% of the sample are located at the top-most (100) and/or bottom-most (0) scores, a ceiling and/or floor effect (respectively) will be considered to exist for the instrument. The higher this value (above 20%) the less likely the instrument/dimension will capture change over time (in follow-up).
3. Examine internal consistency reliability coefficients for each instrument and any hypothesized dimensions of each instrument. Each Cronbach's alpha should be =0.70 (although some investigators use a threshold of =0.80). If a particular instrument's/dimension's alpha is <0.70, there may be one or more items that are included in the instrument that should be dropped. If so, investigators may proceed with an analytically based approach to identifying items from the instrument that cause the reduction in internal consistency (and removing it if investigators believe the content of the item is of less importance than obtaining an internally consistent instrument/dimension). If an originally hypothesized set of items result in a Cronbach's alpha value that is too low (<0.70), and all subsequent subsets of the original set do similarly, there will be no further evaluation of the new instrument/dimension, and it should be removed from all subsequent analysis recommendations (validity assessments can only be completed for an instrument/dimension that has reliability). If a Cronbach's alpha indicates a value that is =0.90 – which indicates that there is item redundancy – investigators will consider permanent removal of 1 of any 2 or more highly intercorrelated (r>0.70) items (depending on item content).
4. Face/content validity assessment will be done with study participants and clinician experts from the HeartSense Advisory Board (the latter subgroup will assess "Intent" and "Knowledge" instruments only). The former subgroup may be asked, once they have completed all instruments, to answer close-ended questions about whether each instrument/dimension had posed questions specific to its stated content; and to answer open-ended questions about whether all questions relevant and important to each instrument/dimension had been asked.
5. Construct validity assessment will proceed using multiple, complementary approaches. A multi-trait, multi-item approach will be completed first, which will assess quantitatively whether items that are hypothesized to belong to a particular instrument/dimension are significantly more correlated to that instrument/dimension than to another hypothesized instrument/dimension. This will provide evidence at the item level that the items in each instrument/dimension assess a unique construct that is not shared with other instruments/dimensions under examination. Scaling success rates will be reported – with intent to achieve a commonly employed threshold for new instruments of =90%. Following this analysis (and any item adjustments it may require in response to results), a multi-trait, multi-method approach will be employed to confirm analytically that each of the instruments/dimensions are measuring constructs different from one another (i.e., heterotrait homomethod correlations). A related and robust method to assess quantitatively whether item responses within a particular instrument/dimension are correlated with each other in a way that is unique from item responses of other instruments/dimensions is the factor analytic approach. All items will be entered into a principal components analysis using the following parameters: correlation matrix; VARIMAX rotation; tolerance, 0.001; iterations, 25; eigenvalue =1.0; and pairwise deletion. We will complete re-analysis using the same parameters, except OBLIMIN rotation (gamma=0) will be used instead of VARIMAX – allowing for inclusion of correlations between factors. If loading across both methods is the same, we will assume that the orthogonality of the initial factor solution was not an artifact of the method of rotation. The initial factor solution should reveal that all items load into factors that mirror the hypothesized instrument/dimensions (or those achieved after completion of subsections 1-5, above). Finally, an hypothesis-testing approach will be completed using differences in the scores of different participant subgroups hypothesized, a priori, to exist for each instrument/dimension. This approach will be very difficult to pursue in the current clinical trial approach, as one of the defining subgroup features is the intervention subgroups. Thus, only sociodemographic and other background (i.e., computer game use) data can be used for this approach, and said data are quite limited (and will not allow for easy construction of subgroups for which hypothetical differences in instrument-assessed constructs might be said to differ).

Only instruments/dimensions that successfully meet criteria of subsections 1-6, above, will be included in subsequent data analysis plans.

4) Clinical Trial Protocol and Intended Analysis Plan

The current version of HEART SENSE is about to undergo a randomized trial in a clinical setting to evaluate its efficacy in terms of knowledge learning, shift of player intent to call 911, and impact upon decisional conflict. We are particularly interested in exploring the role of entertainment and engagement upon these variables. Also of interest are the usability and usfulness of various scenes and types of interventions. To help evaluate these effects we are running a 4-arm trial including the following arms:

B = Baseline (no intervention materials) – Patients are tested based on their common, everyday knowledge to provide a baseline to compare against. No interventions are used.

P = Training content only ("Pamphlet") – Patients are tested for the impact of "usual practice" which involves handing them a pamphlet. The pamphlet used here will be one produced by the American Heart Association entitled "Act in Time to Heart Attack Signs".

M = Engaging story ("Movie") – Patients are tested for the impact of a cartoon movie comprised of one path through the game. This involves passive viewing of the selections and dialogs with no interactivity on the part of the subjects.

G = Entertainment aesthetic ("Game") – Patients are tested for the impact of an interactive game that combines training, engagement, and entertainment aspects.

4.1) Clinical Trial Protocol

In order to evaluate the Heart Sense Game, the Medical School members of the team will conduct a trial with actual patients from the University of Pennsylvania Healthcare System (UPHS). Some aspects of the protocol include:

Purpose: To determine the efficacy of the Heart Sense game software in reducing delay in seeking care for heart attacks.

Duration: November 1, 2003 through February 14, 2004

Subject recruitment and selection: A total of 200 subjects, aged 35-75 years, will be recruited from the clinical practices at the Hospital of the University of Pennsylvania. These subjects, in either of three intervention arms or a fourth, baseline control group will be selected as a random sample from the patient database system. Project staff will contact each candidate by telephone (script attached) to describe the study and to solicit their interest in participating.

Research Design: After recruitment, subjects will be randomized to one of the four arms, and mailed a packet containing instructions for appearing at the study site. Upon appearing for the scheduled intervention session, the Study Coordinator will obtain informed consent from each subject, and introduce the study and the intervention. Subjects will first complete the Demographic questionnaire in an online instrument, and then be exposed to the intervention of that arm B/P/M/G. They will then be given online versions of the EE-KID-UU-WTP relevant to their arm and in that order. After one month, the subjects will be mailed another set of KID questionnaires to ascertain the persistence of any attitude change towards delay. Subjects will receive $25.00 in cash for the initial participation in the arms and another $25 and a thank you letter for completing the follow-up questionnaire one month later. Analysis will consist of evaluating the results on each instrument and comparing pre- and post-intervention scores on primary measures (KID-WTP) as well as for the secondary ones (EEUU).

Inclusion and exclusion criteria

Prospective subjects will be included in the study using the following criteria:

a. Member of a cardiovascular risk group: Enrollment in the study will be restricted to those already having at least one risk factor for myocardial infarction. Prior cardiac history, as well as the presence of more than three risk factors will exclude a candidate subject (see below). Risk factors include diabetes mellitus, obesity, and smoking.
b. Sex: subjects of either sex will be admitted to the study
c. Age: Subjects must be 45-75 years of age to participate

d. Race: Subject of all races and ethnic groups will be included

The following criteria will be used to exclude subjects:

a. History of myocardial infarction: as the focus of this project is to determine the efficacy of the Heart Sense game on reducing delay for a first coronary syndrome, subjects who have a history of myocardial infarction at any time prior to recruitment will be excluded.
b. History of heart failure, arrhythmia, valvular disease, or other cardiac pathology: There is a high likelihood that these individuals have, in the course of cardiac evaluation and therapy, been educated about the risks of delay in seeking care for acute coronary syndrome. As the efficacy of the game itself is being evaluated in this study, and co-intervention (even if previous to recruitment) must be minimized, these individuals will be excluded.
c. The presence of more than three risk factors: Given that individuals with more than three risk factors are likely to have received substantial educational materials and interventions by their physician and others, it should be assumed that they are qualitatively different from those with fewer risk factors. This, these individuals will be excluded from this study.

Randomization and allocation of subjects -- Two levels of randomization will be performed. The first is a simple, non-stratified randomization scheme that will be applied to subjects identified through the UPHS system. This randomization will have a selection fraction of 0.50, and will be implemented using a random number routine provided within the database engine. The Study Coordinator will use this routine to identify and generate a listing of candidate subjects for the study as described above. Patients who are not contacted will be characterized as to demographics using data from the database system for later comparison to assess bias at this stage of the selection process. After a candidate subject has been screened, another randomization procedure will be used to determine to which treatment arm he or she will be allocated. The allocation procedure will ensure that sufficient subjects are recruited for each specific arm of the trial over its course, but also within each week, to ensure that each arm has sufficient subjects to study each week. Each week, there will be at least eight intervention sessions, consisting of two sessions representing each arm. Subject recruitment and assignment procedures will ensure this balance.

Exploratory analyses: Before proceeding with inferential analyses, the data for this project will be fully described including aspects of data quality. Much of this work will be ongoing as standard data management procedures. A summary for each variable or group of variables will be produced for each study group (game, movie, pamphlet, baseline) at regular intervals. Graphical methods (e.g., histograms) will be employed to understand aspects of data quality and to examine distributional assumptions. Means, medians, interquartile ranges, maxima, minima, and standard deviations will be computed for all continuous variables, as data become available. Frequencies will be computed for categorical and ordinal variables. These interim analyses will be conducted only for descriptive and quality control purposes. No hypothesis testing will be performed during the course of the study.

The groups will be compared at baseline with respect to demographic characteristics to assess the success of the randomization. Note that there is no pre-intervention measurement of outcome, so comparisons of the scores on the measurement instruments pre-intervention will not be possible in this context of evaluating comparability of the groups. Rather, comparisons will be limited to pairwise (between-group) contrasts as well as those between immediate post-intervention and follow-up outcomes. Thus, assuming the groups are G=Game, M=Movie, P=Pamphlet, the contrasts of particular interest are: G vs. P, G vs., M, G vs. M+P (all at time of intervention), and then repeated at the one month follow-up. Likewise we wish to compare improvements of M, G, or P to the Baseline arm.

Analyses addressing the training dimension hypotheses -- The first analysis will be a global hypothesis test for differences in mean levels of the primary outcome of interest among the four groups. For sample size purposes, we focus on the "intention" scale, which is incorporated within the KID-WTP (Knowledge, Intention, Decisional Conflict, and Willingness To Pay, Appendix II), although we will also analyze knowledge and decisional conflict. The primary analyses will be the unadjusted intent-to-treat analyses, i.e., with no adjustment for covariates and including each participant in the group to which he/she was randomized, regardless of adherence to the assigned strategy (i.e., regardless of whether he/she completed the game or viewing the movie). Differences in the means after the initial intervention will be tested using standard analysis of variance (ANOVA) methods. Pairwise contrasts between specific groups (e.g., game versus movie) will be tested using the Student's t-test. These are planned contrasts, so we will not be using *post hoc* comparisons, but we have set the experiment-wise type I error rate to 0.01, to allow for multiple comparisons. We note that some of the variables collected (e.g., the composite score for the KID) may exhibit distributions that will require suitable transformations prior to analysis with these

parametric procedures. In addition to considering transformations, results of the t-tests will be confirmed using Wilcoxon rank sum tests and the ANOVA results will be confirmed using the Kruskal-Wallis test. If the results disagree, both will be reported to allow evaluation by the informed reader.

Aesthetics Hypotheses -- The aesthetics hypotheses will focus on outcomes of edutainment that are influenced by different design aspects. For these analyses, we need to turn to the EEUU instruments, which are in Appendix III. The actual instrument starts on page 3 ("Entertainment"). We shall focus on the major dimensions, i.e., entertainment, engagement, usability, and usefulness. Keystrokes during gameplay will also be collected and analyzed to assist in the determination of the rationale for some of the Entertainment answers. Examples of specific hypotheses to be tested follow, but in general we need to compare X vs. Y, X vs. Z and X vs. Y or Z where X, Y, and Z are alternately defined as Movie (M), Game (G), or Pamphlet (P), such that all combinational sets are exhausted. Likewise we wish to compare improvements of M, G, or P to the Baseline arm.

Sample Size Considerations -- One goal of the proposed study is to test the null hypothesis that the two population means are equal. This will be done using the two-sample t-test. The criterion for significance (alpha) has been set at 0.01. The test is 2-tailed, which means that an effect in either direction will be interpreted. Note that by powering the study to be able to contrast pairs of groups, we provide more than adequate power for a test of the global hypothesis of no difference across the four groups. With the proposed sample size of 48 for each of the groups, the study will have power of 80.3% to yield a statistically significant result in the contrast between any two groups. This computation assumes that the mean difference is -10.0 (corresponding to means of 79.0 versus 89.0) and the common within-group standard deviation is 14.0. This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

Precision for estimating the effect size -- A second goal of this study is to estimate the mean difference between the two populations. On average, a study of this design would enable us to report the mean difference with a precision (95.0% confidence level) of plus/minus 6.96 points. For example, an observed difference of -10.0 would be reported with a 95.0% confidence interval of -16.96 to -3.04. The precision estimated here is the median precision. Precision will vary as a function of the observed standard deviation (as well as sample size), and in any single study will be narrower or wider than this estimate.

References

1. Smith, K.L., Cameron, P.A., Meyer, A., & McNeil, J.J. (2002). Knowledge of heart attack symptoms in a community survey of Victoria. *Emergency Medicine, 14*(3), 255-260.
2. Kentsch, M., Rodemerk, U., Muller-Esch, G., Schnoor, U., Munzel, T., Ittel, T.H., et al. (2002). Emotional attitudes toward symptoms and inadequate coping strategies are major determinants of patient delay in acute myocardial infarction. *Zeitschrift fur Kardiologie, 91*(2), 147-155.
3. Avis, N.E., Smith, K.W., & McKinlay, J.B. (1989). Accuracy of perceptions of heart attack risk: What influences perceptions and can they be changed? *America. Journal of Public Health, 79*, 1608-1612.
4. Bleeker, J.K., Lamers, L.M., Leenders, I.M., et al. (1995). Psychological and knowledge factors related to delay of help-seeking by patients with acute myocardial infarction. *Psychotherapy & Psychosomatics, 63*, 151-158.
5. Henderson, S.O., Magana, R.N., Korn, C.S., Genna, T., & Bretsky, P.M. (2002). Delayed presentation for care during acute myocardial infarction in a Hispanic population of Los Angeles County. *Ethnicity & Disease, 12*(1), 38-44.
6. Gibler, W.B., Armstrong, P.W., Ohman, E.M., Weaver, W.D., Stebbins, A.L., Gore, J.M., et al. (2002). Persistence of delays in presentation and treatment for patients with acute myocardial infarction: The GUSTO-I and GUSTO-III experience. *Annals of Emergency Medicine, 39*(2), 123-130.
7. Canto, J.G., Shlipak, M.G., Rogers, W.J., Malmgren, J.A., Frederick, P.D., Lambrew C.T., et al. (2000). Prevalence, clinical characteristics, and mortality among patients with myocardial infarction presenting without chest pain. *JAMA, 283*(24), 3223-3229.
8. Dracup, K., McKinley, S.M., & Moser, D.K. (1997). Australian patients' delay in response to heart attack symptoms. *Medical Journal of Australia, 166*, 233-236.
9. Osganian, S.K., Zapka, J.G., Feldman, H.A., Goldberg, R.J., Hedges, J.R., Eisenberg, M.S., et al. (2002). Use of emergency medical services for suspected acute cardiac ischemia among demographic and clinical patient subgroups: The REACT trial. Rapid Early Action for Coronary Treatment. *Prehospital Emergency Care, 6*(2), 175-185.
10. Silverman, B.G., Holmes, J.H., Kimmel, S., Branas, C., Ivins, D., Weaver, R., et al. (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.
11. Silverman, B.G., Johns, M., & Weaver, R. (May 2003). Satisfying the Perceived Need for Free-Play in Pedagogically Oriented Interactive Dramas, *Conf. on Animated and Social Agents Procs.* New York: IEEE Press.
12. Silverman, B.G., Johns, M., Weaver, R., & Mosley, J. (Nov. 2003). Authoring Edutainment Stories for Online Players (AESOP): Introducing Gameplay into Interactive Dramas. *Internat'l Conf. on Virtual Storytelling*. Toulouse, FR: Springer.
13. Cacioppo, J. T., Petty, R. E., & Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.
14. Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197-253.
15. O'Connor, A.M. (1999). Decisional Conflict Scale. Ottowa: U. of Ottowa Hospital/LHRI, 1999 (4th Edition). The full scale is available at http://www.ohri.ca/programs/clinical_epidemiology/OHDEC/dcsdoc.pdf.
16. Hardin, P.C., & Reis, J. (1997). Interactive multimedia software design: Concepts, process, and evaluation. *Health Education & Behavior, 24*(1), 35-53.
17. Pennsylvania State Data Center. (2000). Pennsylvania Fact Sheet. Penn State Harrisburg Institute for State and Regional Affairs, 1-2.
18. Norman, S. (2000). Women's Contraceptive and Reproductive Experiences (CARE) Study.
19. Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.
20. Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
21. Bandura, A. (1994). Social cognitive theory and exercise of control over HIV infection. In R.J. Diclemente & J.L. Peterson (eds.), *Preventing AIDS: Theories and methods of behavioral interventions* (pp. 25-60). New York: Plenum Press.
22. Fishbein, M. (2000). *The role of theory in HIV prevention*. Philadelphia: Public Policy Center, Annenberg School for Communication, University of Pennsylvania, unpublished work.

23. Fishbein, M., Bandura, A., Triandis, H.C., Kanfer, F.H., Becker, M.H., & Middlestadt, S.E. (1992). Factors influencing behavior and behavior change: Final report -- Theorist's Workshop. Rockville, MD: National Institute of Mental Health.
24. Becker, M. (1974). The health belief model and personal health behavior. *Health Education Monographs, 2*, 324-473.
25. Rosenstock, I.M., Strecher, V.J., Becker, MJ. (1994). The health belief model and HIV risk behavior change. In R.J. Diclemente & J.L. Peterson (eds.), Preventing AIDS: Theories and methods of behavioral interventions (pp. 5-24). New York: Plenum Press.
26. Rhodes, F., Fishbein, M., & Reis, J. (1997). Using behavioral theory in computer-based health promotion and appraisal. *Health Education & Behavior, 24*(1), 20-34.
27. McArdle, J.B. (1972). *Positive and negative communications and subsequent attitude and behavior change in alcoholics.* Unpublished doctoral dissertation, University of Illinois.
28. Fishbein, M., & Coombs, F.S. (1974). Basis for decision: An attitudinal analysis of voting behavior. *Journal of Applied Social Psychology, 4*, 95-124.
29. Sheppard, B.H., Hartwick, J., & Warshaw, P.R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research, 15*, 325-343.
30. Argyris, M. and Schön, D. (1974). *Theory in practice. Increasing professional effectiveness*. San Francisco: Jossey-Bass.
31. Irwin, J. & Baron, J. (2001). Values and decisions. In S.J. Hoch, & H.C.Kunreuther, (eds.), *Wharton on making decisions*. New York: Wiley.
32. Kelly, N.R., Huffman, L.C., Fernando, M.S., & Robinson, T.N., (2003). Effects of videotape to increase use of poison control centers by low-income Spanish-speaking families: A randomized controlled trial. *Pediatrics, 111*, 21-26.
33. Campbell, J. (1973). *The hero with a thousand faces*, Princeton: Bollingen Series/Princeton University Press.
34. Green, G.S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*, 534-537.
35. Meretzky, S. (Nov 20, 2001). Building character: An analysis of character creation. *Gamasutra*. http://www.gamasutra.com/resource_guide/20011119/meretzky_01.htm
36. Decker, D. (Sept. 2002). Anatomy of a screenplay. *First Annual Theory of Story Conference (Storycon)*, Palm Springs: CA. www.anatomyofascreenplay.com/book.htm
37. Whitton, M.C. (June 2003). Making virtual environments compelling. *CACM, 46*(7), 40-46.
38. Green, M. C., & Brock, T. C. (2002). In the mind's eye: Transportation-imagery model of narrative persuasion. In M.C. Green, J. J. Strange, & T. C. Brock (Eds.), Narrative impact: Social and cognitive foundations. Mahwah, NJ: Erlbaum.
39. Green, M., & Brock, T.C. (2000). The role of transportation in the persuasiveness of public narratives. *J. of Personality and Social Psychology, 79*(5), 701-721.
40. Silverman, B. G. (1992). *Critiquing human error: A knowledge based human-computer collaborative approach*. London: Academic Press.

Appendix I - Demographics Instrument (D)

(NOTE 1: This is the only instrument applied before the intervention)

INSTRUCTIONS: As you know, we asked you to be a part of this research project to help us study the best way to educate people in some of the issues about heart disease. We will begin by asking you some questions about yourself and then ask you to _____. Once you have _____, you will be asked to complete a final set of questions. It will be quite helpful to us if you answer all the questions that have been asked to the best of your ability. Please read each question and its possible answers carefully. If you have any questions, please ask the project staff who is helping you today.

1. What is your age? _____
2. Please circle your sex: Male Female
3. Please circle your racial/ethnic group:
   African-American White Asian Other
4. Please circle whether you consider yourself to be of Hispanic heritage? Yes No
5. Do you own a computer? Yes No
6. How often do you use computers at home or at work? (Please circle one answer.)
   Never 1-4 times/year 1-2 times/month 1-2 times/week Daily
7. How often do you play computer games or videogames? (Please circle one answer.)
   Never 1-4 times/year 1-2 times/month 1-2 times/week Daily
8. What computer games do you play? (Please list a few types or names.)
9. Has anyone close to you had a heart attack? Yes No
10. Have you had a heart attack? Yes No

Need for Cognition Questions:
Based on Scales from Cacioppo & Petty (1984)

11. Some people like to have responsibility for handling situations that require a lot of thinking, and other people don't like to have responsibility for situations like that.

What about you? Do you like having responsibility for handling situations that require a lot of thinking, do you dislike it, or do you neither like nor dislike it?

☐ Like ☐ Dislike ☐ Neither like nor dislike it

12. How much do you like/ dislike it?

☐ A lot ☐ Somewhat

13. Some people prefer to solve simple problems instead of complex ones,
whereas other people prefer to solve more complex problems. Which type of problem do you prefer to solve: simple or complex?

☐ Simple ☐ Complex

Appendix II - Training Instruments (KID-WTP)

Knowledge Questionnaire (K)

Completion of the "Knowledge" questionnaire will provide a score (K) that indicates each participant's intent to call 911 in the event s/he is having a heart attack. The total score is obtained from the combined scores of two tasks and four subscales (respectively): accurate generation of heart attack symptoms (AG); accurate recognition of heart attack symptoms (AR); knowledge of symptom pattern possibilities (S); knowledge about risk factors for heart attacks (RF); knowledge about issues surrounding treatment (T); and knowledge about one's plan for action in the event one were to have a heart attack (PA). (Not all elements of each hypothesized knowledge subscale nor all hypothesized knowledge subscales could be tested since inclusion of items in the questionnaire was dependent on what was covered in all planned intervention arms.) Overall, we hypothesize that K = AG + AR + S + RF + T + PA (see various notes below about encoded values).

Items:

1. In the box, below, please TYPE in FOUR symptoms that suggest someone might be having a heart attack.*

* This is a knowledge generation task. Variations on "chest pain," "shortness of breath," "sweating," "shoulder pain," "racing heart," "dizziness," "nausea," and "tiredness" will be accepted. One point per accurately generated symptom will be added, up to a total of +4. One point per inaccurately generated symptom will be deducted, down to a value of -4.

2. In the box, below, please CHECK off SIX symptoms that suggest someone might be having a heart attack.†

Headache	Chest pain	Blood in urine	Forgetfulness
Dizziness	Muscle aches	Blurred vision	Sweating
Tiredness	Shortness of breath	Shoulder pain	Nasal congestion
Racing heart	Diarrhea	Lower back pain	Nausea

† This is a knowledge recognition task. Six symptoms of the following choices will be acceptable: "chest pain;" "shortness of breath;" "sweating;" "shoulder pain;" "racing heart;" "dizziness;" "nausea;" and "tiredness." One point per accurately recognized symptom will be added, up to a total of +6. One point per inaccurately recognized symptom will be deducted, down to a value of -6.

3. A person having a heart attack always has pain in the chest. (S)‡

4. Heartburn can be a symptom of a heart attack. (S)

5. If you think you are having a heart attack, you should . . . (PA)
(Response options, for this item, are: "wait for an hour to see if the symptoms go away;" "call a friend or relative to get his or her opinion;" "go directly to your doctor's office;" and "call 9-1-1 immediately." All choices other than the last one will have a value of "0" encoded. The last, correct choice will have a value of "1" encoded.)

6. There are treatments that help people having a heart attack, but only when they are given early after recognizing the symptoms. (T)

7. Women are at low risk of having a heart attack. (RF)‡

8. An emergency room will refuse to treat someone with a heart attack who does not have some way to pay for care. (T)‡

9. People who exercise do not have heart attacks. (RF)‡

10. People under the age of 60 can be at risk for heart attacks. (RF)

11. People in some racial groups do not have heart attacks. (RF)‡

‡ Except where indicated otherwise, encoded values for response options are "false" = 0; "true" = 1. Encoded values for response options of denoted items, however, are reversed.

Intent Questionnaire (I or stated intent)

Completion of the "Intent" questionnaire will provide a score (I) that indicates each participant's intent to call 911 in the event s/he is having a heart attack. The total score is obtained from the scores of three subscales: recognition of symptoms that indicate a heart attack is likely enough for call to 911 (R); likelihood for being affected by barriers to calling 911 in the presence of heart attack symptoms (B); and perceived "immunity" from having a heart attack (IM). Overall, we hypothesize that I = R + B + IM (see note below about encoded values).

For each statement, please put an **X** in the box that best states whether you agree or disagree:

	Strongly Agree	**Agree**	**Neither Agree nor Disagree**	**Disagree**	**Strongly Disagree**

Items:

1. If I had indigestion that was not relieved by antacids, I would call 911 or go to the emergency room. (R)

2. If I had a mild discomfort in my chest and broke out in a sweat, I would call 911 or go to the emergency room. (R)

3. If I had a dull ache in my shoulder and shortness of breath, I would call 911 or go to the emergency room. (R)

4. If I started having heart attack symptoms when other people were around, I would not call 911 or ask for help because I would be embarrassed about wasting people's time if it turned out I didn't have a heart attack. (B)*

5. If I started having heart attack symptoms when other people were around, I would not call 911 or ask for help because I would be worried that they would think I was weak. (B)*

6. If I started having heart attack symptoms when other people were around, I would call 911 or ask for help even though I might be embarrassed about all the fuss it would cause. (B)

7. If I started having heart attack symptoms while I was busy with an important task (for example, cleaning a dirty house, or mowing an overgrown lawn), I would wait until I was finished before calling 911 or going to the emergency room. (B)*

8. I would not call 911 if I had heart attack symptoms because I don't think a heart attack can happen to me. (IM)*

* Encoded values for response options are "strongly disagree" = 1; "disagree" = 2; "neither agree nor disagree" = 3; "agree" = 4; and "strongly disagree" = 5. Encoded values for response options of items with an asterisk, however, are reversed.

Decision Conflict Instrument

Source: Adapted from [O'Connor, 1999]

For each statement, please put an **X** in the box that best states whether you agree or disagree:

	Strongly Agree	**Agree**	**Neither Agree nor Disagree**	**Disagree**	**Strongly Disagree**

1. If I were having symptoms of having a heart attack, the decision to call 911 would be hard for me to make.

2. If I were having symptoms of having a heart attack, I would be unsure of what to do.

3. If I were having symptoms of having a heart attack, the choice to call 911 would be clear to me.

4. I am aware of the choices I have to make if I had symptoms of a heart attack.

5. I feel that I know the benefits of calling 911 if I am having symptoms of a heart attack.

6. I feel that I know the risks of calling 911 if I am having symptoms of a heart attack.

7. If I were having symptoms of having a heart attack, the decision to call 911 would be hard for me to make.

8. I need more advice and information about the choices available to me if I were having symptoms of a heart attack.

9. I know how important the benefits of calling 911 for a heart attack are to me if I were having symptoms of a heart attack.

10. I feel pressure from others in making the decision to call 911 if I were having symptoms of a heart attack.

11. If I were having symptoms of having a heart attack, I have the right amount of support from others in making the choice to call 911 or not.

12. If I were having symptoms of having a heart attack and I called 911, I would feel that I have made the correct choice.

13. My decision to call 911, if I were having symptoms of having a heart attack, shows what is important to me.

14. If I were having symptoms of having a heart attack and I called 911, I would stick with my decision to be transport3ed to the hospital and be treated.

Intent2: Willingness to Pay Instrument (WTP)
Heart Sense Willingness to Pay Instrument

Page 1

1. Do you have a close personal friend or family member who is at risk of having a heart attack?
   If yes, see question 2
   If no, go to section 2.a

2. Do you think this person might delay calling 9-1-1 like some of the people in the pamphlet/movie/game?
   If yes, proceed to page 2
   If no, go to page 2.a.

Page 2

You have the opportunity to give your friend or family member the ability to instantly call 911 for help via a pushbutton pager that attaches to their keychain.

- If a person has a medical emergency, they just have to push a single button.
- The pager lets the emergency center know where the person is located, and send an ambulance to come automatically to the rescue of the person in trouble
- The pager includes a cancellation signal in case the button is pushed by accident.

You are being paid $25 for your participation today. How much of this fee would you be willing to pay for us to send this device to your friend or family member? (If you would not want to have this device sent to your friend or family member, please enter $0.)

Please enter amount (between $0 and $25): _____

Note: A picture of the imaginary device will be on the webpage.

Page 2.a

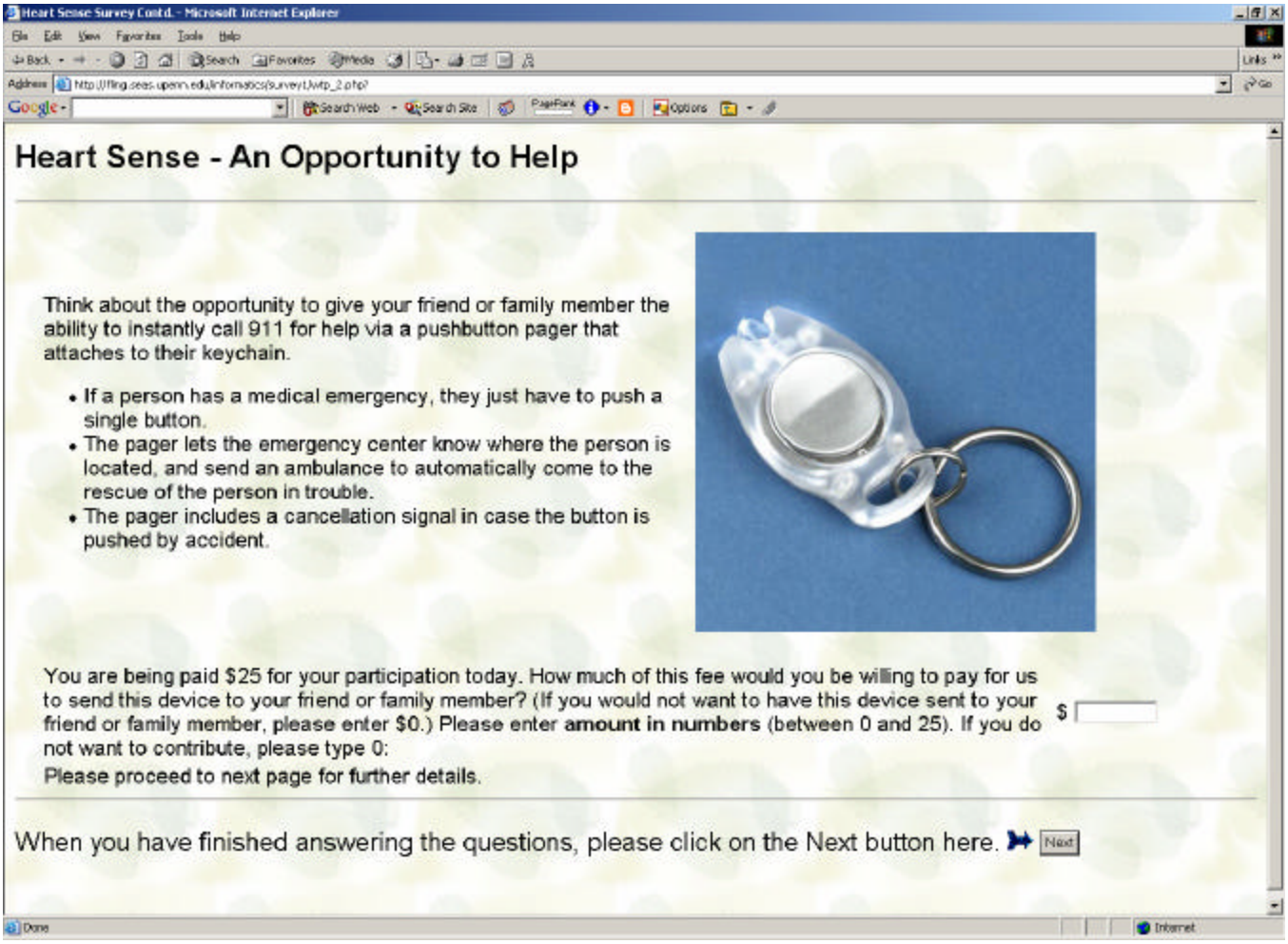
Imagine you have the opportunity to give the friend or family member you are imagining the ability to instantly call 911 for help via a pushbutton pager that attaches to their keychain.

- If a person has a medical emergency, they just have to push a single button.
- The pager lets the emergency center know where the person is located, and send an ambulance to come automatically to the rescue of the person in trouble
- The pager includes a cancellation signal in case the button is pushed by accident.

You are being paid $25 for your participation today. How much of this fee would you be willing to pay for us to send this device to the friend or family member you are imagining? (If you would not want to have this device sent to your friend or family member, please enter $0.)

Please enter amount (between $0 and $25): _____

Note: A picture of the imaginary device will be on the webpage.



Participants will be shown a device identical to the one above. The wording of the question on this particular web page is going to be replaced by the Heart Sense Willingness to Pay Instrument.

Page 3

Why did you decide to pay the amount you did for the pager? Please type the answer here.

Page 4

The prior questions about an emergency pager were part of today's experiment. We are not actually providing the ability to order such devices, and you will receive the full $25 amount you were promised for today's participation. There are necklace devices and home alarm add-ons that can facilitate alerting 911 in the event of an emergency. If you would like information about this to take home, please ask the study coordinator for information as you leave.

Handout if Subjects Request It

Heart Sense

Home Emergency Devices Contact List

Emergency Response Telephones

Ameriphone:	(800) 874-3005	$ 229.00-259.00
Fare Thee Well Phones:	(866) 892-4387	$ 199.00- 79.99

Home Security

Brinks:	(800)-725-3537
ADT:	(888)-238-7088

Heart Sense Willingness to Pay Instrument
Follow-up Questionnaire

1. Since you came in for your visit one month ago, have you asked for information on 911 alert devices? (Yes/No)

2. If Yes, what device(s) have you looked into?

Appendix III - Aesthetic Instruments (EEUU)

ENGAGEMENT (DRAMA) INSTRUMENT*

(Items not relevant to pamphlet: 5, 6, 12-13)

For each statement, please put an **X** in the box that best states whether you agree or disagree:

	Strongly Agree	**Agree**	**Neither Agree nor Disagree**	**Disagree**	**Strongly Disagree**

1. The <game/movie/pamphlet> affected me emotionally. (T)
2. I could picture myself in the events of the <game/movie/pamphlet>.(T)
3. I found my mind wandering during the <game/movie/pamphlet>. (T)
4. When I was done with the <game/movie/pamphlet>, I found it easy to put it out of my mind. (T)

5. I was interested in how Jack's relationships with various characters worked out in the end. (P1)
6. I was interested in how Jack's career plans worked out in the end. (P1)
7. Over time, I was LESS interested in how each new heart attack situation worked out. (P1)

8. I personally identified with at least one character. (P2)
9. Some of the characters reminded me of people I know (P2)
10. The characters' voices were NOT believable. (P2)
11. Various characters became too preachy. (P2)

12. The animation (characters' movement) added life to the <game/movie/pamphlet>. (P3)
13. Sound effects added realism to the <game/movie/pamphlet>. (P3)
14. Settings pictured in the <game/movie/pamphlet> were like everyday places in my life. (P3)
15. Situations that the characters experienced were like real life. (P3)

---

* Definition of "engagement"(ENG): the level to which one is personally involved in and affected by a story. (TRANSPORT, T, seems to define the top level and its been subsumed here). Also, there are three major subdomains also worth measuring:

PLOT (P1) - Degree to which one is interested in themes raised and conflicts resolved.
PEOPLE (P2) - a) Degree to which characters are "like me" and/or "like people close to me."
b) Degree to which the characters are interesting to me.
c) Degree to which character development occurs.
PLACE (P3) - a) Degree to which one can transfer the story experience to one's real life.
b) Degree to which actual sensate experience (i.e., sound, visuals, smells, terrain, sets, etc.) of story is "like real."

Overall, we might state that

ENG = T + P1 + P2 + P3

ENTERTAINMENT (GAMEPLAY) INSTRUMENT*

(NOTE: This is the only instrument that I think won't apply across all arms. In this section, only 1 applies to the pamphlet 1, 2, 5, 6, 13 apply to movie version.)

For each statement, please put an **X** in the box that best describes how much you agree or disagree with it. The correct answer will be different for each person:

	Strongly Agree	**Agree**	**Neither Agree nor Disagree**	**Disagree**	**Strongly Disagree**

1. The <game/movie/pamphlet> was fun. (F)
2. I found myself thinking of ways the <game/movie> could turn out differently if I replayed it. (F)
3. I was always able to make Jack say what I wanted him to say. (C)
4. I was NOT always able to make Jack do what I wanted him to do. (C)
5. I wanted more choices in moving around within scenes. (R)
6. I wanted more choices in moving between scenes (pizza, baseball, etc.). (R)
7. The choices I made in each scene had the results I expected in that scene. (O)
8. The choices I made over the course of the whole game affected the final outcome in the warehouse. (O)
9. I was able to affect Jack's career plans. (AT)
10. I was able to affect Jack's relationship with various characters (AT)
11. I was able to help a number of heart attack victims. (AT)
12. The story had characters that tried to block my goals for Jack (AT).

---

* - Here we are trying to isolate the overall fun score (F) as well as four contributing sub-scores of (R) "Rules" of the game were satisfying, (C) aesthetic of creating your own story via player "Control" and use of tactics of your choosing to deal with dilemmas along the way, (O) fairness of "outcomes" along the way and in the end (did things you do have logical consequences) – social contract between designer and player; and (AT) aesthetic of reward-punishment or accumulate-threaten, thus did the game provide you an opportunity to collect things, threaten you with their loss (scare/thrill), allow you to try and protect them, etc.
Overall, we might say that

ENT = F + R + C + O + AT

KEYSTROKE METRICS:

By collecting Node-ID and time stamps when dialog buttons open and then are selected, we should be able to derive most of the following:

Total clicks through the <game/movie/pamphlet>
Total nodes through the <game/movie/pamphlet>
Elapsed wall clock time
Time/node and Distribution of time/node (mean, std dev)
Dialog path chosen
Ending viewed
# of strayings off chosen path (node count)
# of nodes spent on bad path trying to appear good
# of nodes in ambiguous path (regions)
Whether the ESC key was selected

For each statement, please put an **X** in the box that best states whether you agree or disagree:

	Strongly Agree	**Agree**	**Neither Agree nor Disagree**	**Disagree**	**Strongly Disagree**

USABILITY (UB) *

The following questions concern your experience with the <G/M/P> and not with the questionnaires online:

1) During the <game >, the opening instructions were clear (L) . . ? ? ? ?
2) It was difficult figuring out how to
proceed through the <game/movie/pamphlet> (N) . . .. ? ? ? ?
3) What I was supposed to get out of the stories in the <game/movie/pamphlet> was clear to me
(U) . . . . . .? ? ? ?
4) I wasn't sure what the ultimate goal was of the
<game/movie/pamphlet> (L) . . . . . ? ? ? ?

_____

* Usability concerns the human factors or keystroke "ilities" (usually called human computer interface issues)
Usability: one's functional ability to progress through a task/intervention;
Understandability (U)=ability to grasp what it is one is to do;
Navigability (N)=ability to move around within the realm of what it is one is to do;
Learnability (L)=ability to master controls/ideas that allows one to do what it is one is to do;
Rememberability is not able to be assessed as part of this study protocol.

USEFULNESS (UF) *

1. The <G/M/P> gave me helpful ideas for what to do when someone has a heart attack. (H)
2. I would use this <G/M/P> again to refresh my memory about heart attacks. (H)
3. This <G/M/P> would be helpful for someone with heart problems. (H)
4. The heart attack victims behaved like real people do when told they are having a heart attack. (V)
5. The game/movie/pamphlet allowed me to rehearse what I would do if I was with someone having a heart attack. (W)
6. Various characters acted as guides for what I was supposed to do. (CL)
7. I learned something about heart attack symptoms. (W)
8. I learned something about the importance of calling 911. (W)
9. I now understand better why people delay care for themselves when having a heart attack. (W)

_____

* - Usefulness: The degree to which the content of an intervention is clearly applicable to the particular workings of one's life in ways that are valid, helpful, and workable;
Clear/Logical (CL)=clarity and logic of the content in presenting applicability of intervention to one's life;
Valid/Corresponding with reality (V)=the degree to which the presentation of content corresponds with what happens in real life;
Helpful (H)=shows one what to do;
Workable (W)=allows one to practice/rehearse what it is one's to do.
UF = CL + VC + H +W