

Intelligent User Experience Engineering

Mark Neerincx

Evaluation

Example: SuperAssist

Module 6



Evaluation

- Evaluation is central and any of the other design activities will be followed by an evaluation, e.g.
 - the designer checking through to make sure something is complete and correct
 - a high level design brief that is sent to a client,
 - a formal evaluation of a functional prototype by the future system users

Establish context of evaluation

IMPACT (Benyon et al.)

- Intention: Clarify aims/objectives of evaluation
- Metrics and measure: What, how and why
- People: Target group and participants in evaluation
- Activities: Derive action list from scenarios
- Context: Social and physical aspects
- Technologies: Hard- and software

Two Types of Evaluation

Formative evaluation:

- identifying usage problems with consequences (severity) **and** generating solutions (with priorities)

Summative evaluation:

- assessing the quality of (alternative) user interfaces

Evaluation techniques

- **Model-based** - Evaluator can work through the model - e.g. counting the number of actions needed, or checking for consistency
- **Expert-based** - People experienced in interface design are asked to take the role of less experienced users and describe the potential problems they foresee arising for such users.
- **Observational evaluation** involves watching people and collecting data that provides information about what users do when they interact with a system.
- **Co-operative evaluation** is when the expert observes and helps. People are encouraged to 'think aloud' about the problems they are having.

Expert-Based: Heuristic Evaluation

(Benyon et al., Ch. 3)

1. Visibility
2. Consistency
3. Familiarity
4. Affordance
5. Navigation
6. Control
7. Feedback
8. Recovery
9. Constraints
10. Flexibility
11. Style
12. Conviviality (“polite software”)

Participatory Heuristic Evaluation

(Benyon et al., Ch. 21)

1. System status
2. Task sequencing
3. Emergency exits
4. Flexibility and efficiency of use
5. Match between system and the real world
6. Consistency and standards
7. Recognition rather than recall
8. Aesthetic and minimalist design
9. Help and documentation
10. Help users recognize, diagnose and recover from errors
11. Error prevention
12. Skills
13. Pleasurable and respectful interaction
14. Quality work
15. Privacy

Other Expert-Based: Heuristic Evaluation

10 General Guidelines *(Nielsen)*

- simple & natural dialog
- speak user's language
- minimize memory load
- be consistent
- provide feedback
- clearly marked exits
- provide shortcuts
- good error messages
- prevent errors
- good help & documentation

8 Golden Rules *(Shneiderman)*

- consistency
- shortcuts
- feedback
- closure of dialogs
- error prevention and handling
- reversal of actions
- internal locus of control
- memory load reduction

And another—CE-based—approach...

User Interface at Two Levels

Task level

- based on users' goals and information needs, the system's functions and information provision are specified or assessed

Communication level

- the control of the functions and the presentation of the information is specified or assessed (i.e. the “look-and-feel”)

13 Guidelines

Task Level

1. User fit
2. Goal conformance
3. Information needs conformance
4. User's complement
5. Work context

Communication Level

6. Consistency
7. Compatibility
8. Usage context
9. Structure and pattern
10. Feedback and mode-awareness
11. Interaction load
12. Integrated support
13. User control and tailoring

Evaluation techniques (continued)

- Survey evaluation
 - Using similar methods to requirements engineering - interviews, questionnaires, etc. - but the focus is on seeing if you have got it right.
- Experimental evaluation
 - May be performed in a usability laboratory, so that an evaluator can manipulate a number of factors associated with interface design and study their effects on various aspects of user performance.
 - May be set up in a computer laboratory with little interruption from evaluator where people undertake benchmark tests

Analytical vs. Empirical Evaluation

=> Analytical Evaluation: theory, model, experience-based

- heuristic evaluation: possible problems
- claims analysis: design tradeoffs
- (model-based approach (e.g. GOMS): limited application)

=> Empirical Evaluation: severity ratings help prioritize

- field studies
- lab tests
- controlled experiments

⇒ Usability test: Effectiveness, efficiency and satisfaction.

⇒ User experience sampling: usability and additional measures (e.g. trust, emotion, situation awareness) during realistic usage period

Identifying Usability Problems

	Truth	
UEM claims that	Real Problem Exists	No Problem Exists
A&B are problems	Hit	False Alarm
C&D are NOT problems	Miss	Correct Rejection

Experiments comparing Usability Evaluation Methods:

- summing up hits & false alarms
- ignoring misses and correct rejection

Quality of the Test Methods

Reliability

Obtaining similar results for repeated measurements

- Address user diversity => representative sample

Validity

Results reflect the “claimed” usability aspects

- Realistic tasks in realistic work environment
- Take care for “confounding”

Test Plan

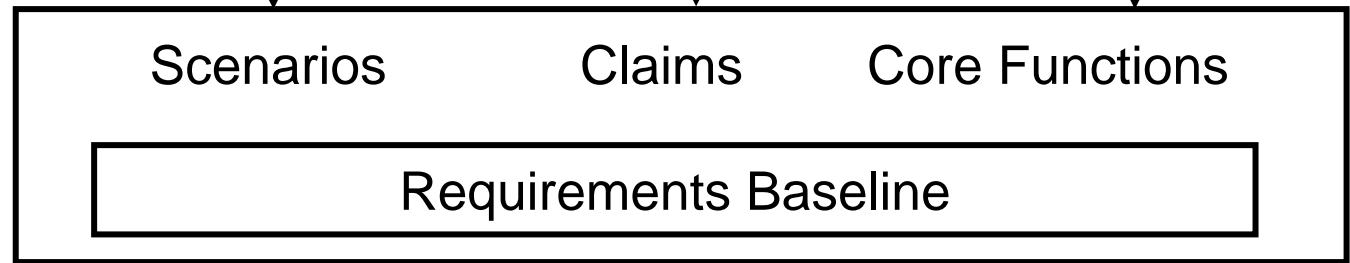
- Objective (cf. formative/summative)
- Participants (expert/novice, pay-off ratio, design for all)
- Tasks
- Procedure (preparation, introduction, test, debriefing)
- User experience measures (effectiveness, efficiency, satisfaction; e.g. trust, emotion, situation awareness)
- Pilot
- Setting (in lab/on location)
- Data analysis
- Report and presentation

sCE for SuperAssist

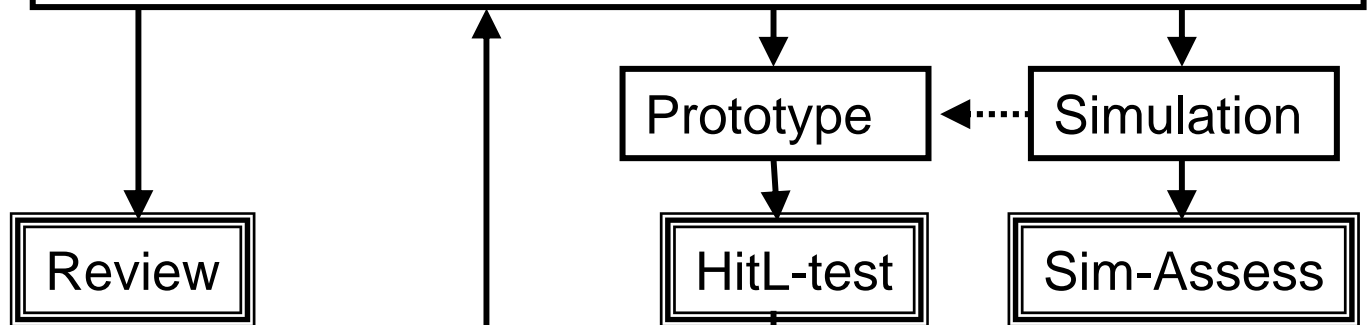
Derive



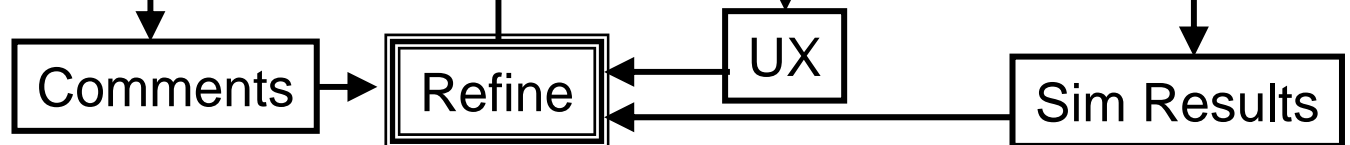
Specify



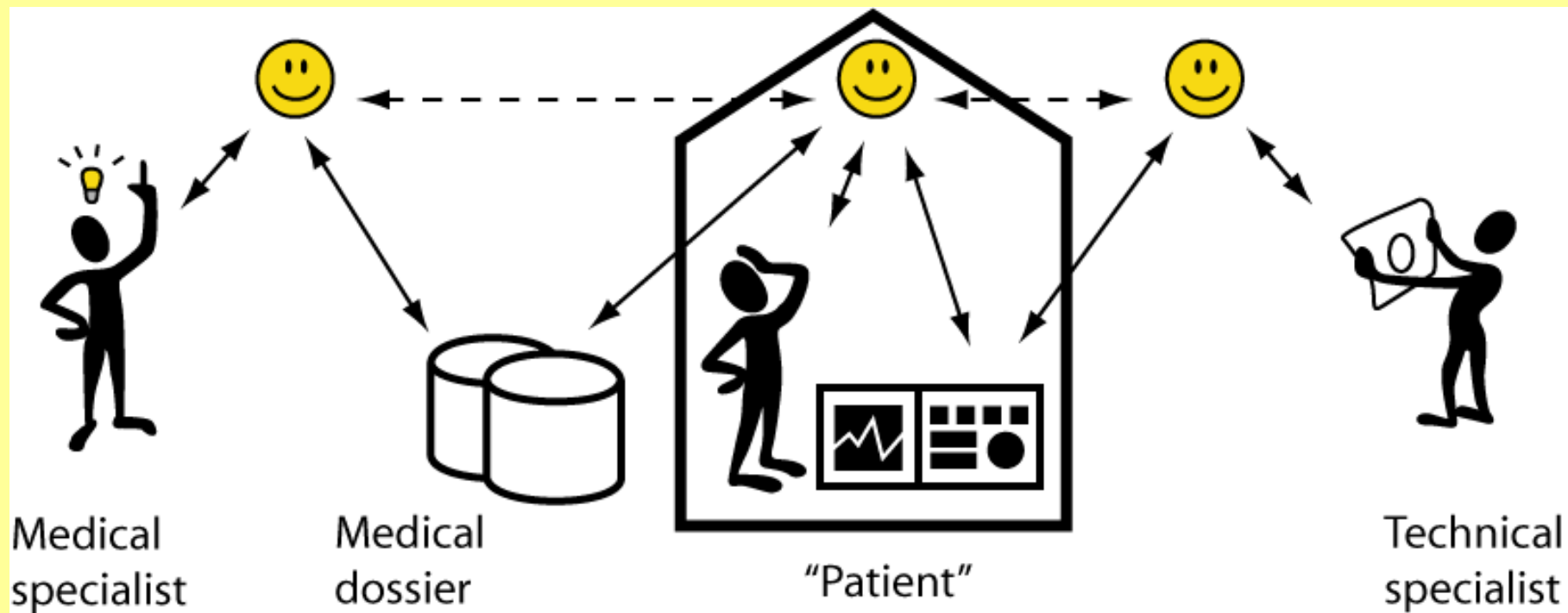
Test



Refine



SuperAssist Concept



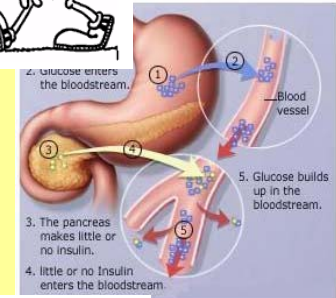
Scenario: Diabetes Type II

John (aged 58), enjoys his full time occupation as an attorney. Combining his career with his social and family life leaves little room for maintaining a healthy lifestyle.

Lately, he experiences some trouble with his health. He experiences increased thirst, frequent urination at night, and moodiness. During the last visit, the physician assessed John has **Diabetes Type II**.

The physician strongly recommends him to perform more **self-care activities**, including: maintaining a healthy diet, performing exercise regularly, monitoring his glucose level and taking medication.

Key issues for John are **combining self-care with his daily tasks** while maintaining a good quality of life.



SuperAssist

Design Specifications – Core Functions

Support troubleshooting medical instruments

Stimulate adherence to self-care objectives

Features:

- Retrieving, sharing and managing information
- Maintaining diary and calendar
- Communication

Assistant Feedback Styles

Assistant and User Characteristics

	Cooperative Feedback Style	Directive Feedback Style
Assistant	Coaching Educating Advising Oriented towards satisfaction and long-term development	Directing Reporting Dictating Oriented towards quick problem solving
User	High participation level Committing	Low participation level Complying

Assistant Feedback Styles

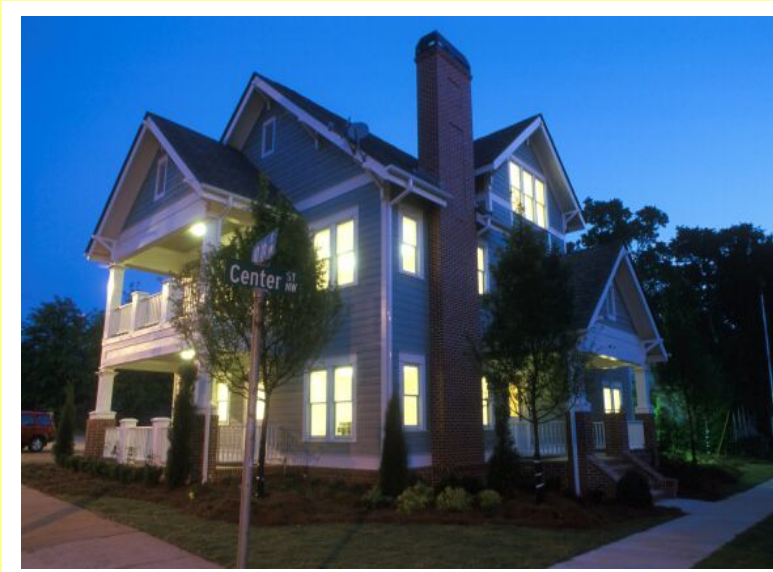
Advantages and Disadvantages

	Cooperative Feedback Style	Directive Feedback Style
Advantages	<p>Learn new competencies and develop understanding</p> <p>Better performance in long-term</p> <p>User-assistant complementing</p>	<p>User needs few competencies</p> <p>Better performance in short-term</p> <p>Vigorous acting due to expert assistant</p>
Dis-advantages	<p>Assistant support can become tedious and patronizing</p>	<p>Vulnerable to mistakes when participation is required</p> <p>User loses idea of control</p>

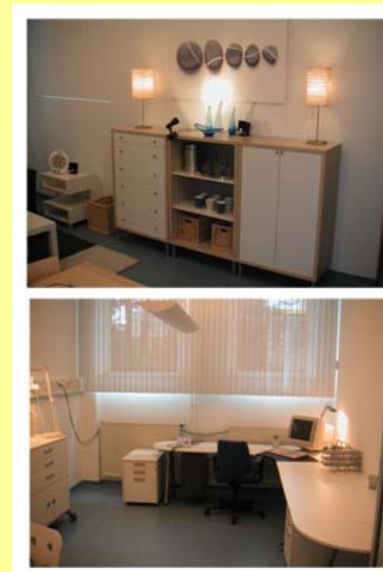
Experiment in Smart Home Lab

Blanson Henkemans, O.A., Rogers, W.A., Fisk, A.D., Neerincx, M.A., Lindenberg, J. & van der Mast (2008). Usability of an adaptive computer assistant that improves self-care tasks and health literacy of older adults. *Methods of Information Medicine*, Vol. 47(1), 82-88.

Home like atmosphere for assessment of natural behavior



GT AwareHome



TNO/TUD Experience Lab

Adaptive computer assistant

Patient Situation

- Normal
- Health-Critical

Assistant Type

- Fixed
- Adaptive

		Patient Situation	
		Normal Situation	Health-Critical Situation
Assistant Type	Fixed	Cooperative feedback style	Cooperative feedback style
	Adaptive	Cooperative feedback style	Directive feedback style

0:00

Perform Task

1:00

Perform Task

2:00

Perform Task

3:00

Exercise: Aerobic Exercise 31-40 min
 Symptom: Sweating
 Symptom: Blurry vision
 Glucose Level: 3.0
 Symptom: Sweating

Perform Task

4:00

Perform Task

5:00

You are experiencing :
 * a minor hypoglycaemic attack.


To improve your health situation,
 I have the following suggestions ==>

Please check the box next to which solution(s)
 you have chosen and press the "Finish" button
 when done

Yes No Finish

My status

Minor hypoglycaemic attack
 Blood Pressure normal



Electronic Patient Record

View Old Diary

Eat a piece of fruit

Eat a cookie

Start Chat with Nurse Stop Chat with Nurse

[Empty chat area]

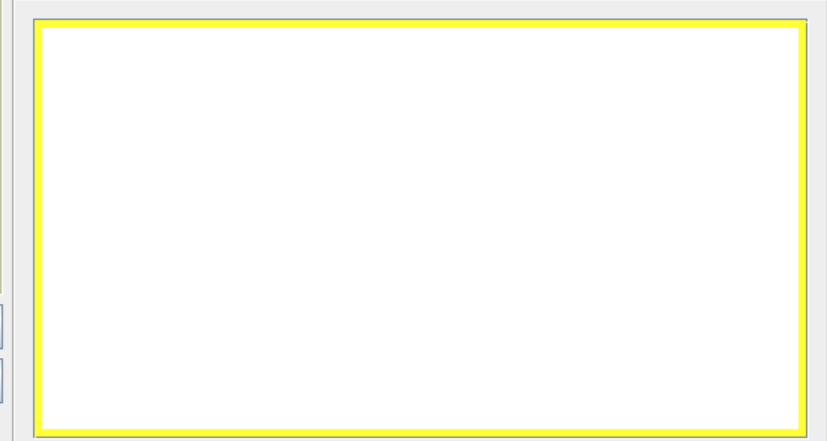
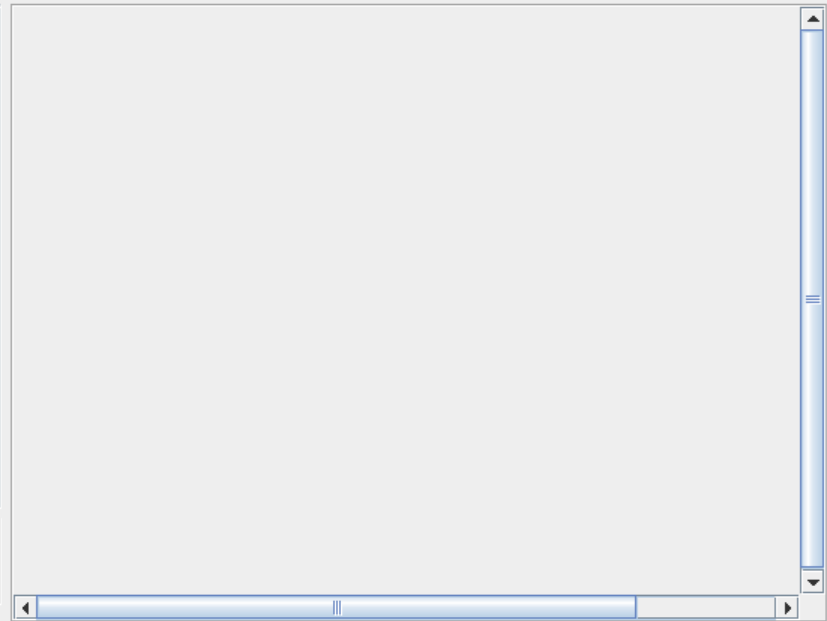

[Input field] Send

14:00	
	Perform Task
15:00	
	Perform Task
16:00	
	Perform Task
17:00	Exercise:Aerobic Exercise 31-40 min Symptom: Sweating Symptom: Blurry vision Glucose Level: 3.0
	Perform Task
18:00	
	Perform Task
19:00	

According to your diary it seems that you're feeling not so well. Is this true?

My status

Minor hypoglycaemic attack
Blood Pressure normal



0:00

Perform Task

1:00

Perform Task

2:00

Perform Task

3:00

Exercise: Aerobic Exercise 31-40 min
 Symptom: Sweating
 Symptom: Blurry vision
 Glucose Level: 3.0
 Symptom: Sweating
 Meal: Fruit servings: 1

Perform Task

4:00

Perform Task

5:00

Hypoglycaemic attacks are caused by a low level of glucose in your blood. It can make you feel uncomfortable. However, this can be prevented by taking a few simple steps:


- * Watching what you eat / Minding your nutrition
- * Trying to organize your daily schedule better and prevent stressful situations
- * Take measurements more often for an accurate overview of your glucose level

Press the "Finish" button.

Yes No Finish

My status

Minor hypoglycaemic attack
 Blood Pressure normal



Electronic Patient Record

View Old Diary

Start Chat with Nurse

Stop Chat with Nurse

Chat area (empty)

Send


14:00	Yellow task area
	Perform Task
15:00	Yellow task area
	Perform Task
16:00	Yellow task area
	Perform Task
17:00	Exercise: Aerobic Exercise 31-40 min Symptom: Sweating Symptom: Blurry vision Glucose Level: 3.0
	Perform Task
18:00	Yellow task area
	Perform Task
19:00	Yellow task area

To diagnose your health situation, please indicate which of the following symptoms you are experiencing.

Please place a check mark next to the relevant symptom(s) and press the "Finish" button when done.

My status

Minor hypoglycaemic attack
Blood Pressure normal



- You feel agitated.
- You feel like laughing nervously.
- You are irritated/in a bad mood.
- You feel aggressive.
- You are behaving differently.
- You feel tired.
- You feel like you might lose consciousness.
- You are sweating more than normal.
- You feel shaky.

Computer Assistant Hypotheses

Adaptive computer assistant more effective and time efficient

Preference for adaptive assistant

Increased knowledge of diabetes

Participants

Twenty-eight older adults

- 15 male
- 13 female
- Mean age 67
- Majority with college degree or higher

Variables

Effectiveness

- Errors made

Efficiency

- Time on task
- Mental effort (NASA TLX)

Satisfaction

- Preference for assistant type

Moderating effects of personal characteristics

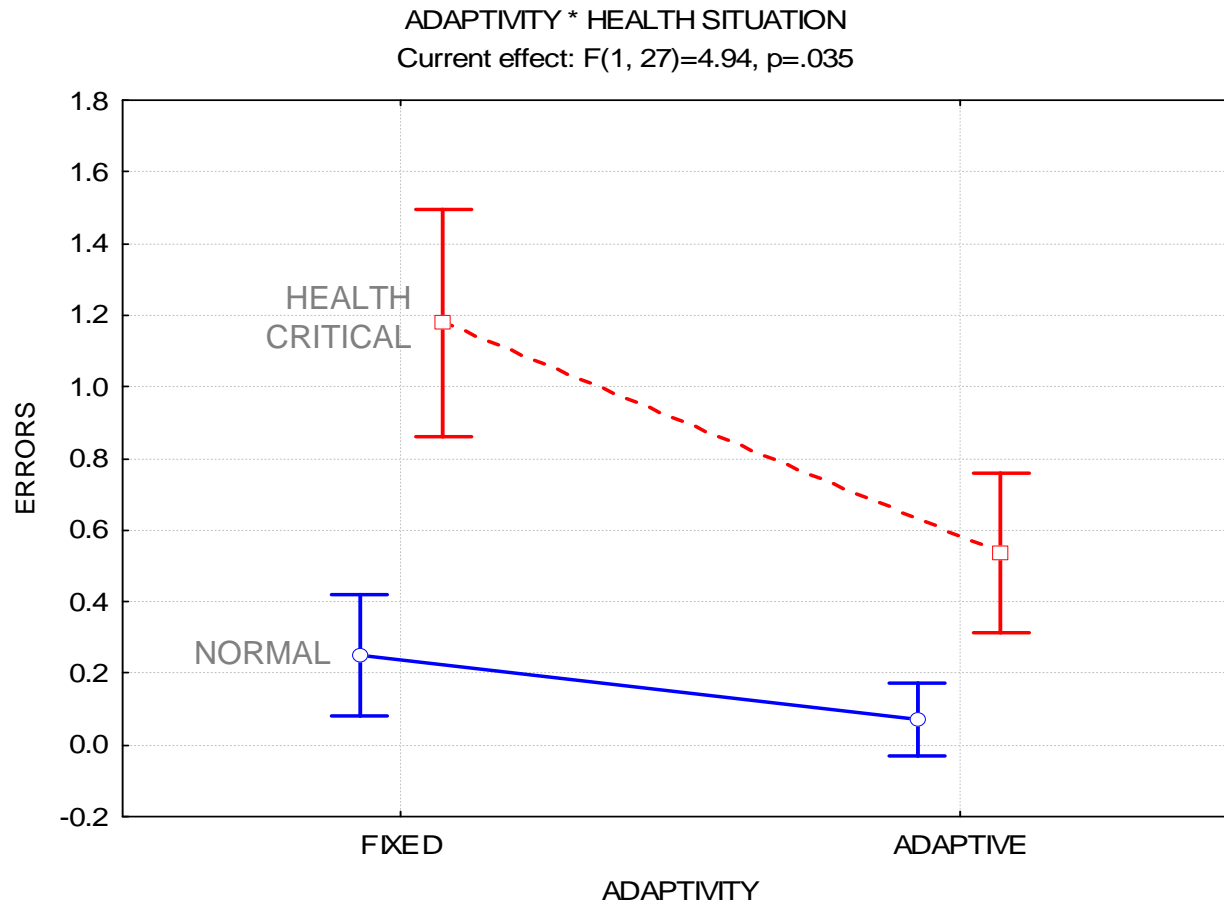
- Cognitive abilities
- Personality traits

Educational value

- A diabetes knowledge survey after each condition

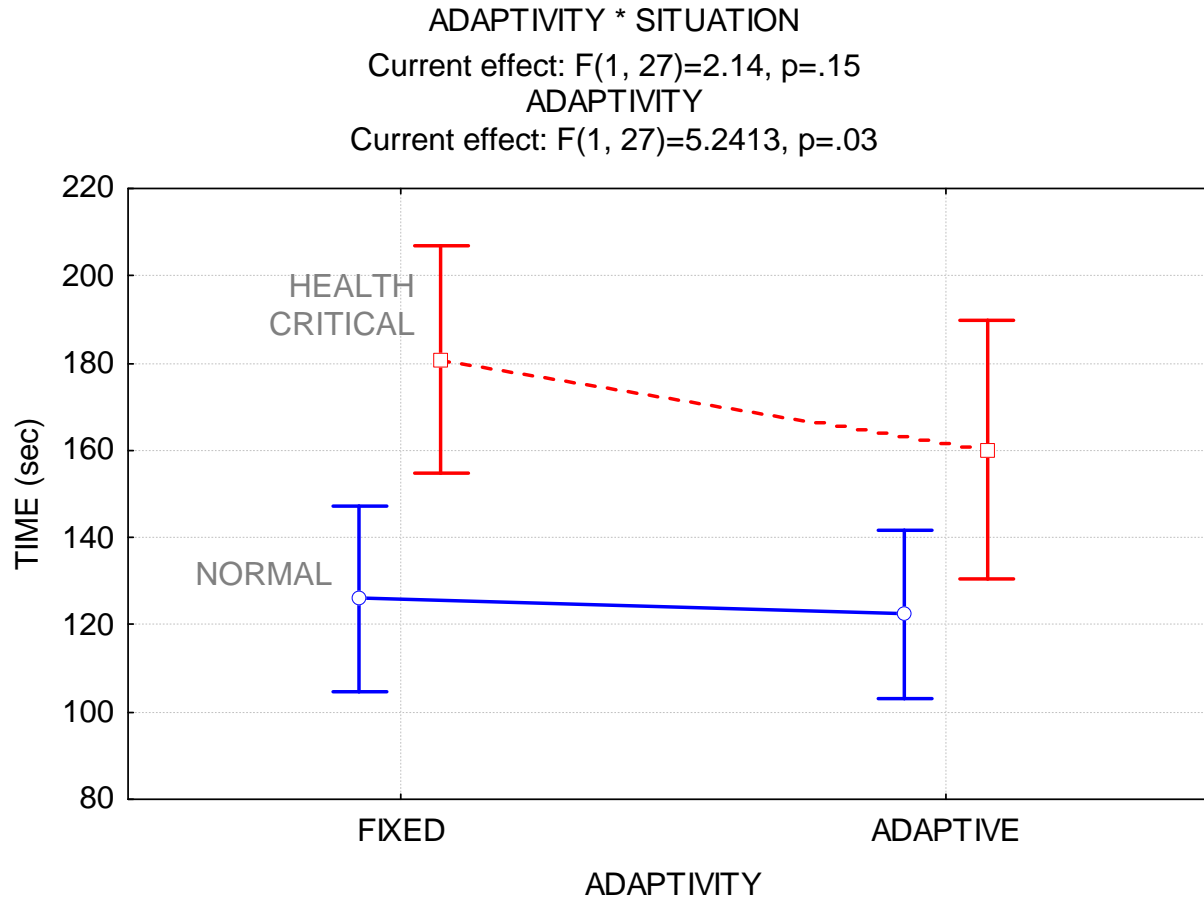
Results

Effectiveness Measured by Quantity of Errors



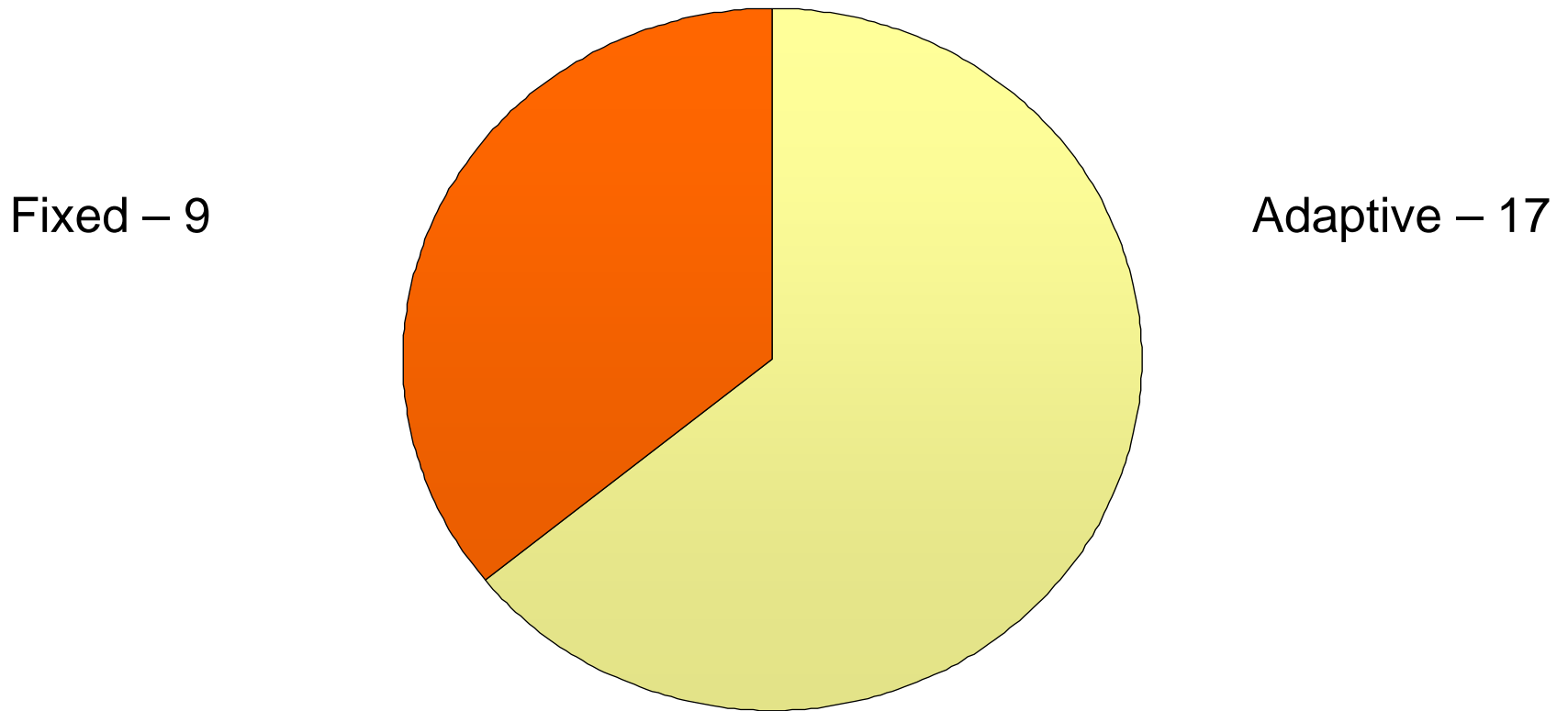
Results

Efficiency Measured by Time



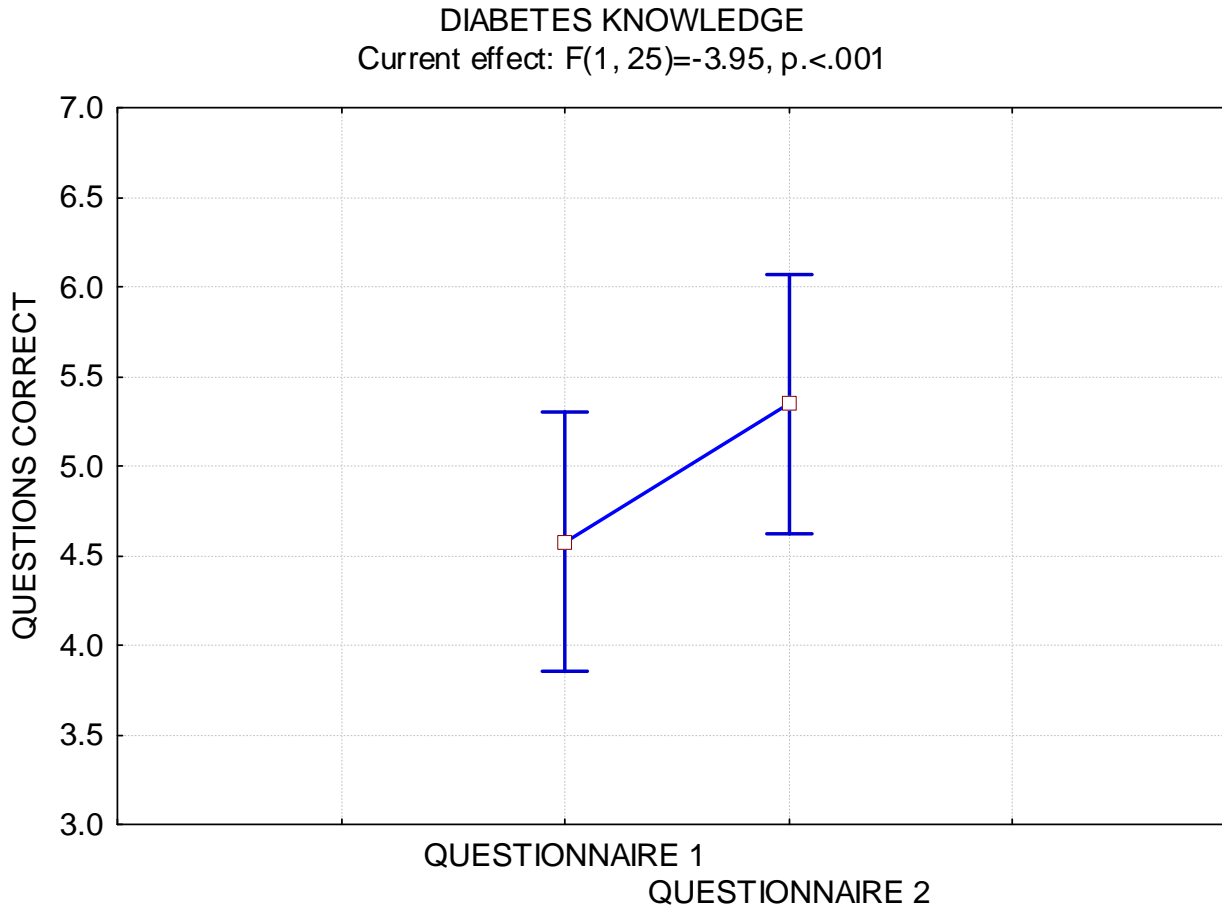
Results

Satisfaction measured by Preferences for a Fixed or Adaptive Assistant



Results

Educational Value measured by increases in Diabetes Knowledge



Summary of Results

Adaptive computer assistant is indicated to be more effective and time efficient

Preference for adaptive assistant

- Strong preference with males
- Personal characteristics influence preference

Increased knowledge of diabetes

Recommendations for improvement of interface

Sustainable Self-Care

Blanson Henkemans, O.A., van der Boog, P.J.M., Lindenberg, J., van der Mast, C.A.P.G., Neerincx, M.A., and Zwetsloot-Schonk, B.J.H.M. (forthcoming). An Online Lifestyle Diary with Persuasive Computer Assistant for Support of Self-Management. *Technology & Health Care*.

How can a Computer Assistant support sustainable use of an online lifestyle diary?

- Actual patients
- Over a longer period of time
- At home

Design

DieetInzicht

www.DieetInzicht.nl offers support for maintaining a healthy diet

Objective information in relation to nutrition

Developed in by LUMC internist Paul van der Boog and dieticians

Since 2006, approximately 18,000 users

Design

Personal Lifestyle Goals

Diary Use

- Use diary at least 4 time a week

Diet

- Eat the right amount of fat per day

Physical Activities

- Light exercise, 30 minutes per day



Diet Page

Goal Overview



Here you can see what goals you have achieved or not.

Press the goal button to see more details about the goal.

Goal Achieved

Diary Goal Yes

Diet Goal No

Activity Goal Yes

Filling in Diet Diary - Monday August 25, 2008

Select Date

Diet Misc

Breakfast

- 2 snee Brood- tarwe (70 gr.)
- 1 beleg (dik) Boter halfvol (8 gr.)
- 1 broodbeleg Jam (15 gr.)
- 1 grote beker Thee zonder melk, z suiker (225 gr.)

Morning snack

- 1 automaat Koffie bereid (130 gr.)
- 1 middel Banaan (130 gr.)

Lunch

- 4 snee Brood- tarwe (140 gr.)
- 1 klontje Boter halfvol (13 gr.)
- 1 middel Appel m schil (150 gr.)
- 1 broodbeleg Chocoladehagelslag (15 gr.)
- 1 stuks Kippenei gekookt (50 gr.)
- 1 stuks Hollandse nieuwe haring (75 gr.)
- 1 soepkop Groentesoep helder (250 gr.)

Products Copy/Delete

Add Products

Moment

Category

Search product

Description

Quantity

Description	Quantity
Kippenei gekookt	<input type="text"/> stuks (50 gram)
Drumsticks bereid mz	<input type="text"/> grote (80 gram)
Drumsticks bereid zz	<input type="text"/> grote (80 gram)
Drumsticks rauw	<input type="text"/> grote (100 gram)
Ei gebakken mz	<input type="text"/> stuks (50 gram)
Ei gebakken zz	<input type="text"/> stuks (50 gram)

- Lifestyle Diary
- Home
- Personal Page
- Diet Page**
- Activity Page
- Report Page
- Products page
- Manual
- Study Regulations
- Contact
- Ingelogd als: blanson
- Log Out

- [Help](#)
- [Activity Page](#)
- [Report Page](#)



Design

Persuasive Computer Assistant

Animated iCat

Cooperative Feedback

Motivational Interviewing (Miller & Rollnick, 1991; Looije et al., 2009)

- Expressing empathy
- Being cheering and complimenting
- Exploring differences between goals and diary entries
- Supporting self-efficacy and optimism

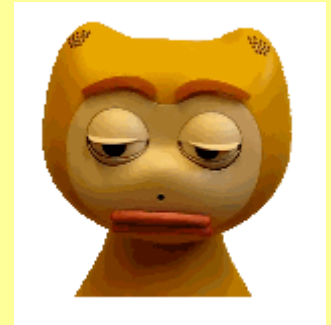
Design

Persuasive Computer Assistant Example

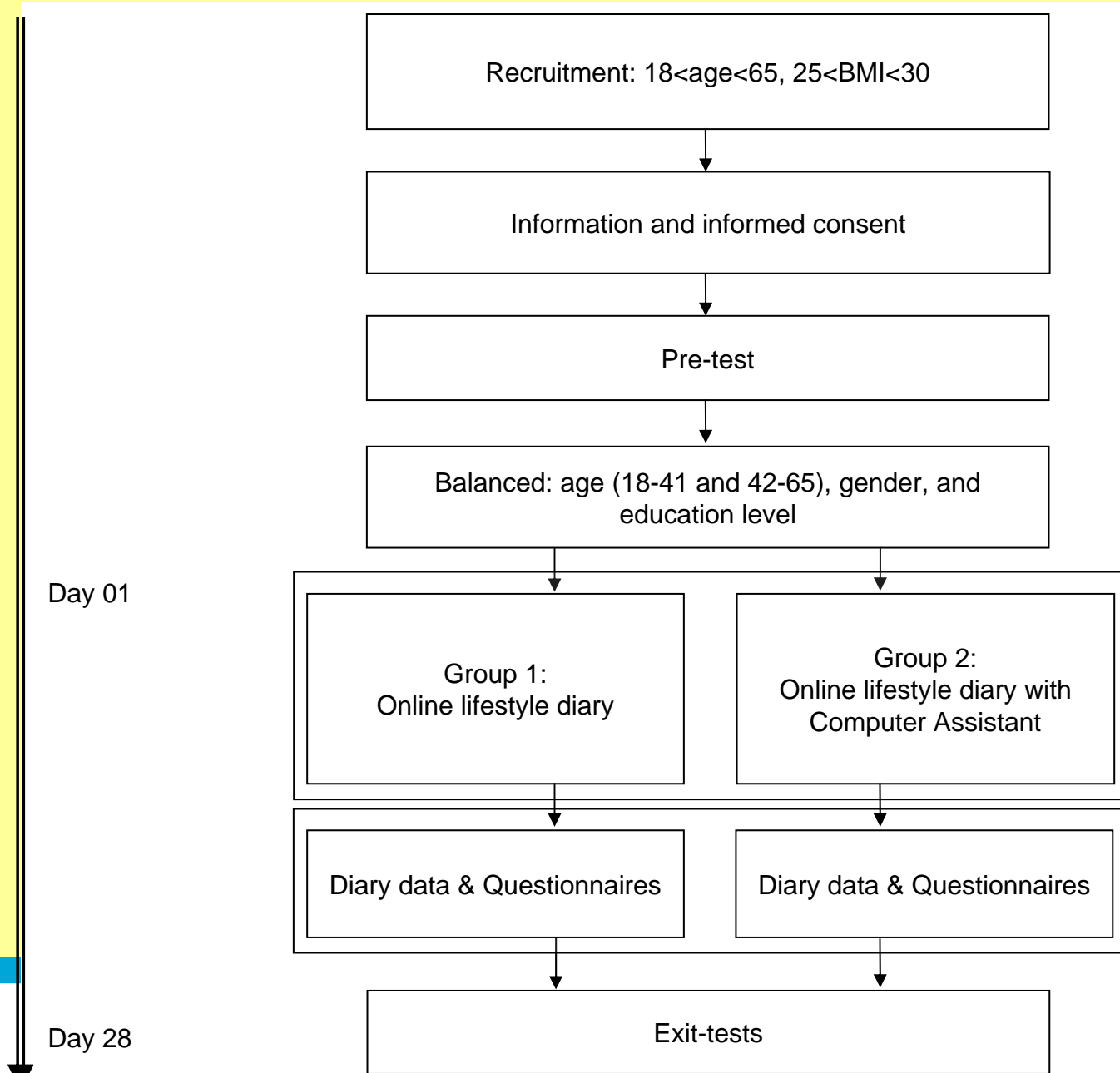
*You achieved you goal for today.
Congratulations! Try to persist the coming
days.*



*You did not succeed in achieving your goal for
today. Too bad, but don't let that discourage
you. It goes step by step. Try again the
coming days. For suggestions on how to
achieve your goal, read some literature [here](#).*



Method Randomized Controlled Trial



Method

Participants

	Recruited	Maintained diary >5 days (male/female)	Completed exit test (male/female)
Computer assistant	97	65 (15/50)	18 (3/15)
No computer assistant	94	53 (6/47)	17 (3/14)
Total	191	118 (21/97)	35 (6/29)

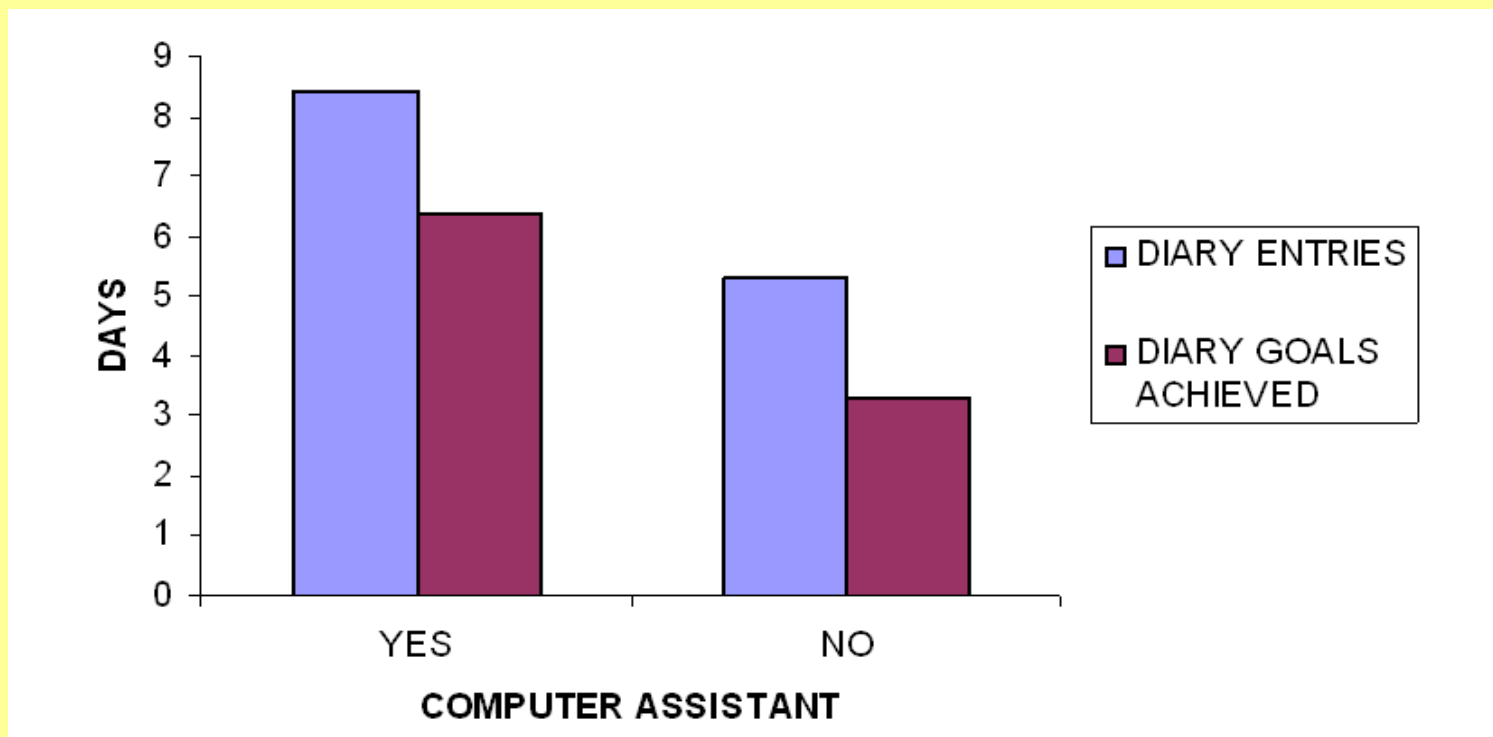
Method

Dependent Variables

	Pre-study (Day 01)	Study (Day 01–28)	Exit survey (Day 28)
Demographics	118		
Computer Experience	118		
Locus of Control Scale	118		
Vocabulary	118		
Lifestyle knowledge	118		35
Motivation	118	50 (on Day 14)	35
Body Mass Index (BMI)	118		35
Diet and physical activities		118	
Online diary evaluation			35

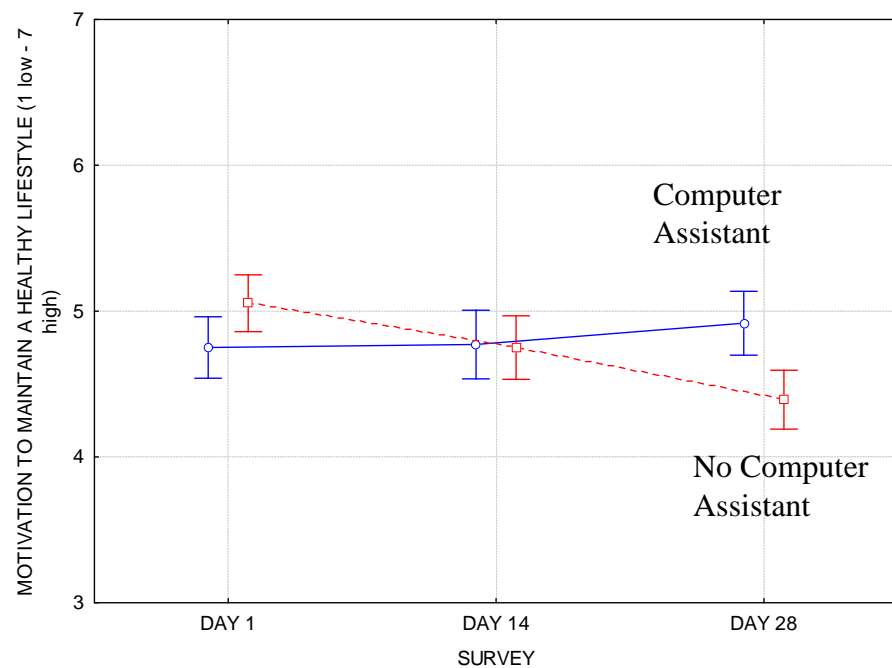
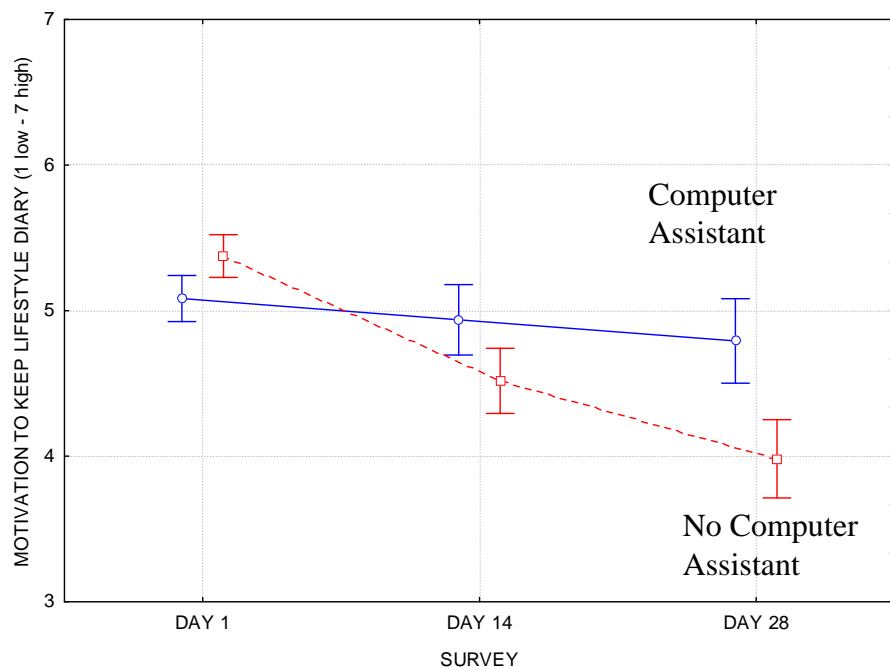
Results

Diary Adherence & Goal achievement ($N=118$)

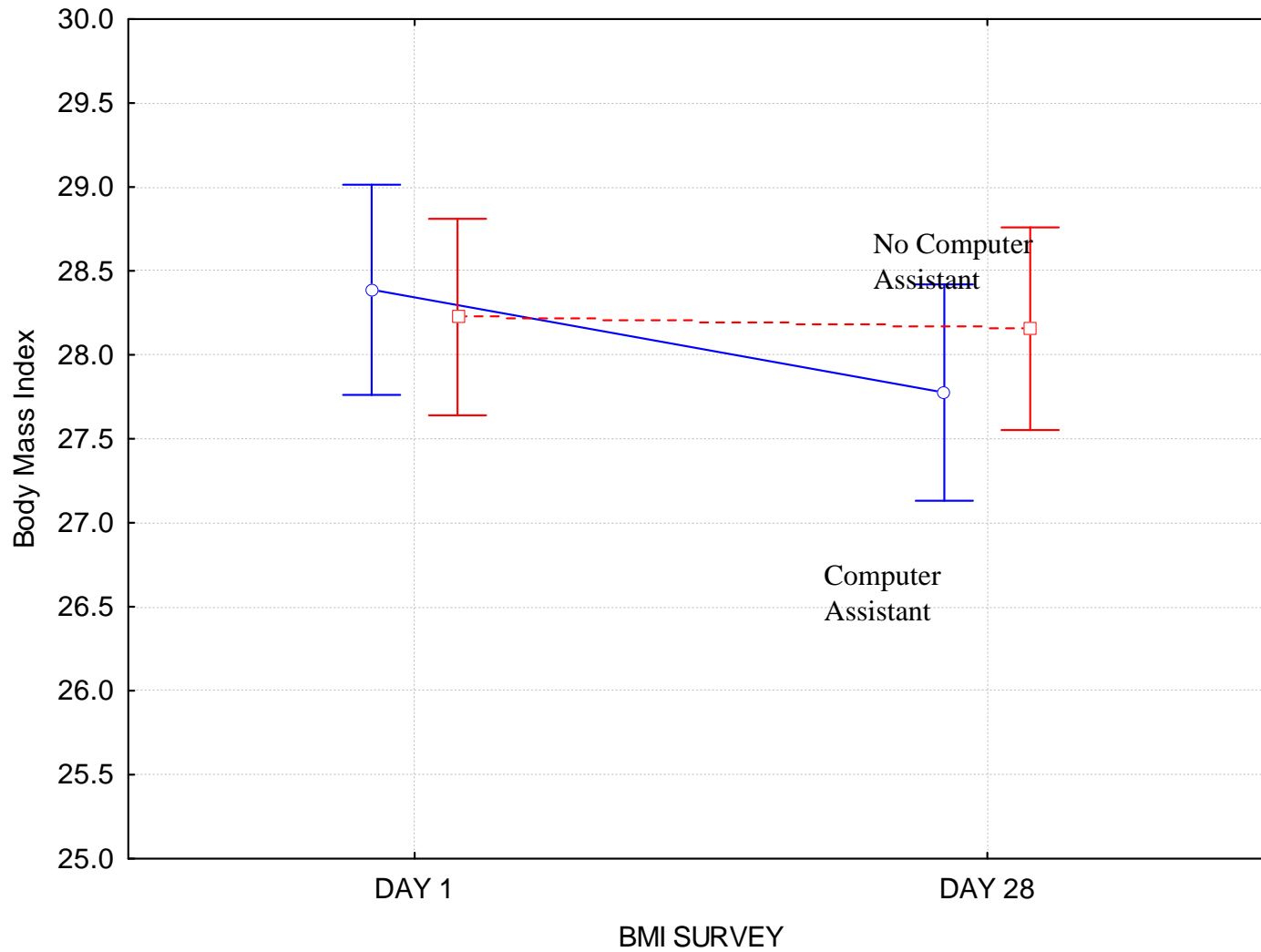


Results

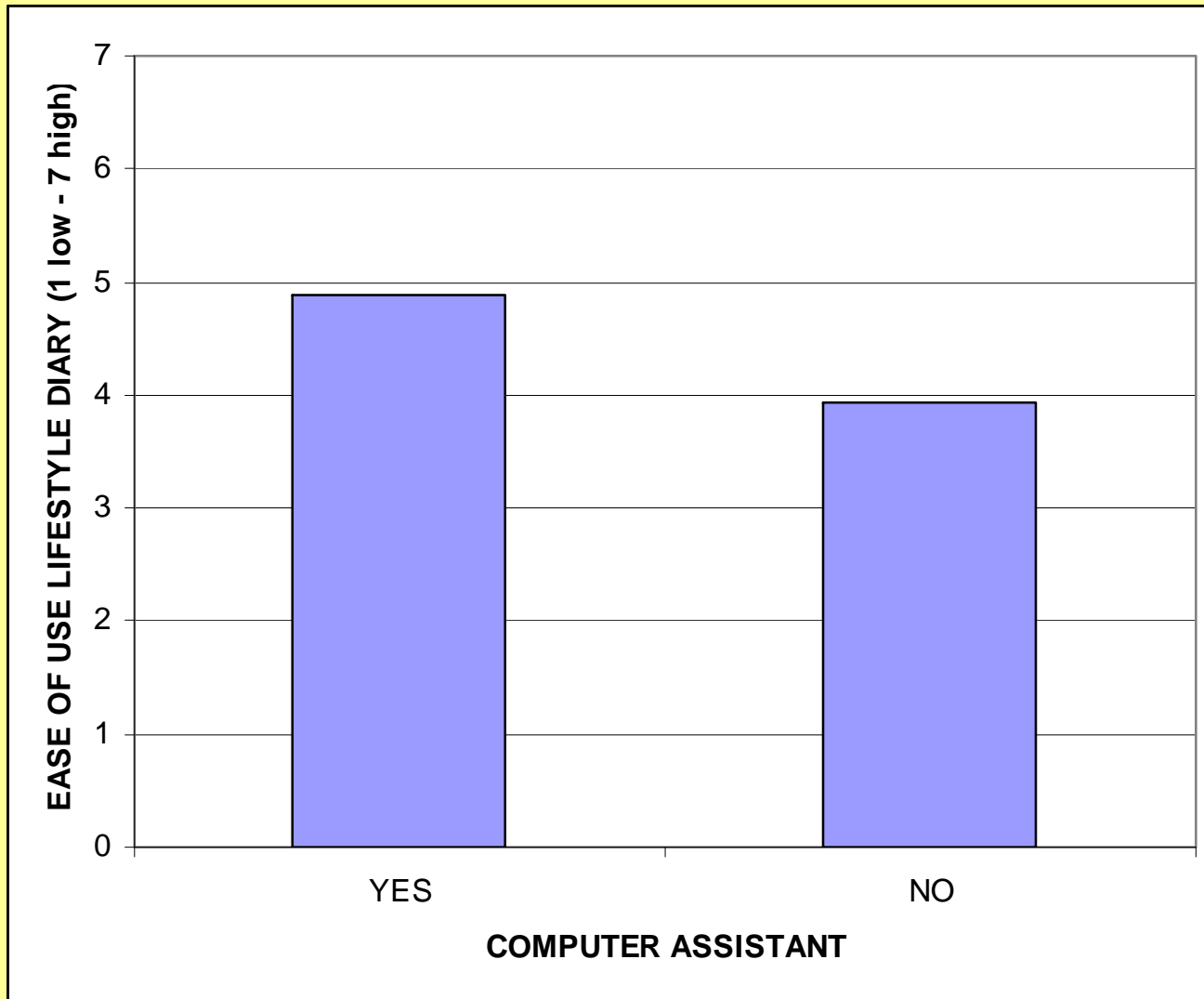
Motivation ($N=35$)



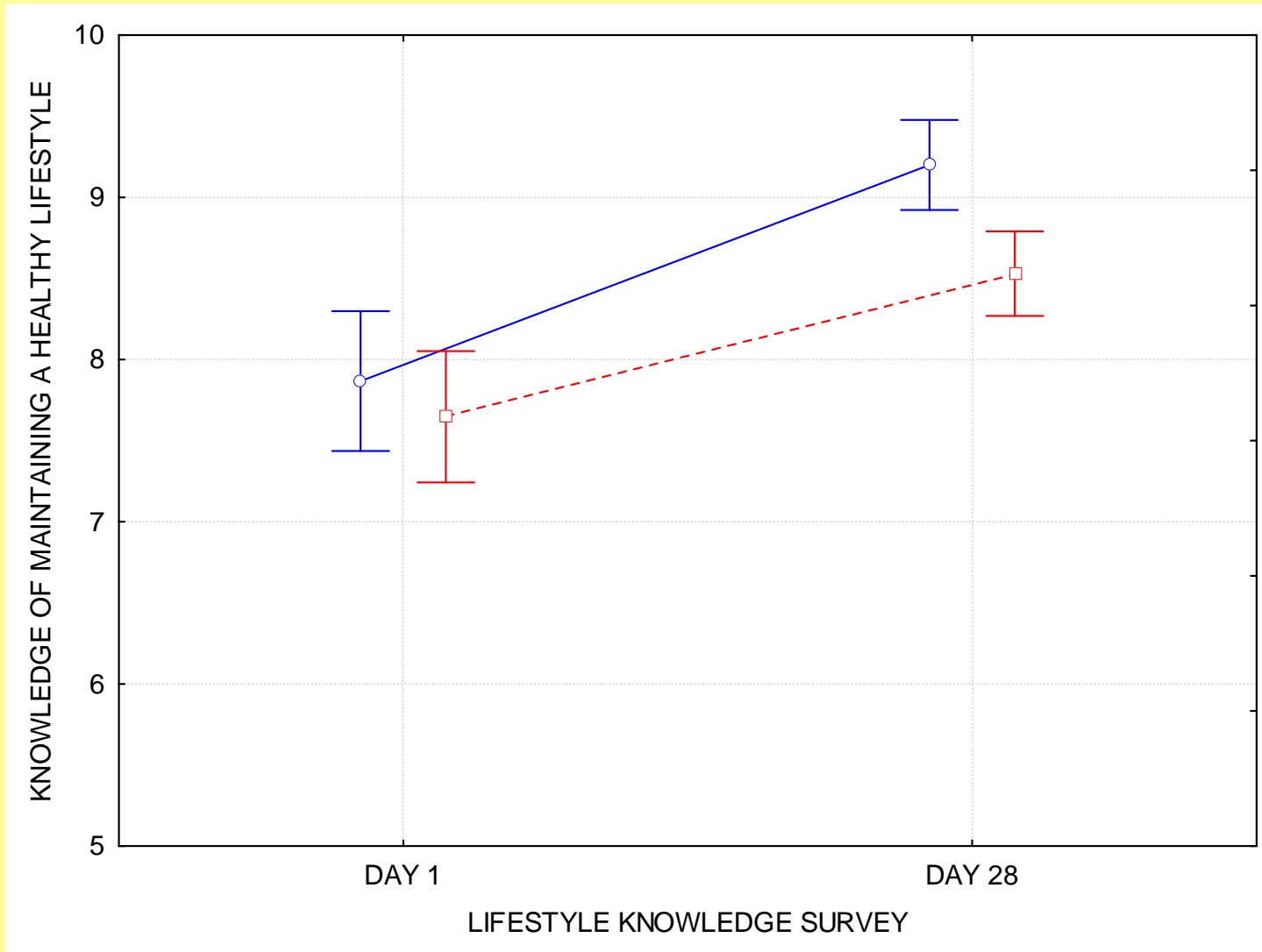
Results BMI ($N=35$)



Results usability ($N=35$)



Results Health Literacy (N=35)



Results

Personal Characteristics

Participants with a high internal locus of control and who scored high on vocabulary and computer experience, entered the diary more accurately

Younger participants, who scored higher on computer experience and education level, were more motivated to maintain a healthy lifestyle

Discussion

Persuasive Computer Assistant for adherence to lifestyle diary

- Animated iCat
- Cooperative feedback
- Motivational Interviewing

Improved health outcomes with motivated people

- BMI
- Motivation

PCA can contribute to maintaining a healthy lifestyle

Acknowledgments

SuperAssist

- Partly financed by SenterNovem, IOP-MMI

Research team

- Olivier Blanson-Henkemans (Olivier.BlansonHenkemans@TNO.nl)
- Mark Neerincx
- Jasper Lindenberg
- Rosemarijn Looije
- Charles van de Mast
- Bertie Zwetsloot-Schonk
- Paul van der Boog
- ...

So, good evaluation is?

Address user experience in its breadth

- Effectiveness, efficiency, satisfaction, learnability, ...

Take account of

- Context and user diversity

Apply different methods during the development process

The evaluation should provide insight in problems *and* causes in order to support planning changes to correct the problem

Literature

Current Lecture (module 6):

- Maguire, M. (2001). Methods to support human-centred design. *International journal of human-computer studies*, 55(4), 587 - 634.

[Next Lecture (module 7):

- Prototype and test plan presentation of each group]

Module 8 (Quarter 4):

- Smets, N.J.J.M., Abbing, M.S., Neerincx, M.A., Lindenberg, J, and van Oostendorp, H. (2008). Game-based evaluation of personalized support for astronauts in long duration missions. *Proceedings of the 59th International Astronautical Congress (IAC2008)*. Paris, France: IAF. DVD: ISSN 1995-6258.