

Pengenalan Evaluasi



Dua tipe dasar evaluasi

- **Evaluasi formatif** dilakukan pada tahap pengembangan yang berbeda-beda untuk mengecek produk telah memenuhi kebutuhan pengguna
- **Evaluasi sumatif** menilai kualitas produk yang sudah selesai

Fokus kita adalah pada evaluasi formatif

Apa yang dievaluasi

Perancangan dan evaluasi berulang adalah proses yang terus menerus untuk menguji:

- Ide awal model konseptual
- Prototipe awal sistem baru
- Kemudian, prototipe yang lebih lengkap

Perancang perlu mengecek mereka telah mengetahui kebutuhan pengguna

Mengapa perlu mengevaluasi?

“Perancangan berulang, dengan siklus perancangan dan pengujian yang diulang-ulang, merupakan metodologi yang paling valid agar sukses dalam menghasilkan produk secara konsisten. Jika Anda tidak menyertakan pengujian *user* sebagai bagian yang terintegrasi dengan proses perancangan Anda, itu seperti Anda membuang seember uang ke selokan.”

(Bruce Tognazzini – AskTog.com)

Kapan mengevaluasi?

- Dalam keseluruhan perancangan
- Mulai dari deskripsi awal, sketsa kebutuhan *user*, dsb sampai dengan produk akhir
- Proses perancangan yang meliputi siklus berulang dari 'perancangan-pengujian-perancangan ulang'
- Evaluasi adalah bahan utama untuk perancangan yang sukses

Paradigma Evaluasi

Setiap evaluasi dipandu secara langsung ataupun tidak langsung oleh sekumpulan keyakinan, yang kadang didukung oleh teori. Keyakinan dan metode yang berhubungan dengan keyakinan tersebut disebut sebagai **'paradigma evaluasi'**

Empat Paradigma Evaluasi

- 'quick and dirty'
- pengujian *usability*
- studi lapangan
- evaluasi prediktif

Quick and dirty

- **evaluasi 'quick & dirty'** menggambarkan kebiasaan yang dilakukan oleh perancang untuk mendapat umpan balik secara informal dari *user* atau konsultan untuk mengkonfirmasi gagasan mereka sesuai dengan kebutuhan *user* dan disukai
- evaluasi *quick & dirty* dilakukan kapanpun
- penekanannya adalah pada masukan untuk proses perancangan yang lebih cepat, dan tidak pada temuan yang didokumentasi secara hati-hati

Pengujian Usability

- Pengujian *usability* melibatkan perekaman performansi *user* tertentu dalam melakukan pekerjaan tertentu pada situasi yang terkendali. Observasi lapangan juga dapat digunakan.
- Saat *user* melakukan pekerjaan, mereka diamati dan direkam dengan menggunakan video, dan penekanan tombol dicatat
- Data ini digunakan untuk menghitung waktu performansi, mengidentifikasi kesalahan, dan membantu menjelaskan mengapa *user* melakukan apa yang mereka lakukan
- Kuesioner dan wawancara *user satisfaction* digunakan untuk memperoleh opini *user*

Studi Lapangan

- Studi lapangan dilakukan di suasana yang alami
- Tujuannya adalah untuk memahami apa yang dilakukan oleh *user* secara alami, dan bagaimana teknologi berdampak terhadap mereka
- Dalam perancangan produk, studi lapangan dapat digunakan untuk:
 - mengidentifikasi peluang teknologi baru
 - menentukan kebutuhan perancangan
 - memutuskan cara terbaik untuk memperkenalkan teknologi baru
 - mengevaluasi teknologi yang sedang digunakan

Evaluasi Prediktif

- Para pakar mengaplikasikan pengetahuan mereka pada *user* tertentu, terkadang dipandu dengan aturan heuristik untuk memprediksi permasalahan *usability*
- Pendekatan yang lain melibatkan model berbasis teori
- Fitur utama evaluasi prediktif adalah *user* tidak perlu ada
- Relatif cepat dan murah

DECIDE: kerangka kerja untuk memandu evaluasi

- Determine the *goals* the evaluation addresses.
- Explore the specific *questions* to be answered.
- Choose the *evaluation paradigm* and *techniques* to answer the questions.
- Identify the *practical issues*.
- Decide how to deal with the *ethical issues*.
- Evaluate, interpret and present the *data*.

Determine the goals (Tentukan Sasaran)

- Apa sasaran utama evaluasi?
- Siapa yang menginginkannya dan mengapa?
- Penentuan sasaran menentukan paradigma evaluasi
- Contoh sasaran:
 - Identifikasi metafor terbaik untuk mendasari perancangan
 - Cek untuk menjamin antarmuka akhir konsisten
 - Selidiki bagaimana teknologi berdampak pada pekerjaan
 - Kembangkan *usability* produk yang sudah ada

Explore the questions (Eksplorasi Pertanyaan)

- Semua evaluasi membutuhkan sasaran dan pertanyaan untuk memandu mereka, sehingga waktu tidak terbuang percuma pada studi yang tidak terdefinisi dengan baik
- Contoh: sasaran evaluasi yaitu menyelidiki mengapa banyak pelanggan lebih suka membeli tiket pesawat (kertas) daripada tiket elektronik, dapat diturunkan menjadi beberapa sub-pertanyaan:
 - Bagaimana kebiasaan pelanggan terhadap tiket baru tersebut?
 - Apakah mereka peduli terhadap keamanan?
 - Apakah antarmuka tiket elektronik tidak menarik?
- Kasus:
 - Pertanyaan apa yang dapat Anda tanyakan untuk perancangan telepon selular?

Choose the evaluation paradigm & techniques

(Pilih Paradigma & Teknik Evaluasi)

- Paradigma evaluasi berpengaruh kuat pada teknik yang digunakan, bagaimana data dianalisa dan dipresentasikan
- Contoh: studi lapangan tidak melibatkan pengujian atau pemodelan

Identify practical issues (Identifikasi isu-isu praktis)

Contoh, bagaimana:

- memilih *user*
- menaati anggaran
- menaat jadwal
- mencari evaluator
- memilih peralatan

Decide on ethical issues (Tentukan isu-isu etis)

- Buat form persetujuan
- Partisipan mempunyai hak untuk:
 - mengetahui sasaran evaluasi
 - apa yang akan dilakukan terhadap hasil evaluasi
 - mendapat perlindungan (privasi) atas informasi personal
 - tidak dikutip tanpa persetujuan
 - meninggalkan tempat kapan pun mereka suka
 - diperlakukan secara sopan

Evaluate, interpret & present data (Evaluasi, interpretasi & presentasi data)

- Bagaimana data dianalisa dan dipresentasikan tergantung pada paradigma dan teknik yang digunakan
- Hal-hal berikut juga harus dipertimbangkan:
 - Reliability: dapatkah evaluasi direplikasi?
 - Validitas: apakah terukur seperti yang Anda pikirkan?
 - Bias: apakah proses menghasilkan bias?
 - Lingkup: dapatkah temuan digeneralisasi?
 - Validitas ekologi: apakah lingkungan mempengaruhi evaluasi? contoh: efek Hawthorn

Interviews

- Unstructured - are not directed by a script. Rich but not replicable.
- Structured - are tightly scripted, often like a questionnaire. Replicable but may lack richness.
- Semi-structured - guided by a script but interesting issues can be explored in more depth. Can provide a good balance between richness and replicability.

Basics of interviewing

- Remember the DECIDE framework
- Goals and questions guide all interviews
- Two types of questions:
 - 'closed questions' have a predetermined answer format, e.g., 'yes' or 'no'
 - 'open questions' do not have a predetermined format
- Closed questions are quicker and easier to analyze

Things to avoid when preparing interview questions

- Long questions
- Compound sentences - split into two
- Jargon & language that the interviewee may not understand
- Leading questions that make assumptions e.g., why do you like ...?
- Unconscious biases e.g., gender stereotypes

Components of an interview

- *Introduction* - introduce yourself, explain the goals of the interview, reassure about the ethical issues, ask to record, present an informed consent form.
- *Warm-up* - make first questions easy & non-threatening.
- *Main body* - present questions in a logical order
- *A cool-off period* - include a few easy questions to defuse tension at the end
- *Closure* - thank interviewee, signal the end, e.g, switch recorder off.

The interview process

- Use the DECIDE framework for guidance
- Dress in a similar way to participants
- Check recording equipment in advance
- Devise a system for coding names of participants to preserve confidentiality.
- Be pleasant
- Ask participants to complete an informed consent form

Group interviews

- Also known as 'focus groups'
- Typically 3-10 participants
- Provide a diverse range of opinions
- Need to be managed to:
 - ensure everyone contributes
 - discussion isn't dominated by one person
 - the agenda of topics is covered

Analyzing interview data

- Depends on the type of interview
- Structured interviews can be analyzed like questionnaires
- Unstructured interviews generate data like that from participant observation
- It is best to analyze unstructured interviews as soon as possible to identify topics and themes from the data

Questionnaires

- Questions can be closed or open
- Closed questions are easiest to analyze, and may be done by computer
- Can be administered to large populations
- Paper, email & the web used for dissemination
- Advantage of electronic questionnaires is that data goes into a data base & is easy to analyze
- Sampling can be a problem when the size of a population is unknown as is common online

Questionnaire style

- Varies according to goal so use the DECIDE framework for guidance
- Questionnaire format can include:
 - 'yes', 'no' checkboxes
 - checkboxes that offer many options
 - Likert rating scales
 - semantic scales
 - open-ended responses
- Likert scales have a range of points
- 3, 5, 7 & 9 point scales are common
- Debate about which is best

Developing a questionnaire

- Provide a clear statement of purpose & guarantee participants anonymity
- Plan questions - if developing a web-based questionnaire, design off-line first
- Decide on whether phrases will all be positive, all negative or mixed
- Pilot test questions - are they clear, is there sufficient space for responses
- Decide how data will be analyzed & consult a statistician if necessary

Encouraging a good response

- Make sure purpose of study is clear
- Promise anonymity
- Ensure questionnaire is well designed
- Offer a short version for those who do not have time to complete a long questionnaire
- If mailed, include a s.a.e.
- Follow-up with emails, phone calls, letters
- Provide an incentive
- 40% response rate is high, 20% is often acceptable

Advantages of online questionnaires

- Responses are usually received quickly
- No copying and postage costs
- Data can be collected in database for analysis
- Time required for data analysis is reduced
- Errors can be corrected easily
- Disadvantage - sampling problematic if population size unknown
- Disadvantage - preventing individuals from responding more than once

Problems with online questionnaires

- Sampling is problematic if population size is unknown
- Preventing individuals from responding more than once
- Individuals have also been known to change questions in email questionnaires

Questionnaire data analysis & presentation

- Present results clearly - tables may help
- Simple statistics can say a lot, e.g., mean, median, mode, standard deviation
- Percentages are useful but give population size
- Bar graphs show categorical data well
- More advanced statistics can be used if needed

Case study1: Designing mobile communicators

Two examples, for very different audiences:

- Nokia's mobile communicator
- Philips communicator for children

Designing Nokia's mobile communicator

- design cycle:
iterative user-centered approach
- which methods:
ethnographic research
scenarios and task models
- confidential product issues:
first in the market is key
evaluation must be very limited and no real users

Designing Nokia's mobile communicator (contd)

- physical aspects:
 - screen size
 - number of buttons versus functionality
- consistency issues
 - internal consistency (within mobile software)
 - external consistency (with desktop software)
- user testing
 - none before release
 - summative testing & questionnaires after

Designing Philips' communicator for children

- design cycle:
iterative and evolutionary
- which methods:
low-fidelity prototyping
participatory design
interface metaphors
- physical aspects:
color, shape, size, robustness
pen input
bags to protect screen

Designing Philips' communicator for children

- user involvement:
children involved throughout
prototypes evaluated constantly
invaluable insights for the designers
- lessons learned:
agree on assumptions in requirements
think of follow-on projects early on
users are not designers
act quick and dirty if necessary

Why was using different methods valuable?

- The evaluators were able to build-up a broad picture of usability problems.
- Using GOMS and heuristic evaluation they could explore the potential benefits of the redesigned system.
- User testing enabled them to confirm that the redesigned system offered better usability.
- User satisfaction questionnaires confirmed that users preferred the redesigned system.