

**Cornell University**  
**Computing and Information Science**

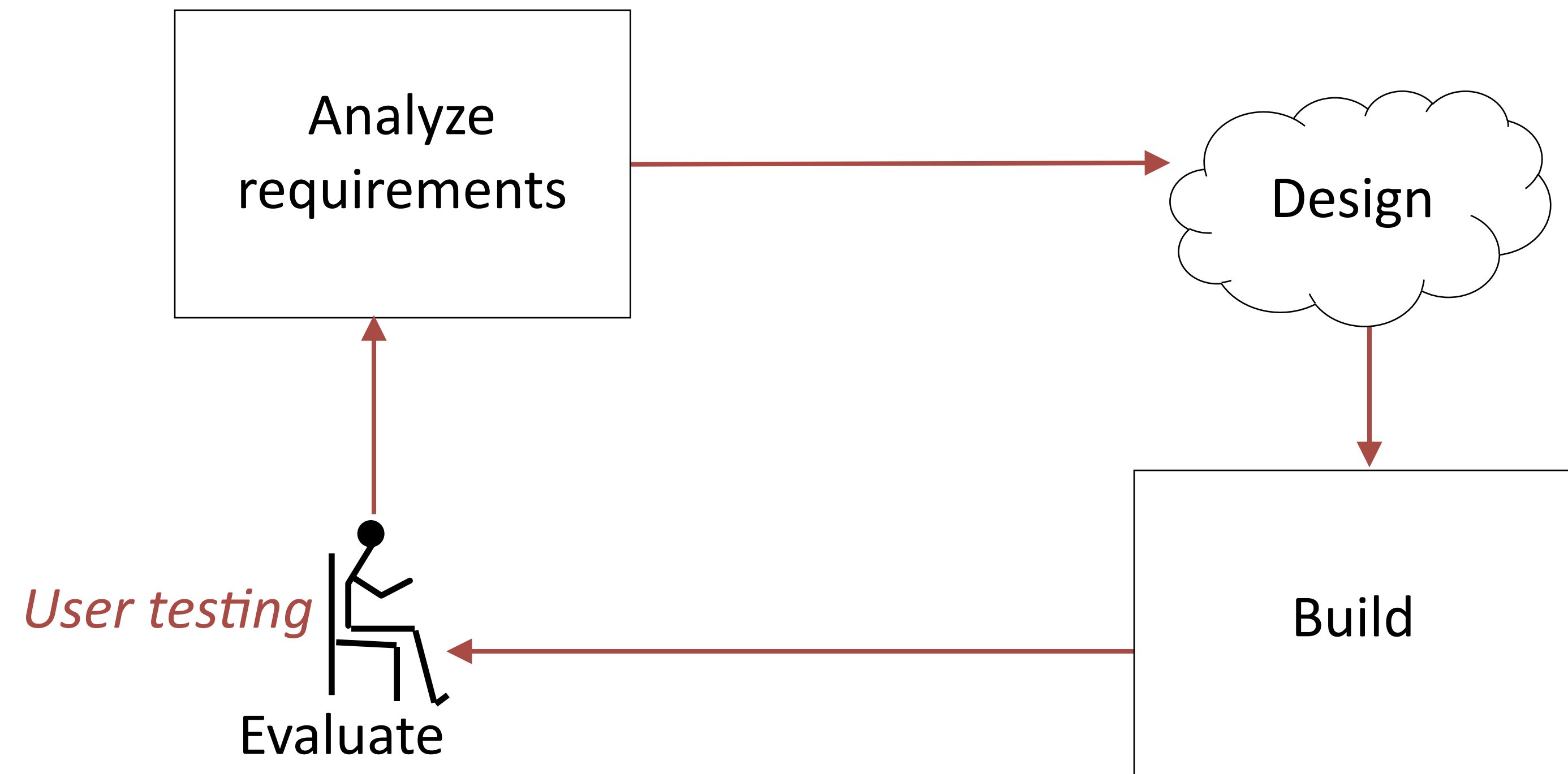
---

**CS 5150 Software Engineering**  
**11. Evaluation and User Testing**

William Y. Arms

# The Analyze/Design/Build/Evaluate Loop

---



Whenever possible, the design and evaluation should be done by different people.

# Evaluation

---

If your system has users, the schedule should include time for user testing and time to make changes after the user testing is completed.

## When to do evaluation

- Iterative improvements during development.
- Making sure that a system is usable before launching it.
- Iterative improvements after launch.

## Methods of evaluation

- Empirical evaluation with users (**user testing**)
- Measurements on operational systems
- Analytical evaluation: without users (*not in CS 5150*)

# Evaluation

---

## How do you measure usability?

Usability comprises the following aspects:

### Effectiveness

The accuracy and completeness with which users achieve certain goals

**Measures:** quality of solution, error rates

### Efficiency

The relation between the effectiveness and the resources expended in achieving them

**Measures:** task completion time, learning time, number of clicks

### Satisfaction

The users' comfort with and positive attitudes towards the use of the system

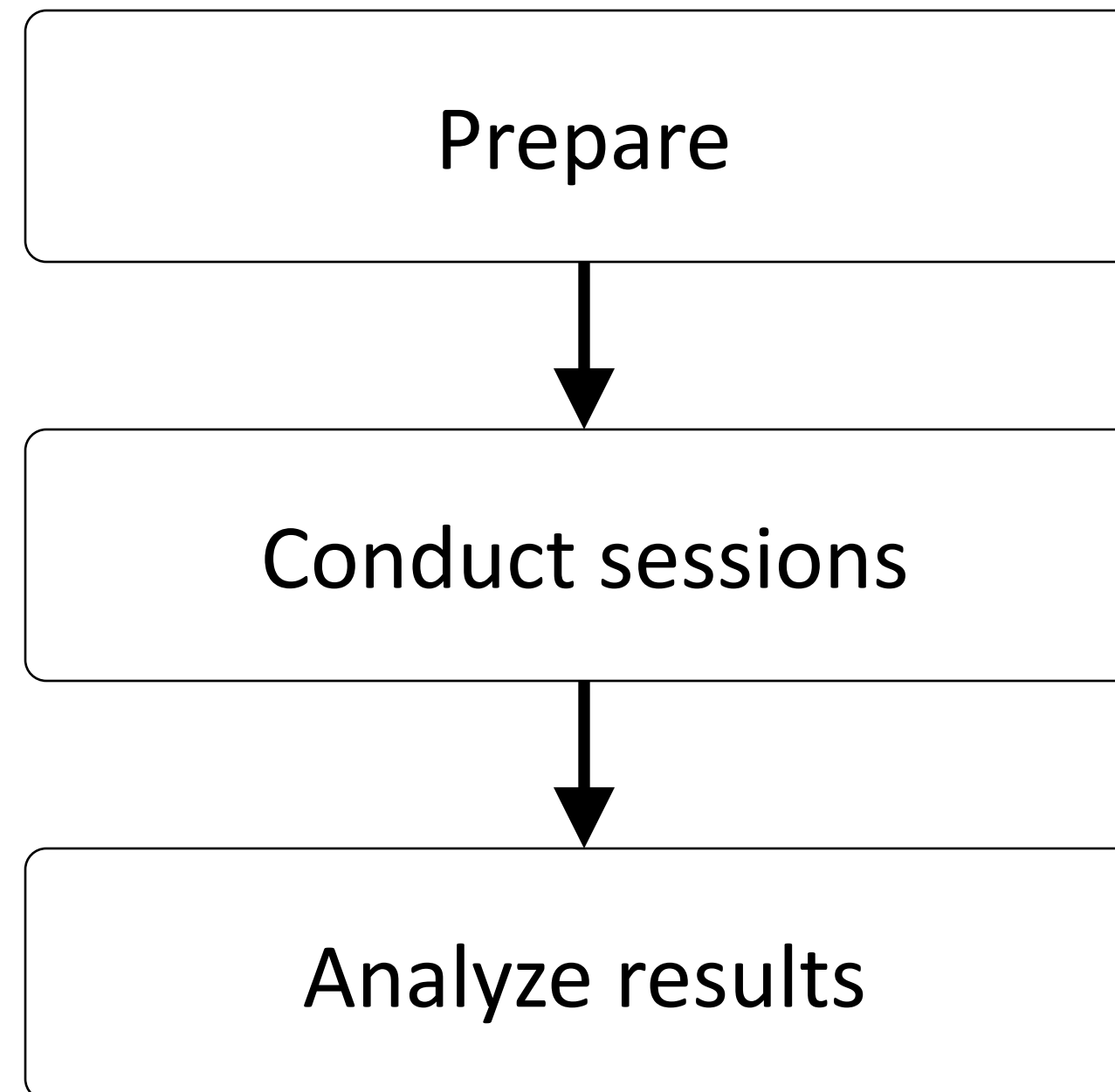
**Measures:** attitude rating scales

*From ISO 9241-11*

# Evaluation with Users

---

Stages of evaluation with users:



User testing is time-consuming, expensive, and essential.

# Evaluation with Users: Preparation

---

Determine **goals** of the usability testing

*“Can a user find the required information in no more than 2 minutes?”*

Write the **user tasks**

*“Given a new customer application form, add a new customer to the customer database.”*

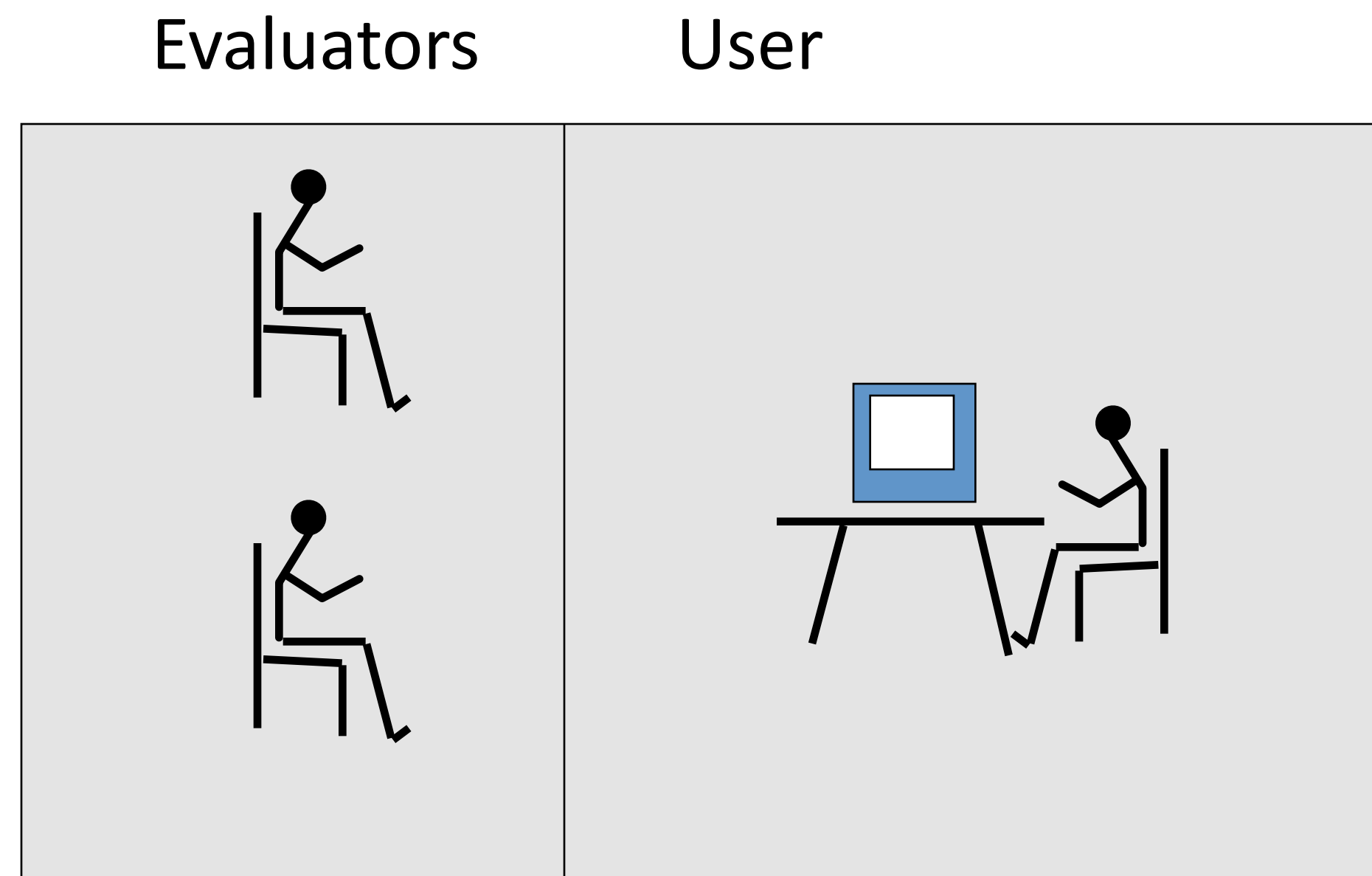
Recruit **participants**

Use the descriptions of users from the requirements phase to determine categories of potential users and user tasks

# Usability Laboratory

---

**Concept: monitor users while they use system**



*one-way  
mirror*

# Evaluation with Users: Sessions

---

## Conduct the session

Usability Lab

Simulated working environment

## Observe the user

Human observer(s)

Video camera

Audio recording

## Inquire satisfaction data





# Evaluation: Number of Users

---

## Number of users

A great deal can be learned from user testing with a small number of users, even as few as five people.

- Try to find different types of user (young/old, experienced/beginners, etc.).
- Prepare carefully.
- Combine structured tests with free form interviews.
- Have at least two evaluators for every test.

# Results Analysis

---

## Test the system, not the users

Respect the data and users' responses. Do not make excuses for designs that failed.

If possible, use statistical summaries.

Pay close attention to areas where users:

- were frustrated

- took a long time

- could not complete tasks

Note aspects of the design that worked and make sure they are incorporated in the final product.

# Eye Tracking at Google

---



*Dan Russell*  
2007



# Eye Tracking at Google

The image shows a screenshot of a Google search results page for the query "children's unicycle". The search bar at the top contains the text "children's unicycle" and a "Search" button. Below the search bar, there are tabs for "Web", "Images", "Video", "News", "Maps", and "more". The "Web" tab is selected. The search results are listed below the "Web" tab, with six results numbered 1 through 6. Red arrows indicate the eye-tracking path, starting from the search bar and moving down through the results. The results are as follows:

- 1** [Unicycle UK.com - F.A.Q. - What size?](#)  
12" wheel unicycle: this is a small children's unicycle size. It's good for children who are too small to ride a 16" unicycle, but it needs smooth ground ...  
[www.unicycle.uk.com/FAQ.asp?iCategory=53](#) - 23k - [Cached](#) - [Similar pages](#)
- 2** [Selecting a unicycle: Unicycle.com NZ : buy a unicycle or learn ...](#)  
16" wheel unicycle: this is a children's unicycle, the small wheel makes it only suitable for smooth areas. Best used indoors or on smooth ground. ...  
[www.unicycle.co.nz/View.php?action=Page&Name=Selectingaunicycle](#) - 22k - [Cached](#) - [Similar pages](#)
- 3** [100 Miles for Kids - The Goal](#)  
'The Afghan Mobile Mini Circus for Children is an established ... attempt to break the GUINNESS WORLD RECORD for the ONE HOUR UNICYCLE DISTANCE RECORD. ...  
[www.unicycle4kids.org/](#) - 9k - [Cached](#) - [Similar pages](#)
- 4** [Unicycles page at Juggling world](#)  
This is a children's unicycle, the small wheel makes it only suitable for very smooth areas. Best used indoors or on smooth ground; not so good outdoors ...  
[www.jugglingworld.biz/shop/products\\_unicycles.html](#) - 100k - [Cached](#) - [Similar pages](#)
- 5** [Buy a Unicycle: Unicycle.com AU : buy a unicycle or learn unicycling](#)  
Check out a Unicycle Learners Pack for an easy and economical way to take your first steps into the One Wheeled World ... Suitable as a Children's Unicycle ...  
[www.unicycle.au.com/View.php?action=Page&Name=Unicycles](#) - 10k - [Cached](#) - [Similar pages](#)
- 6** [Article - News - A unicycle ride for children](#)  
Adam Brody, 21, of San Juan Capistrano, led a charity event Saturday that benefits the Orangewood Children's Foundation. The Unicycle Club of Southern ...  
[www.ocregister.com/ocregister/news/homepage/article\\_1293795.php](#) - 31k - [Cached](#) - [Similar pages](#)

Dan Russell  
2007



1 0.212

2 0.253

3 0.283

4 0.284

5 0.283

6 0.617

7 0.567

8 0.267

9 0.200

10 0.200

11 0.233

12 0.517

13 0.450

18 0.334

19 0.334

20 0.250

21 0.250

msn.

MSN Home My MSN Sign In

Friday, Apr 05

Search Web: Search

Ben Affleck, Britney & more: Fashion Police 'spice' it up

Which city has the most unmarried men?

Free DVD player with any phone purchase

Buy for only \$4 and

Explore

Buy Books

Buy Spring Tulips

Buy Music & Movies

Go Wireless

Hotel Deals from \$39

Internet Access Specials

Outlet Sale! 30-80% Off

Send Money

Woman in yoga pose (Image Credit: Jack Ambrose/Getty Images)

Stay in Vegas from \$45

Manson's legal angst

Small biz ideas that pay

Don't worry, be happy

Tax filer's checklist

Post your resume online

Cartoon-maker for kids

5 trendy gym workouts

Bikram yoga to BOSU

balls: fun spring crazes

Shape-obsessed: Has Vogue gone too far?

'O Brother' to Brandy: 10 chart-topping CDs

MSN Top Headlines

MSNBC News

U.S. envoy meets with Arafat

W. Bush's first plan

pus tons

Surviving the sneezin' season

Slate: Opinion & Commentary

Friday Fun

Brad Pitt 'Spy' scenes

Find your Chinese sign

New lip-reading cell phones?

Rap with NHL experts

Sick of AOL?

Image Credit: Euan Miles/C Images

50 new reasons to switch to MSN

CNBC on MSN Money

Stock Quotes

Datek | Ameritrade | ShareBuilder | Trade Now | File Your Taxes

Name	Last	Change	% Chg
Dow	10,235.17	unch	0.00%
Nasdaq	1,789.75	unch	0.00%

Save Big on Books, Music and DVDs at Amazon.com!

MSN Channels

Autos

Business

Careers

City Guides

Entertainment

Games

Health

House & Home

Do-it-yourself Guide

Kids

Learning & Research

Love & Relationships

Music & Radio

News

Sports by ESPN

Tech & Gadgets

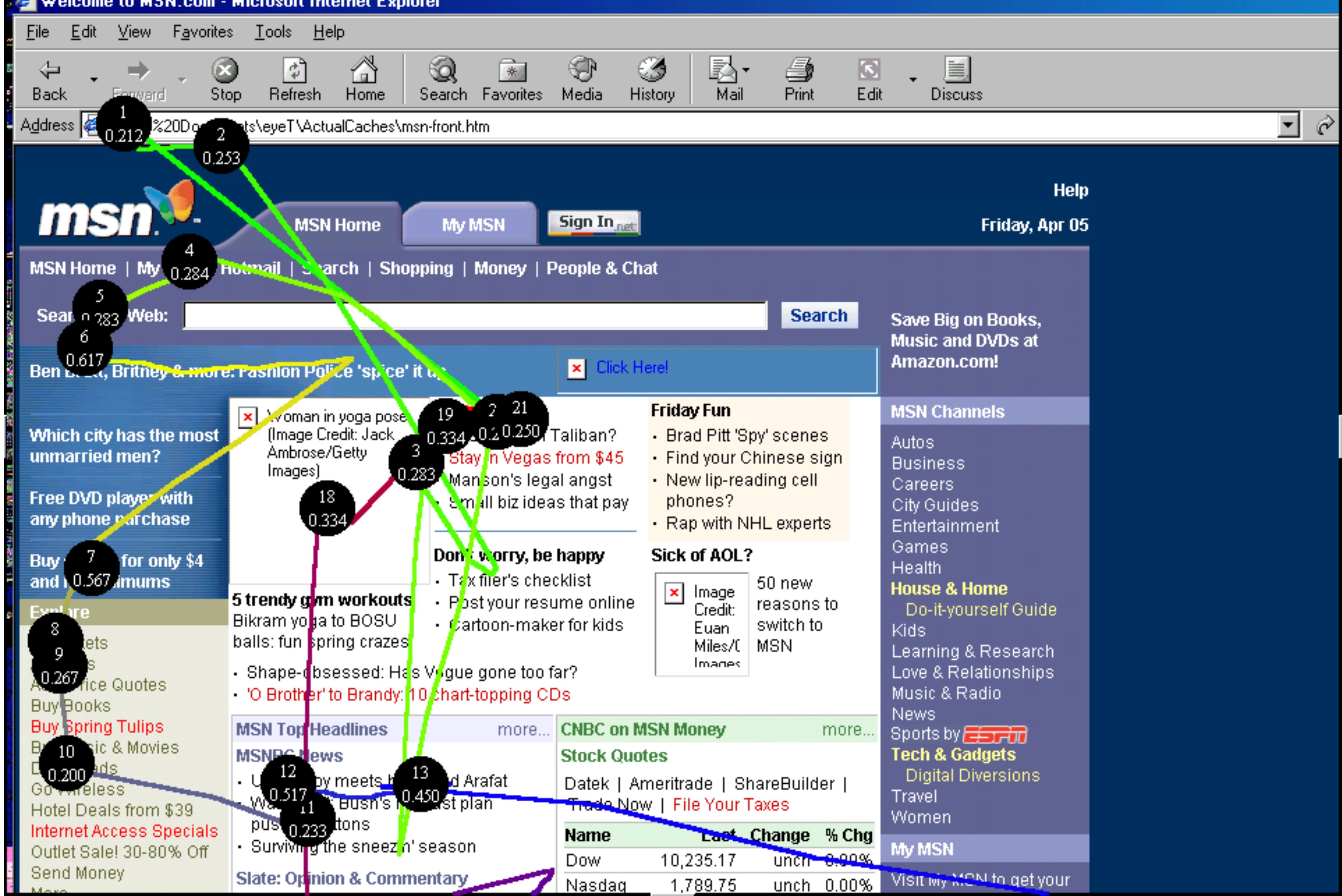
Digital Diversions

Travel

Women

My MSN

Visit my MSN to get your



# A CS 5150 Project: Methodology

---

The next few slides are from a CS 5150 presentation

## ***How we're user testing:***

- One-on-one, 30-45 min user tests with staff levels
- Specific tasks to complete
- No prior demonstration or training
- Pre-planned questions designed to stimulate feedback
- Emphasis on testing system, not the stakeholder!
- Standardized tasks / questions among all testers

# A CS 5150 Project: Methodology

---

## ***How we're user testing:***

- Types of questions we asked:
- Which labels, keywords were confusing?
- What was the hardest task?
- What did you like, that should not be changed?
- If you were us, what would you change?
- How does this system compare to your paper based system
- How useful do you find the new report layout? (admin)
- Do you have any other comments or questions about the system?  
(open ended)







# A CS 5150 Project: Results

*What we've found: Issue #2, Inconspicuous Edit/ Confirmations!*

New Entry

Search/Edit [2]

Export [3]

Your entry has been recorded once:  
Category: Reference  
Label: 1 to 5 minutes  
Medium: In Person  
Notes: gsd  
You did not refer to a librarian/unit for this entry.  
April 5, 2006, 12:54 AM

edit

New Entry

Category: Reference

Label: 1 to 5 minutes

Medium: In Person

Notes:

# A CS 5150 Project: Results

*What we've found: Issue #3, Confirmation Terms*

The screenshot shows a web application interface. At the top, there are three buttons: "New Entry", "Search/Edit [2]", and "Export [3]". Below these buttons is a confirmation message box with a red border. The message states: "Your entry has been recorded once:" followed by the details: "Category: Reference", "Label: 1 to 5 minutes", "Medium: In Person", and "Notes: gsd". Below these details, it says "You did not refer to a librarian/unit for this entry." and "April 5, 2006, 12:54 AM". There is an "edit" link at the bottom right of the message box. Below the message box is a form titled "New Entry" with a green header. The form contains four fields: "Category:" with a dropdown menu showing "Reference", "Label:" with a dropdown menu showing "1 to 5 minutes", "Medium:" with a dropdown menu showing "In Person", and "Notes:" with a text input field.

New Entry

Search/Edit [2] Export [3]

Your entry has been recorded once:

Category: Reference  
Label: 1 to 5 minutes  
Medium: In Person  
Notes: gsd

You did not refer to a librarian/unit for this entry.  
April 5, 2006, 12:54 AM [edit](#)

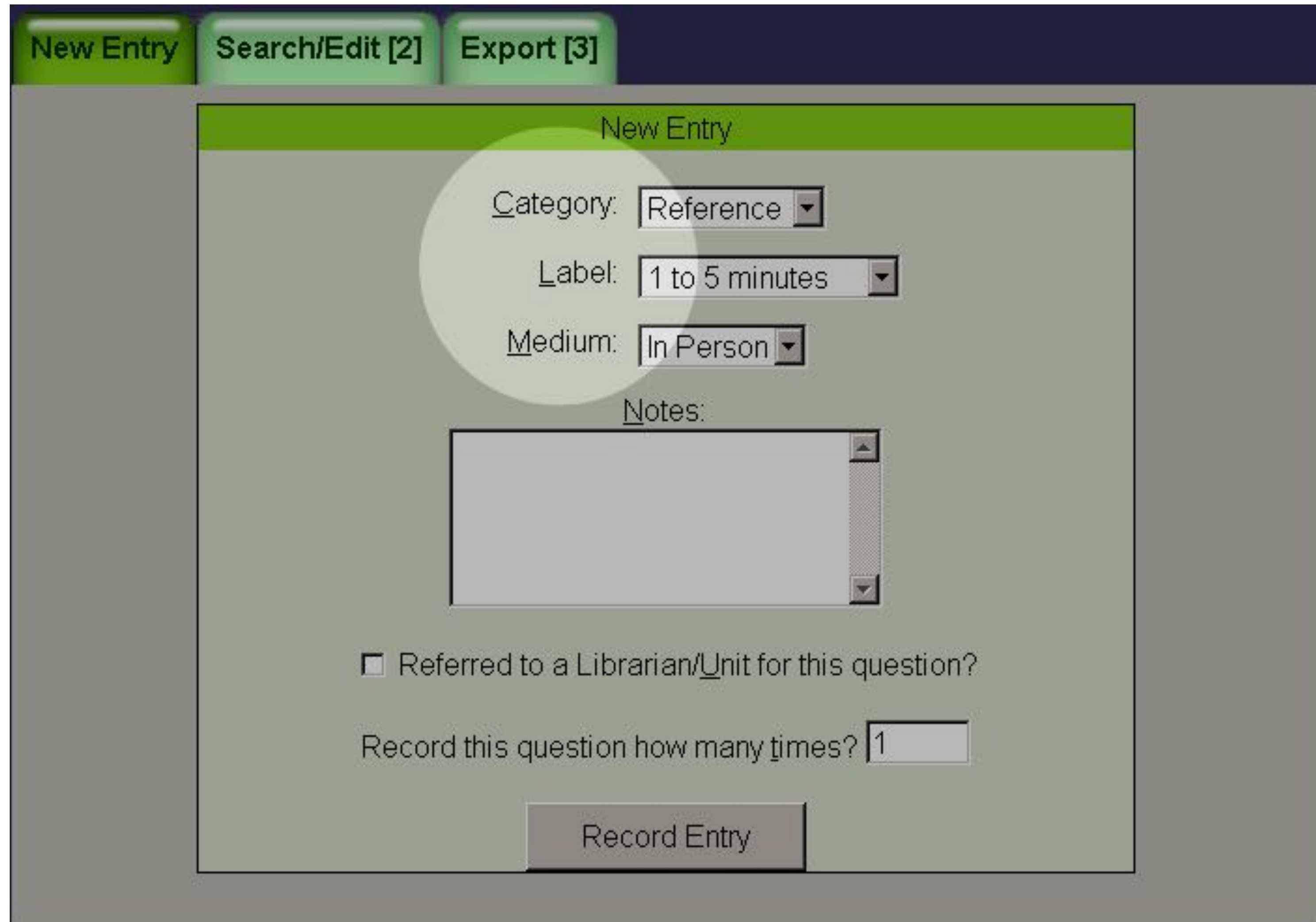
New Entry

Category: Reference  
Label: 1 to 5 minutes  
Medium: In Person  
Notes:

# A CS 5150 Project: Results

---

*What we've found: Issue #4, Entry Semantics*



The screenshot shows a web application interface for creating a new entry. At the top, there are three buttons: 'New Entry' (highlighted in green), 'Search/Edit [2]', and 'Export [3]'. Below these is a 'New Entry' form. The form has a title bar 'New Entry' and contains the following fields:

- Category: Reference (dropdown menu)
- Label: 1 to 5 minutes (dropdown menu)
- Medium: In Person (dropdown menu)
- Notes: (text area)
- ☐ Referred to a Librarian/Unit for this question?
- Record this question how many times? 1 (text input)
- Record Entry (button)

A light green circle highlights the 'Category', 'Label', and 'Medium' dropdown menus, indicating the focus of the study.



# A CS 5150 Project: Results

*What we've found: #5, Search Results Disambiguation & Semantics*

New Search

Select	Edit	Date/Time	Category	Type of Question	Medium	Staff
<input type="checkbox"/>	edit	2006-04-03 16:12:00	Reference	16 to 30 minutes	In Person	Librarian
<input type="checkbox"/>	edit	2006-04-03 16:04:00	Directional	1 to 5 minutes	In Person	Librarian
<input type="checkbox"/>	edit	2006-04-05 01:03:35	Reference	16 to 30 minutes	E-mail	Librarian

Reset

Delete

# Evaluation based on Measurement

---

**Basic concept:** log events in the users' interactions with a system

Examples from a Web system

- Clicks (when, where on screen, etc.)
- Navigation (from page to page)
- Keystrokes (e.g., input typed on keyboard)
- Use of help system
- Errors

May be used for statistical analysis or for detailed tracking of individual user.

# Evaluation based on Measurements

---

## Analysis of system logs

- Which user interface options were used?
- When was the help system used?
- What errors occurred and how often?
- Which hyperlinks were followed (click through data)?

## Human feedback

- Complaints and praise
- Bug reports
- Requests made to customer service

# The Search Explorer: a User Session

project uhuv

## Search Explorer

[Reset selections >>](#)

**Session Information . . .**


Session Duration: 00:03:56  
Total # Queries: 2  
Total # Clicks: 7

**Query Information . . .**

Query: haleys comet  
All resources searched  
All formats included  
Results per Page: 20  
Search Type: Simple

**Result Information . . .**

Title: Comet Definitions  
Time spent: 00:01:01  
At time: 05:40:22



Search Sessions	Queries for session 2	Results for query 2 [20 results]
1. <a href="#">Mon Apr 28 23:26:26 EDT 2003</a>	1. <a href="#">haley's comet</a>	
2. <a href="#">Tue Apr 29 00:37:27 EDT 2003</a>	2. <a href="#">haleys comet</a>	
3. <a href="#">Tue Apr 29 10:56:34 EDT 2003</a>		3. <a href="#">Interacting electrons in magnetic fields: Tracking p</a> <a href="#">meta</a>
4. <a href="#">Tue Apr 29 10:58:50 EDT 2003</a>		4. <a href="#">Comets</a> <a href="#">meta</a>
5. <a href="#">Tue Apr 29 13:21:25 EDT 2003</a>		5. <a href="#">Comets and meteor showers</a> <a href="#">meta</a> Click #2
6. <a href="#">Thu May 01 15:09:26 EDT 2003</a>		6. <a href="#">Comet Definitions</a> <a href="#">meta</a> Click #3
7. <a href="#">Thu May 01 20:42:31 EDT 2003</a>		7. <a href="#">Comet Observation Home Page</a> <a href="#">meta</a> Click #5,7
8. <a href="#">Mon May 05 17:25:02 EDT 2003</a>		8. <a href="#">Comets</a> <a href="#">meta</a>

[JPL HOME](#) [EARTH](#) [SOLAR SYSTEM](#) [STARS & GALAXIES](#) [TECHNOLOGY](#)

## Comet Definitions

The following are generally accepted definitions for terms related to cometary astronomy.

### absolute magnitude (H<sub>0</sub>)

The brightness of a comet when it is at 1 AU from both the Earth and Sun. As this virtually never happens, this quantity is calculated from the comet's light curve. Unfortunately, this quantity is far from absolute. It can be different pre- and post-perihelion. It can also change from apparition to apparition (for periodic comets).

<http://encke.jpl.nasa.gov/define.html>



# Refining the Design based on Evaluation

---

## Do not allow evaluators to become designers

Designers are poor evaluators of their own work, but know the requirements, constraints, and context of the design:

- Some user problems can be addressed with small changes
- Some user problems require major changes
- Some user requests (e.g., lots of options) are incompatible with other requests (e.g., simplicity)

Designers and evaluators need to work as a team



**Cornell University**  
**Computing and Information Science**

---

**CS 5150 Software Engineering**  
**11. Evaluation and User Testing**

End of Lecture