



# KIỂM TRA VÀ ĐÁNH GIÁ

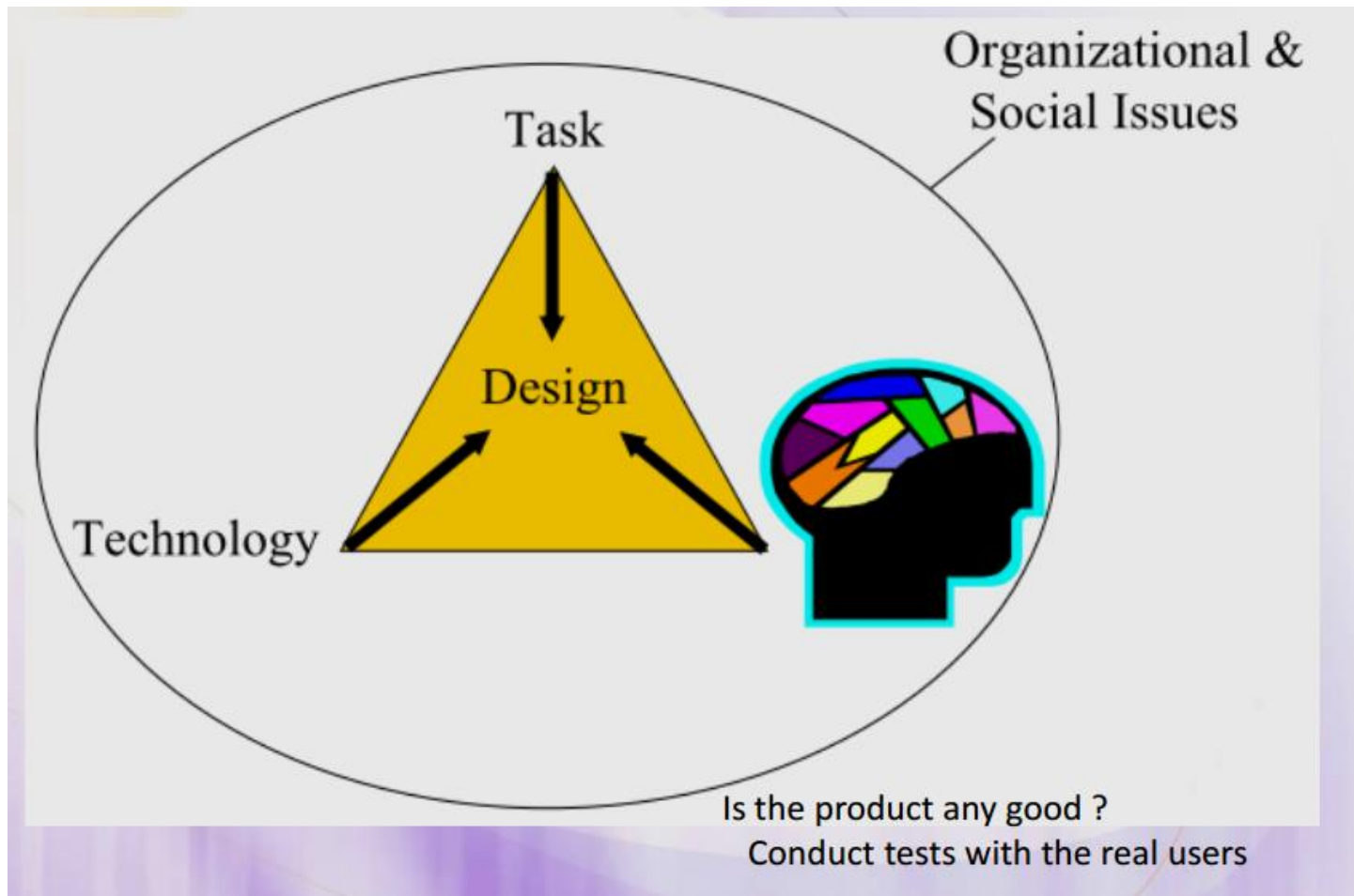




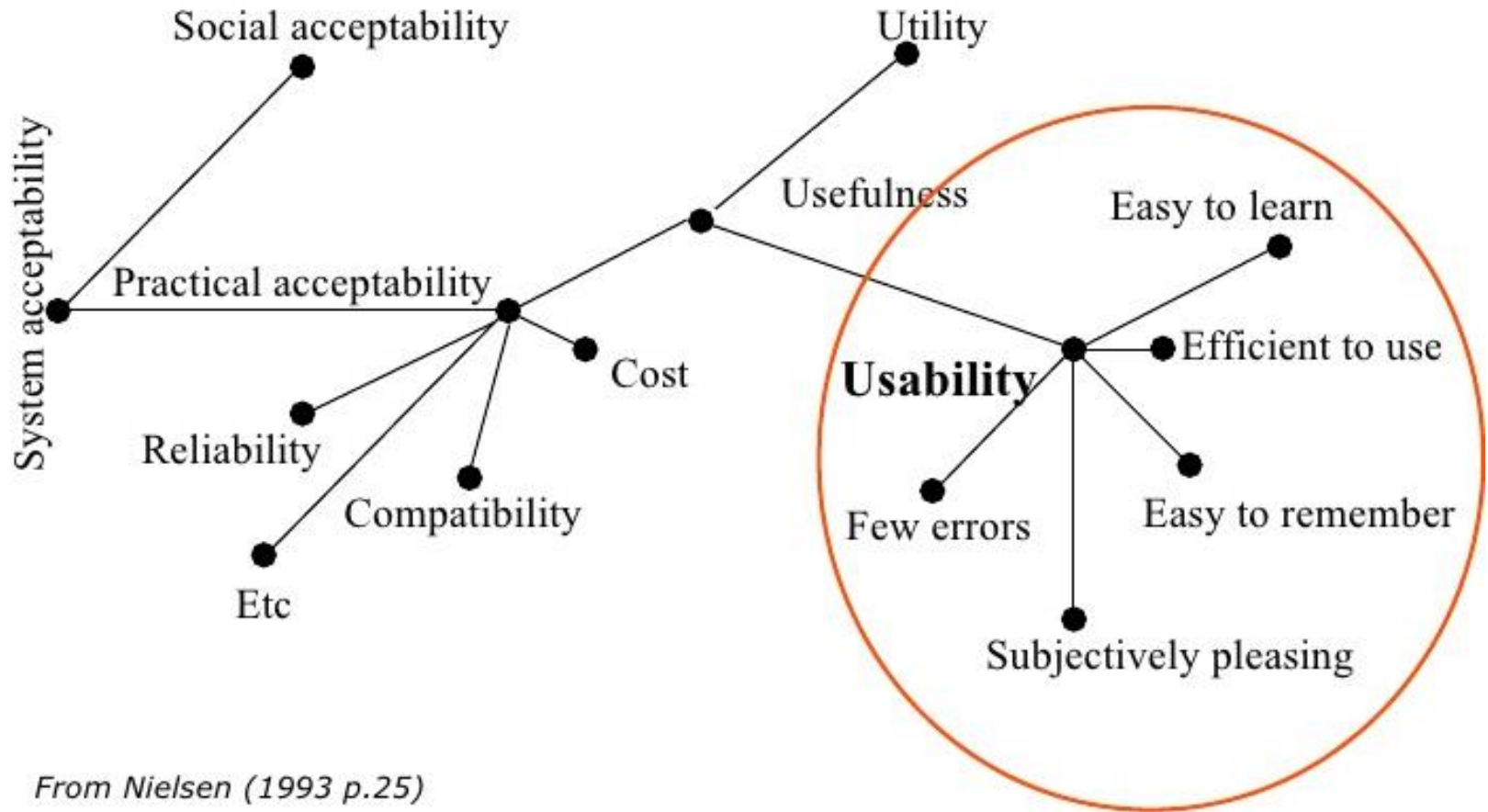
# Nội dung

- Các khái niệm cơ bản
- Các mô thức đánh giá
- Các kỹ thuật đánh giá
- Ví dụ và bài tập

# 1. Các khái niệm cơ bản



# Khả năng hệ thống được chấp nhận



From Nielsen (1993 p.25)




# Đánh giá về tính dùng được, kiểm tra về chức năng

- Kiểm tra: kiểm tra chức năng của hệ thống xác định & sửa lỗi, v.v.
  - là quan trọng
- Đánh giá: kiểm tra khả năng sử dụng của hệ thống có thể đạt được mục tiêu về:
  - Hiệu quả
  - Hiệu quả
  - Năng suất
  - Sự an toàn
  - Sự hài lòng của người dùng



# Kế hoạch kiểm thử

- Mục đích kiểm tra
  - Tuyên bố vấn đề / mục tiêu kiểm tra
  - Hồ sơ người tham gia (tiêu chí bao gồm / loại trừ)
  - Phương pháp / kỹ thuật được sử dụng
  - Danh sách các nhiệm vụ được sử dụng
  - Môi trường thử nghiệm (hiện trường so với phòng thí nghiệm) và vật liệu (HW / SW, tài nguyên, máy ghi âm, biểu mẫu báo cáo, bảng câu hỏi)
  - Vai trò của người thử nghiệm (giám sát, huấn luyện viên, v.v.)
  - Các biện pháp đánh giá cần thực hiện (định tính so với định lượng, chủ quan hay khách quan)
  - Nội dung của báo cáo và cách thức báo cáo -> tập trung nhóm, họp không chính thức
- 

# Các mô thức đánh giá

**Nhanh (1):** các cuộc thảo luận không chính thức với người dùng bất cứ lúc nào có thể sử dụng mẫu thử. Nhận phản hồi từ người dùng hoặc nhà tư vấn để xác nhận rằng ý tưởng của họ phù hợp với nhu cầu của người dùng và được yêu thích.

**Nghiên cứu thực địa (3):** đi đến trang web của người dùng và sử dụng khảo sát hoặc quan sát người dùng sử dụng giao diện người dùng. Người dùng tự nhiên làm gì và công nghệ tác động đến họ như thế nào

**Kiểm tra tính dùng được (2):** quan sát người dùng và ghi lại hiệu suất của các người dùng điển hình trên nhiệm vụ trong cài đặt được kiểm soát. Giải thích lý do tại sao người dùng làm những gì họ đã làm (tính toán thời gian thực hiện, xác định lỗi). Khơi gợi ý kiến của người dùng (phỏng vấn, bảng câu hỏi)

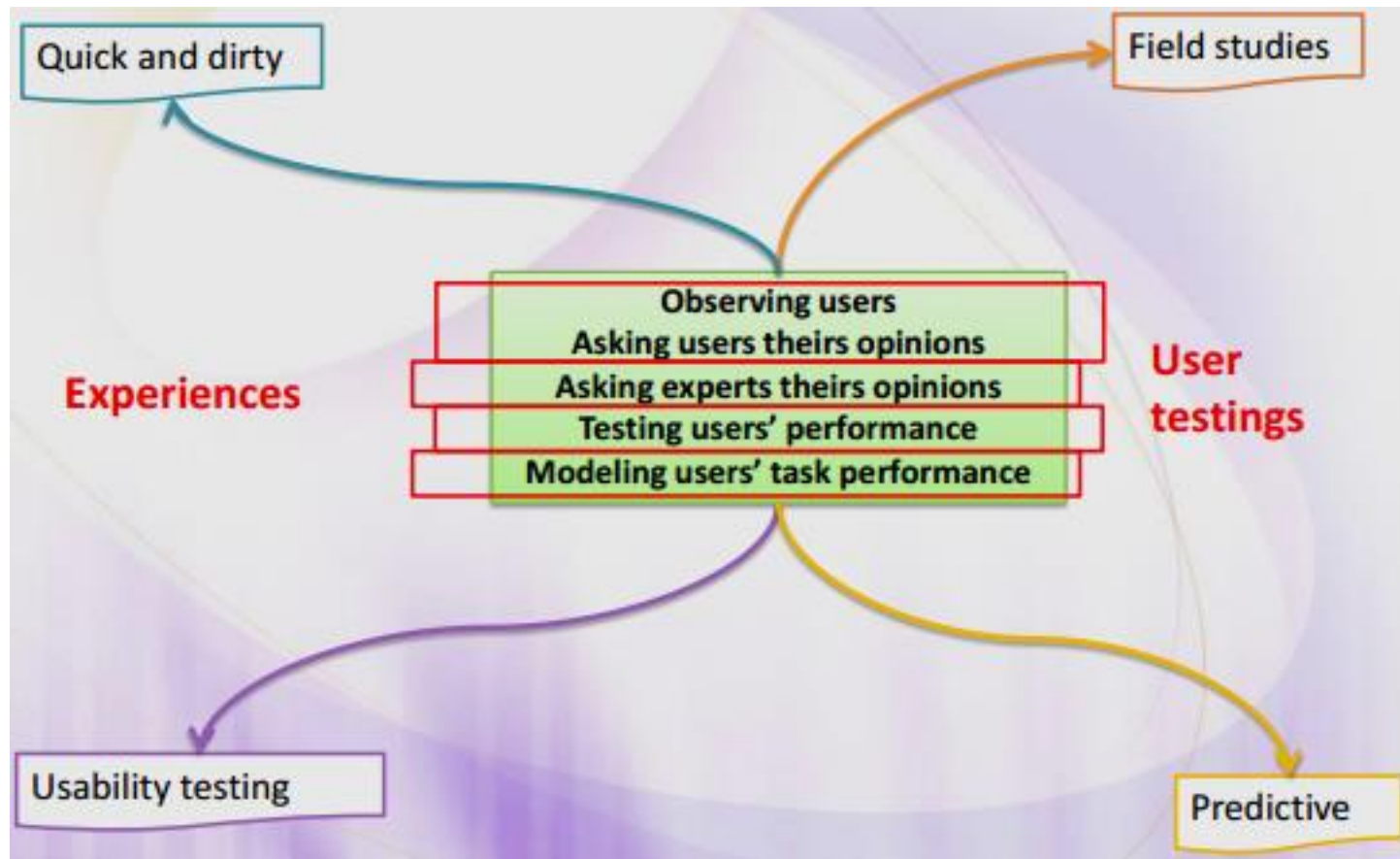
**Dự đoán (4):** người dùng không cần có mặt. Các chuyên gia sử dụng kiến thức của họ về người dùng điển hình (sử dụng đánh giá heuristic) để dự đoán khả năng sử dụng các vấn đề. Một cách tiếp cận khác liên quan đến các mô hình dựa trên lý thuyết.

# Khác nhau giữa các mô thức đánh giá

Mô thức	Quick and dirty (1)	Usability testing (2)	Field studies (3)	Predictive (4)
<i>Vai trò của người dung trong đánh giá</i>	Hoạt động tự nhiên	Để tiến hành các thao tác	Hoạt động tự nhiên	Không có
<i>Ai điều khiển việc đánh giá</i>	Người đánh giá có vai trò tối thiểu	Người đánh giá có vai trò lớn	Mối quan hệ giữa người đánh giá và người dùng	Chuyên gia Người đánh giá
<i>Nơi đánh giá</i>	Môi trường tự nhiên hoặc phòng thí nghiệm	Phòng thí nghiệm	Môi trường tự nhiên	Phòng thí nghiệm hoặc giả định
<i>Khi nào việc đánh giá được sử dụng</i>	Bất kỳ lúc nào	Với mẫu thử hoặc sản phẩm	Sớm	Với mẫu thử
<i>Loại dữ liệu được thu thập từ đánh giá</i>	định tính; không chính thức thảo luận	Định lượng; thống kê	Định tính, phác thảo	Danh sách các vấn đề
<i>Dữ liệu đưa trở lại thiết kế như thế nào</i>	Bản phác thảo và trích dẫn	Báo cáo về biểu diễn	Mô tả tại xưởng, các báo cáo và phác thảo	Báo cáo
<i>Triết học hoặc lý thuyết đánh giá</i>	Người dung làm trung tâm	Khoa học / thử nghiệm	Nhân chủng học	Kinh nghiệm thực tế, Học thuyết



# Các mô thức đánh giá





# Các chủ đề

- Trải nghiệm
- Kiểm tra của người dùng



# Phân loại

- Đánh giá đúng quy cách: trước khi thực hiện
  - giúp hình thành ý tưởng
  - một phần của thiết kế lắp/RAD
  - rất khó quản lý
- Đánh giá tổng kết: khi kết thúc quá trình phát triển
  - kết quả khó hiểu
- Đánh giá sau tổng kết: sau khi thực hiện
  - giám sát sự chấp nhận

# Các kĩ thuật đánh giá

## Đánh giá hình thành

- nhận thức
  - hướng dẫn
  - bút chì và giấy
  - mô phỏng
  - hợp tác xã
  - đánh giá
  - dựa trên kịch bản
  - Đánh giá
- Chi phí thấp, làm thường xuyên
  - Tìm các vấn đề về khả năng sử dụng SƠM

## Đánh giá tổng kết

- kĩ thuật dựa trên phòng thí nghiệm
  - ghi nhật ký và theo dõi
  - dân tộc học
- Chi phí cao, tập trung
  - Thể hiện khả năng sử dụng khi kết thúc

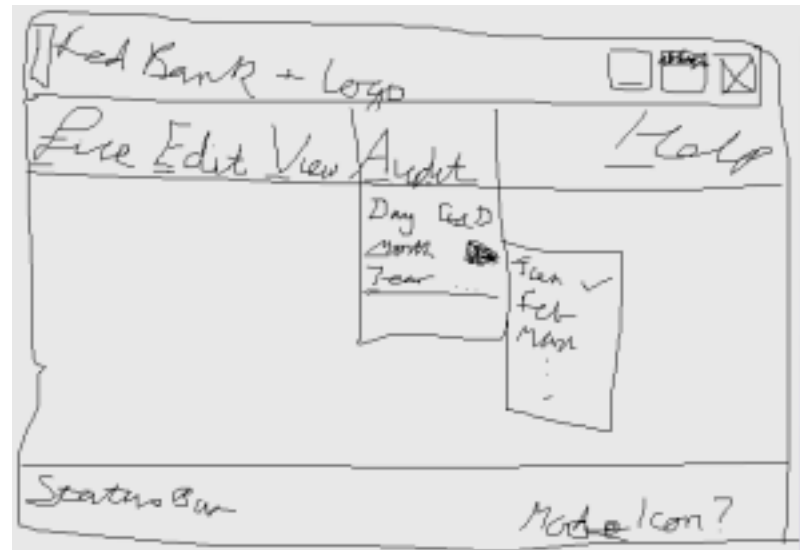
## Đánh giá sau tổng

**kết:** Nếu không phát hiện ra vấn đề:

- lớn và định kỳ
  - chi phí đào tạo cao;
  - số lượng lớn tài liệu;
  - số nhân viên lớn;
  - nhiều nhân viên hỗ trợ;
  - cuối cùng cũng cần thiết kể lại.
- Tầm quan trọng của thiết kế giao diện.

# Đánh giá hình thành: Bút chì và mô hình giấy

- Bắt đầu với các trang overhead projector:
  - chồng lên để hiển thị menu, v.v.
  - NHƯNG thành kiến của người dùng với độ tin cậy thấp?
- Chuyển lên độ tin cậy cao hơn:
  - NHƯNG hãy sẵn sàng vứt bỏ nó
  - toàn bộ là để kiểm tra giả định
- Thu hút người dùng tham gia sớm:
  - không tin tưởng ý kiến của nhà thiết kế ...



# Đánh giá hình thành: Hợp tác



- Thu hút người dùng xem mẫu thử:
  - KHÔNG hỏi những gì họ nghĩ.
  - Yêu cầu thực hiện một nhiệm vụ
- Yêu cầu người dùng nghĩ:
  - Nếu họ nói rằng họ đang bối rối, hãy ghi lại để thiết kế lại sau này.
- Làm việc này ở đâu?
  - Ở môi trường cuối?
  - hoặc ở văn phòng thiết kế?



# Đánh giá hình thành

- Hướng dẫn
  - Chuyên gia đầu tiên:
    - xác định bối cảnh của nhiệm vụ quan trọng
    - đặt bối cảnh của nhiệm vụ
    - giả định về người dùng
  - Thông qua 'nhiệm vụ':
    - dự đoán những cạm bẫy cho người dùng
    - sử dụng chúng trong thiết kế / thử nghiệm.
  - Hướng dẫn nhận thức:
    - phân loại các hoạt động tinh thần
    - Sớm mất mục tiêu
    - Loại bỏ mục tiêu lớn
- Dựa trên kịch bản
- Người dùng mô tả điều họ muốn:
  - dấu vết tương tác lý tưởng;
  - giống như một bản ghi từ một vở kịch.
- Nhà thiết kế loại bỏ điều này:
  - tạo thành các trường hợp thử nghiệm;
  - người dùng phải có thể...
- Tốt cho hợp đồng:
  - biết người dùng / hệ thống phải làm gì ...

# Đánh giá tổng hợp: Kỹ thuật dựa trên phòng thí nghiệm

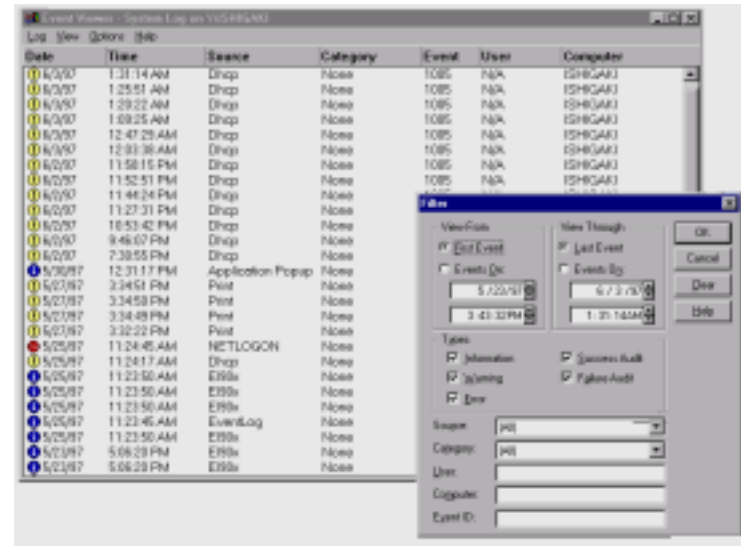
- Kiểm tra giả thuyết cụ thể;
- Ít lỗi hơn ở một nhiệm vụ cho trước;
- Nhanh hơn với dữ liệu giống nhau.
- Các vấn đề:
  - nhiệm vụ thể hiện như thế nào như thế nào?
  - người dùng thể hiện như thế nào?
  - bối cảnh thể hiện như thế nào?
- Tính dùng được trong thí nghiệm:
  - người dùng ra khỏi môi trường bình thường.
- So sánh:
  - người dùng bị gián đoạn





# Đánh giá tổng kết: Ghi nhật ký và theo dõi

- Giám sát một số người dùng với hệ thống trong môi trường làm việc
- Có thể ghi lại tốc độ, lỗi, v.v.
- Các vấn đề:
  - làm gì với khối lượng lớn dữ liệu?
  - cho người dùng biết về sự giám sát?
  - còn tài nguyên nào để khắc phục sự cố không?



- Điều này có hợp pháp hoặc có đạo đức?



# So sánh

- Kiểm tra thực nghiệm:
  - đánh giá các phương pháp đánh giá
- Như mong đợi:
  - các điểm mạnh / điểm yếu khác nhau;
  - dựa trên phòng thí nghiệm ít hiệu quả hơn về chi phí;
  - phạm vi bao phủ ít hơn.
- Ta làm gì không quan trọng nhưng
  - cần làm việc gì đó....




# Các chủ đề

- Trải nghiệm
- Kiểm tra của người dùng



# Tại sao phải kiểm tra người dùng?

- Không thể biết giao diện người dùng tốt như thế nào cho đến khi?
    - Con người sử dụng nó!
  - Các phương pháp khác dựa trên những người đánh giá
    - họ có thể biết quá nhiều
    - họ có thể không biết đủ (về nhiệm vụ, v.v.)
  - Khó dự đoán người dùng thực sẽ làm gì
- 



# Sẽ làm gì?

- Thời gian làm việc
  - Mọi người mất bao lâu để hoàn thành các công việc cơ bản?
  - Ví dụ: tìm thứ gì đó để mua, tạo tài khoản mới và đặt hàng.
- Sự chính xác
  - Mọi người đã mắc bao nhiêu sai lầm?
  - Và họ đã mất hay có thể phục hồi với thông tin chính xác?
- Gợi lại
  - Sau này người đó nhớ được bao nhiêu?
- Phản hồi có cảm xúc
  - Người đó cảm thấy thế nào về các nhiệm vụ đã hoàn thành?
    - Tự tin?
    - Căng thẳng?
    - Người dùng có giới thiệu hệ thống này cho bạn bè không?




# Đánh giá tính dùng được

- Thu hút người dùng trực tiếp trong các thử nghiệm trong phòng thí nghiệm về tính dùng được
- Giao thức Thinking Aloud
  - Người dùng được yêu cầu nói ra suy nghĩ, cảm xúc và ý kiến của họ khi sử dụng hệ thống.
  - Hai biến thể: phản hồi quan trọng & báo cáo định kỳ.
  - Hữu ích để hiểu mô hình tinh thần của người dùng, sự tương tác với hệ thống và thuật ngữ.
- Học và khám phá
  - Người dùng làm việc theo đôi để đạt được mục tiêu chung với hệ thống được thử nghiệm.
  - Người dùng được khuyến khích "think-aloud".
- Phương pháp huấn luyện
  - Những người tham gia đã hỏi một huấn luyện viên chuyên nghiệp.
  - Người tham gia-huấn luyện viên & người tham gia-máy tính tương tác được quan sát.




# Đánh giá tính dùng được

- Phương pháp đồ bóng
    - Một người dùng thành thạo ngồi cạnh người dùng và giải thích hành vi của người dùng cho người kiểm tra tính dùng được.
    - Thông thường là không thích hợp khi người dùng đưa ra bất kỳ phản hồi nào trong phiên kiểm tra
  - Phương pháp giảng dạy
    - Người dùng tương tác với hệ thống để đạt được một số kiến thức chuyên môn.
    - Một người dùng mới được đưa vào và người dùng “chuyên gia” được yêu cầu giải thích cho người dùng mới làm thế nào để thực hiện nhiệm vụ.
  - Đo lường hiệu suất
    - Nhận dữ liệu định lượng khi người tham gia thực hiện một nhiệm vụ.
    - Giảm thiểu sự tương tác giữa người tham gia và người thử nghiệm trong quá trình thử nghiệm vì nó có thể ảnh hưởng đến dữ liệu định lượng.
    - Thiết kế cẩn thận số lượng người tham gia để đảm bảo đủ sức mạnh thống kê.
- 



# Đánh giá tính dùng được

- Cách thức đặt câu hỏi
    - Tester đặt câu hỏi trực tiếp cho người tham gia.
    - Những người tham gia hỏi về kinh nghiệm trong quá khứ của họ liên quan đến hệ thống được thử nghiệm.
  - Kiểm tra từ xa
    - Tester được tách biệt về thời gian / không gian với người dùng.
    - Dữ liệu được lấy từ nhật ký / bản ghi / mạng.
  - Kiểm tra hồi cứu
    - Người thử nghiệm và những người tham gia cùng nhau xem lại phiên đã ghi, người thử nghiệm hỏi điều gì đang xảy ra.
    - Nên sử dụng cùng với các phương pháp khác.
- 





# Bản báo cáo kiểm tra khả năng sử dụng

- Trang bìa: đặt tên cho nhóm và chuyên gia về khả năng sử dụng
- Giới thiệu: Mô tả giao diện người dùng, mục tiêu thử nghiệm và mô tả ngắn gọn về các thử nghiệm (1 trang)
- Kế hoạch thử nghiệm: bản gốc bạn đã tạo để thử nghiệm (khoảng 1 trang cho mỗi tình huống)
- Kết quả: Đây phải là các âm mưu và biểu đồ có giải thích (những gì nó cần, ~ 2 trang)
- Kết luận: Các vấn đề về khả năng sử dụng và đề xuất để cải thiện giao diện người dùng (1 trang)