Week 12-1
Introduction to Evaluation

SFWRENG 4HC3/6HC3 Human Computer Interfaces

** Slides adapted from previous and current instructors of COMPSCI/SFWRENG 4HC3/6HC3*

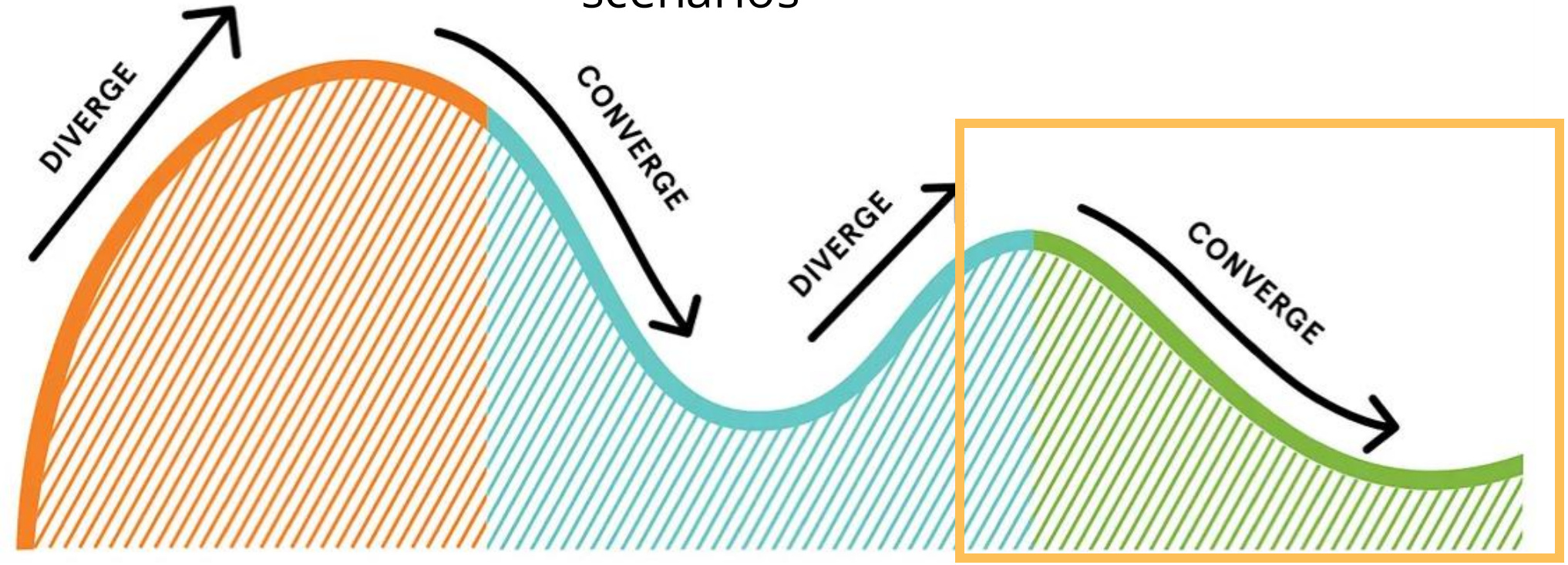
McMaster University

Human-Centered Design Process

Inspiration:
observation, interviews, survey

Ideation:
Participatory design, personas, scenarios

Implementations:
Evaluation



Week 12 Overview

- **Monday**
  - **Introduction to Evaluation**
  - **Evaluation: Performance Modeling**
- **Wednesday**
  - Evaluation: Inspection
- **Friday**
  - Evaluation: Heuristics
  - Evaluation with Users

Evaluation: Overview

> “Iterative design, with its repeating cycle of design and testing, is the only validated methodology in existence that will consistently produce successful results**. If you don’t have user-testing as an integral part of your design process you are going to throw buckets of money down the drain**.”

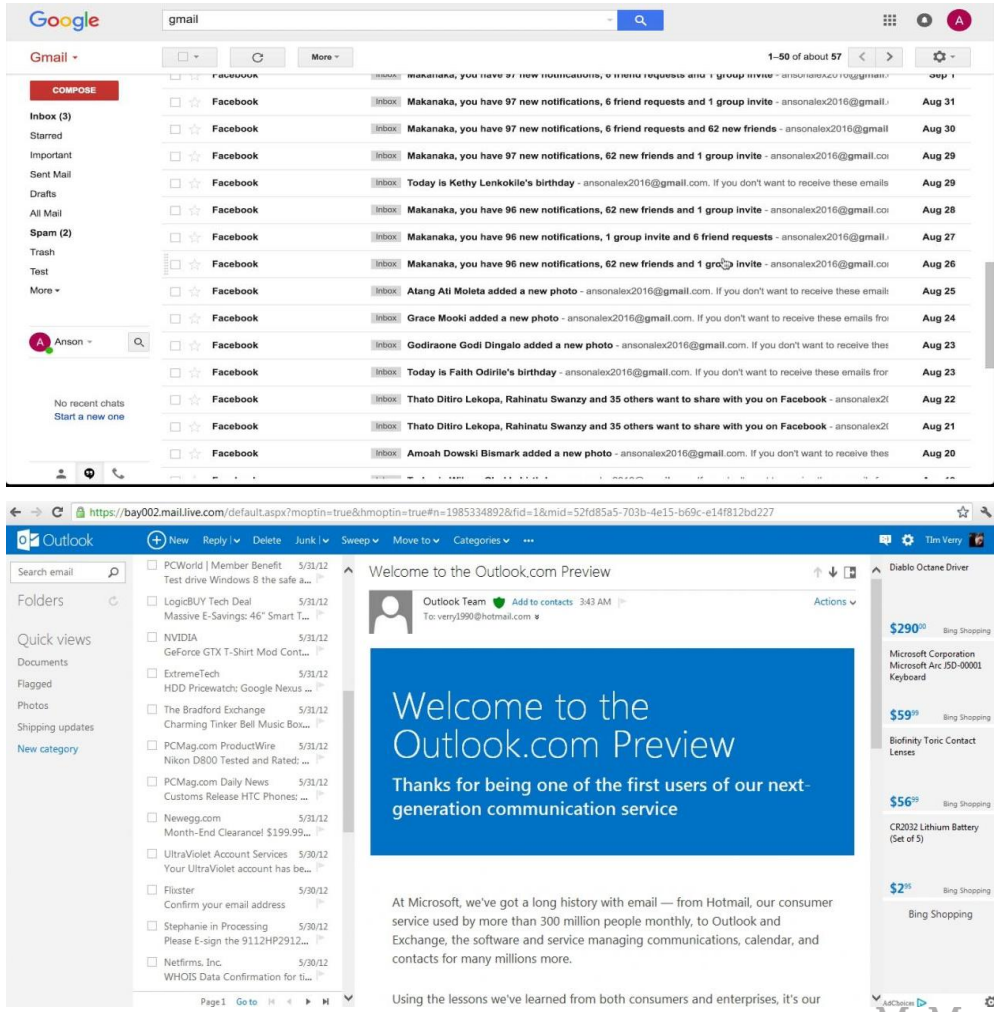
- Bruce "Tog" Tognazzini, usability consultant (Apple, Sun Microsystems, WebMD, etc)

Evaluation: Overview

Is the design good?

- What is “good”?

Which design is better?



Evaluation: Considerations

There are **a range of evaluation methods and settings**

Your choice of method should be informed by the following:

- Why
- What
- Where
- When

Evaluation: Why

What are **the goals** for the evaluation?

- What information do you currently need about your system to **move forward**?

Examples:

- Confirm your understanding of user requirements
- Assess your system against key usability and experience goals

Evaluation: What

What will you be evaluating?

Examples:

- A conceptual model
- An early paper prototype
- A high-fidelity prototype, perhaps with limited functionality
- A complete system

Evaluation: Where

Where will the evaluation take place?

Examples:

- In a controlled environment (e.g., lab space)
- In the field
- Online

Evaluation: When

Evaluation should occur throughout the design process

- From the first descriptions, sketches etc. of users needs through to the final product

Two main types of evaluation

- **Formative evaluation** is done at different stages of development to check that the product meets users' needs
- **Summative evaluation** assesses the quality of a finished product

Evaluation: Other Factors

- Budget
- Time constraints
- Access to user population
- Access resources
  - Space
  - Equipment…

Bottom line: Any evaluation is better than no evaluation!

"Discount usability" – Jakob Nielsen: can learn a lot from even just 5 people

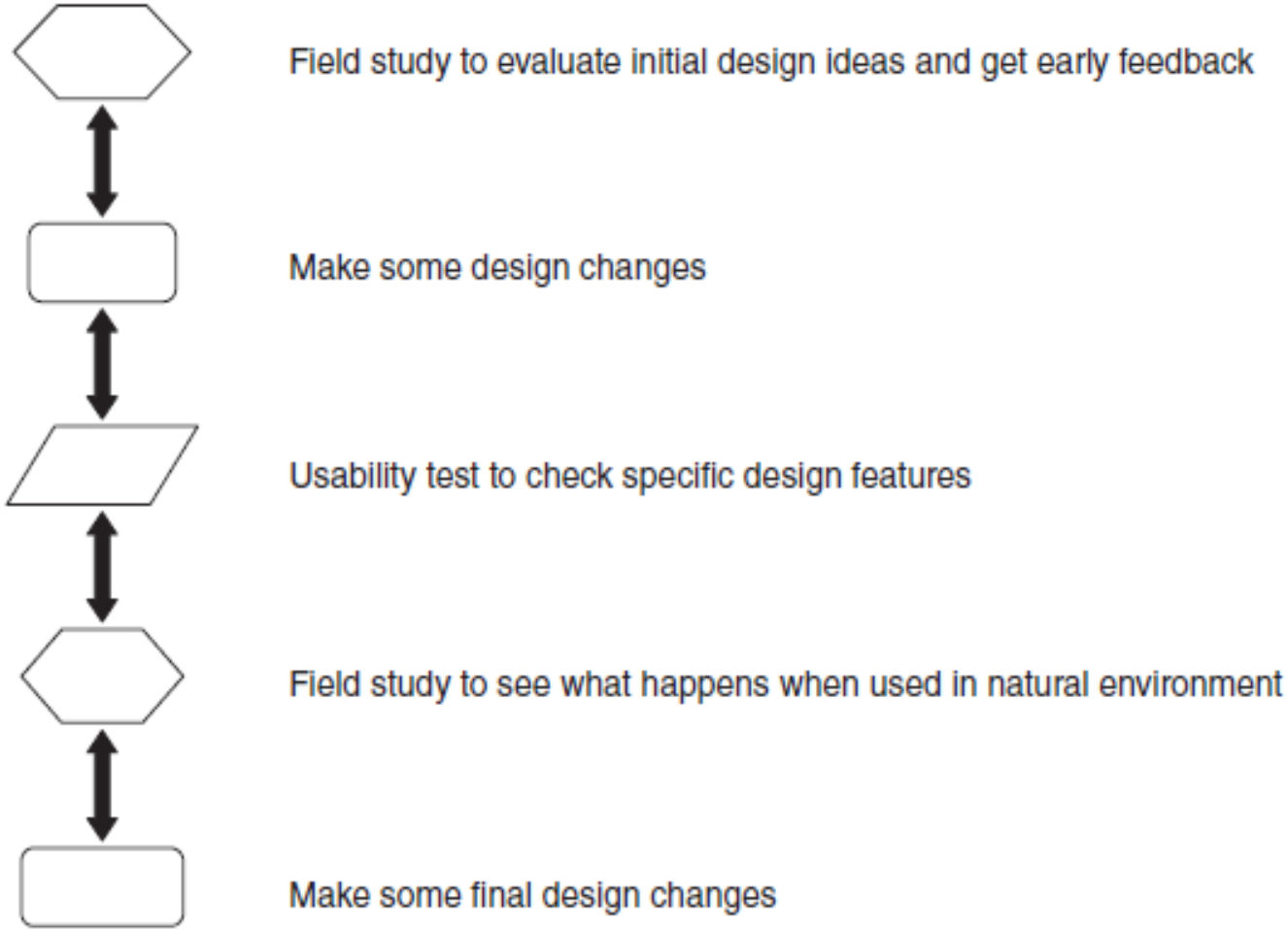
Evaluation: All Stakeholders

- Management
- Actual users
- Maintainers
- Others impacted

Evaluation: Types

- **Controlled settings involving users**
  - Usability Testing, Experiment
- **Natural settings involving users**
  - Field Deployment
- **Settings not involving users**
  - Performance Modeling (Fitt's Law, Hick's Law, KLM)
  - Inspections: Cognitive Walkthrough, Heuristics Evaluation

Evaluation: Types



Evaluation Types Complement One Another

Figure 13.1 Example of the way laboratory-based usability testing and field studies can complement each other

Evaluation: Involving Users

- Evaluation participants need to **be told why** the evaluation is being done, **what** they will be asked to do, and their **rights**
- **Informed consent forms provide this information**
- The design of the informed consent form, the evaluation process, data analysis and data storage methods are typically approved by a high authority (e.g., Institutional Ethics Review Board)

Evaluation: Data Analysis

So you do an evaluation and get data, now what?

Evaluation: Data Analysis

The **goals** of the evaluation dictate what kinds of data you will collect and focus on

Examples:

- Summaries of usability results, e.g., numbers from usability criteria
- Summaries, e.g., from interviews finding important themes for more early, exploratory studies
- Statistics and hypothesis testing for tightly controlled experiments

Think and plan: Why, What, Where

Evaluation: Data Analysis

- **Ecological validity:** does the environment of the evaluation distort the results?
- **Validity:** does the evaluation measure what it was intended?
- **Reliability:** does the method produce the same results on separate occasions?
- **Biases:** Are there biases that distort the results?
- **Scope:** How generalizable are the results?

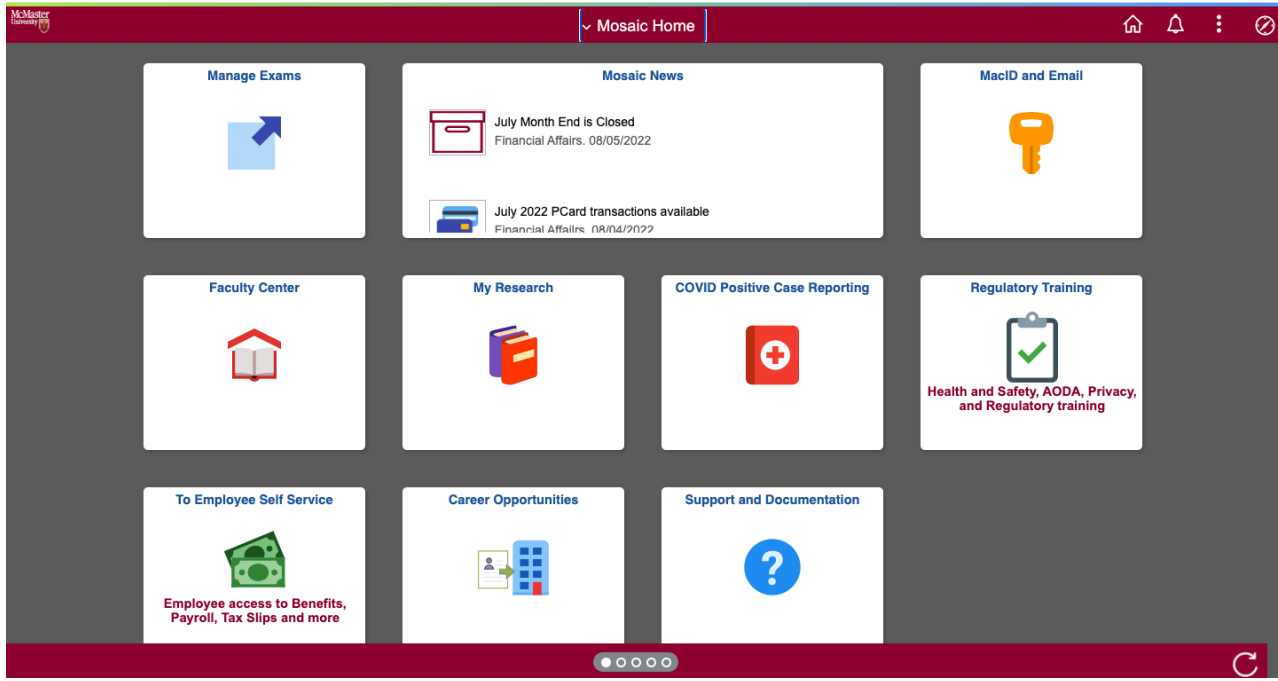
Week 12 Overview

- **Monday**
  - ~~**Introduction to Evaluation**~~
  - **Evaluation: Performance Modeling**
- **Wednesday**
  - Evaluation: Inspection
- **Friday**
  - Evaluation: Heuristics
  - Evaluation with Users

Evaluation: Analytical Evaluation

Sometimes we **can intuitively see** that there **are "things that are wrong"** with the design without the need of running a usability test

- But can we formalize this idea of "using our intuition" so that it is more **systematic**?



Evaluation: Analytical Evaluation

- Evaluations **without involving users**
- **Experts** assess the system using **structured techniques known to be effective at uncovering usability flaws**
- **Theoretical models** of human performance can predict actual use

Evaluation: Analytical Evaluation

- Doesn’t this contradict UCD?
- **Know your expertise!**
  - You’re not the user, so you need them
  - The user is not an HCI expert
- We can see clear problems that should be fixed before getting users
- Users sometimes too expensive to get involved?
  - Overall a bad argument (discount usability)

Analytical Evaluation: Types

Performance modelling

- Fitts' Law
- Hick-Hyman Law
- Keystroke Level Modelling

Inspections

- Cognitive walkthrough
- Heuristic evaluation

Analytical: Performance Modeling

Using **models of human behavior** to generate **quantitative predictions** of certain **interface actions** or **sequences of actions**

Analytical: Fitts' Law

One of most tested, lasting models in HCI (WIMP interfaces)

Models target acquisition performance (e.g., moving mouse to target)

$$T = a + b \log_2 \left[\frac{A}{W} + 1\right]$$

- T = Time
- A = Amplitude (distance)
- W = Width (size) of target
- a/b = empirically derived constants

$$time = a + b\ log_2 \left[\frac{distance}{width} + 1\right]$$

Analytical: Fitts' Law

How can you use the law?

- Fitts' Law predicts how long it will take users to acquire targets once they know which target to select

$$time = a + b\ log_2 \left[\frac{distance}{width} + 1\right]$$

What about decision time?

Analytical: Hick-Hyman Law

Hick-Hyman Law (or Hick's Law) models the time it takes users to **decide** between **n familiar alternatives**

Analytical: Hick-Hyman Law

When items are **equi-probable**:

$$DT = a + b(\log_2 n)$$

When certain items are **more likely** to be chosen than others:

$$DT = a + b\left(\sum_{i=1}^{n} p_i \log_2\left({}^{1}/_{p_i}\right)\right)$$

Analytical: Hick-Hyman Law

Models **decision time**, not **searching time**

If the user is not familiar with the interface elements, they often need to **investigate** each one (and make decisions). **It's not a pure search task.**

- Time to search through n items is more linear, rather than logarithmic

Analytical: Keystroke Level Model

Given a task consisting of **a sequence of steps**

- How long will it take the user to perform those steps given a specific interface?

Keystroke Level Model (KLM)

- Models **performance** given a **sequence of steps** for an **expert user**

Analytical: Keystroke Level Model

How is performance calculated?

- Individual steps described using **operators**.
- Given values are **accepted reasonable defaults** (by experts)
- Sum up times for individual steps

K → keystroking = 0.35s

P → pointing = 1.10s (Fitts' Law for greater precision)

B → Button press or release (mouse) = 0.10 seconds (BB for mouse click = 0.20seconds)

H → homing hands = 0.4s

(placing hands on desired device)

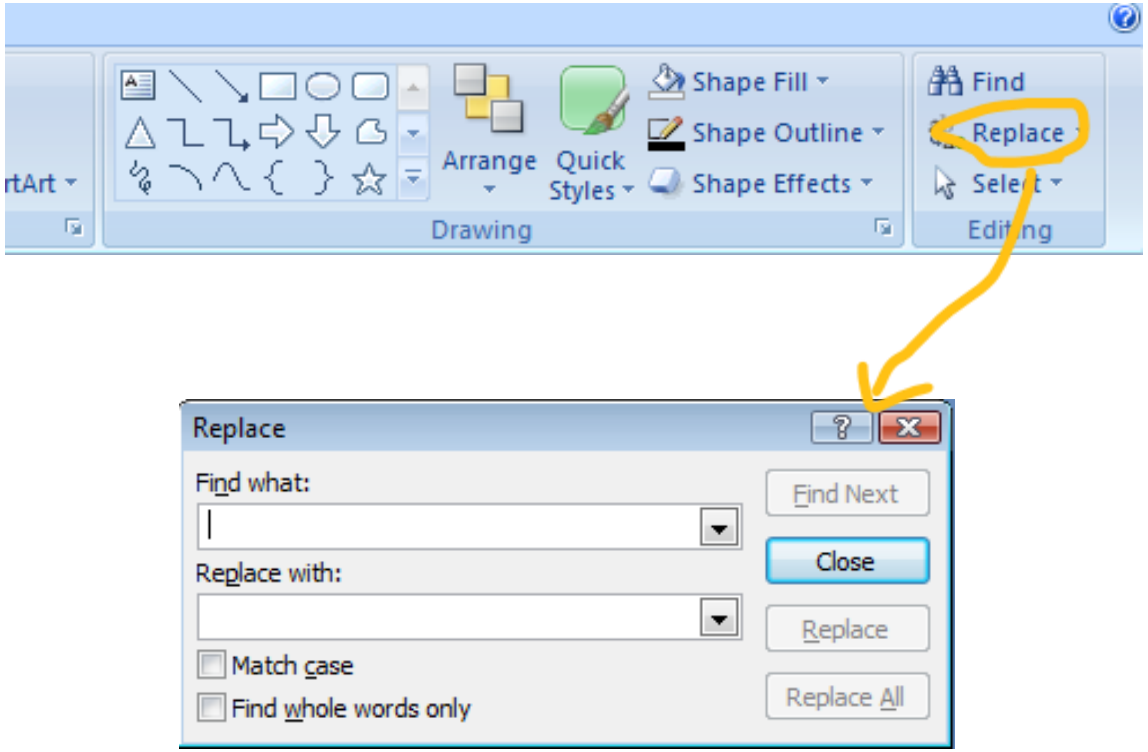
D → drawing = variable with length of line

M → Mental prep = 1.35s

R → response operator by system = 1.2s

KLM: Example

How long would it take the user to replace all occurrences of a 4-letter word with a new 4-letter word



KLM: Example

Description	Operation	Time (sec)
Reach for mouse	H	0.40
Move pointer to “Replace” button	P	1.10
Click on “Replace” button	BB	0.2
Home on keyboard	H	0.40
Type old word	M, K4	2.75
Reach for mouse	H	0.4
Move pointer to correct field	P	1.10
Click on field	BB	0.2
Home on keyboard	H	0.4
Type new word	M, K4	2.75
Reach for mouse	H	0.4
Move pointer to Replace All	P	1.10
Click Replace All	BB	0.2
Total		**11.4**

Performance Modeling: Advantages

- Can evaluate **components of interface** prior to building it
- Good for comparing different interface possibilities
- Can get the kinks out of interface prior to full **user testing/experimentation**

Performance Modeling: Problems

- Difficult to model **complex tasks**
  - For example, consider designing a KLM for your complete project
- Most models consider **only expert behaviour**
- For really **accurate** predictions, **coefficients** (those a's and b's) need to be determined empirically

Week 12 Overview

- **Monday**
  - ~~**Introduction to Evaluation**~~
  - ~~**Evaluation: Performance Modeling**~~
- **Wednesday**
  - Evaluation: Inspection
- **Friday**
  - Evaluation: Heuristics
  - Evaluation with Users