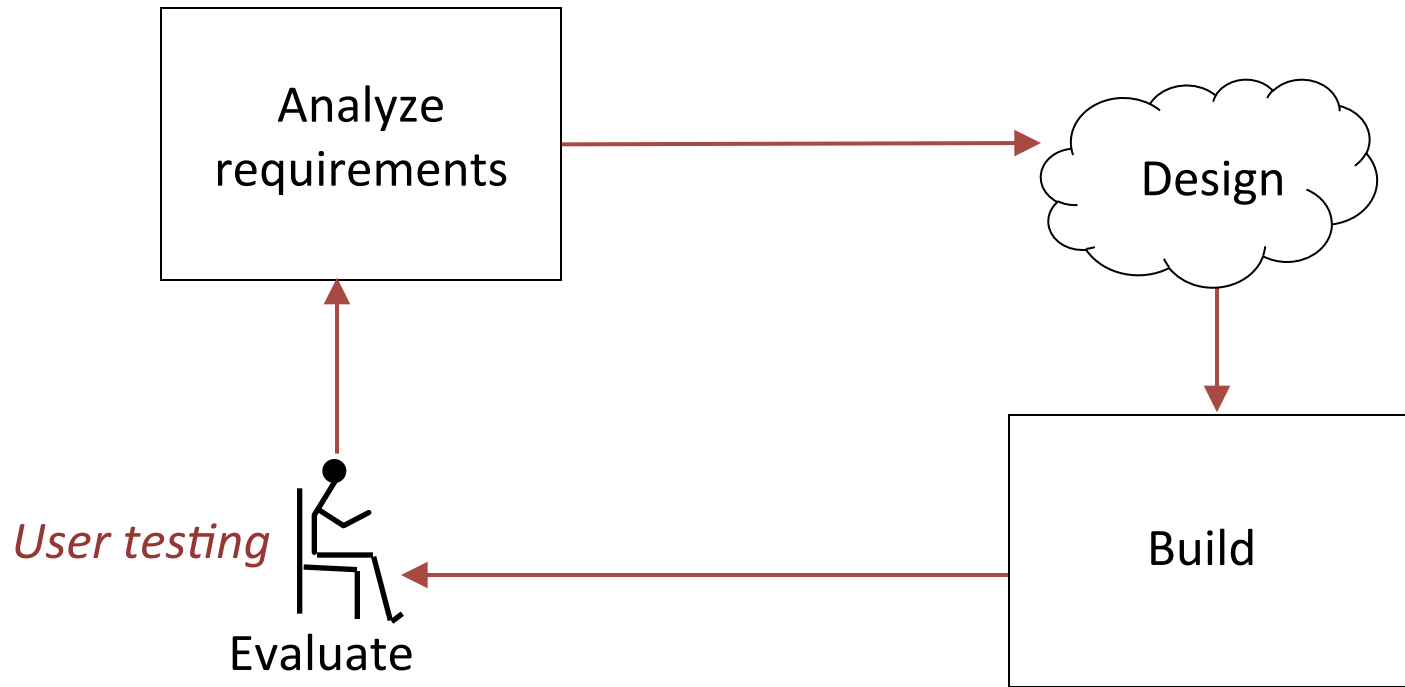


Cornell University
Computing and Information Science

CS 5150 Software Engineering
Evaluation and User Testing

William Y. Arms

Usability: The Analyze/Design/Build/Evaluate Loop



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 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 Application: Reconstruct a User Session

project uhuu
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

1. [Mon Apr 28 23:26:26 EDT 2003](#)
2. [Tue Apr 29 00:37:27 EDT 2003](#)
3. [Tue Apr 29 10:56:34 EDT 2003](#)
4. [Tue Apr 29 10:58:50 EDT 2003](#)
5. [Tue Apr 29 13:21:25 EDT 2003](#)
6. [Thu May 01 15:09:26 EDT 2003](#)
7. [Thu May 01 20:42:31 EDT 2003](#)
8. [Mon May 05 17:25:02 EDT 2003](#)

Queries for session 2

1. [haley's comet](#)
2. [haleys comet](#)

Results for query 2 [20 results]

3. [Interacting electrons in magnetic fields: Tracking p](#) meta
4. [Comets](#) meta
5. [Comets and meteor showers](#) meta
Click #2
6. [Comet Definitions](#) meta
Click #3
7. [Comet Observation Home Page](#) meta
Click #5, 7
8. [Comets](#) meta

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

Comet Definitions

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

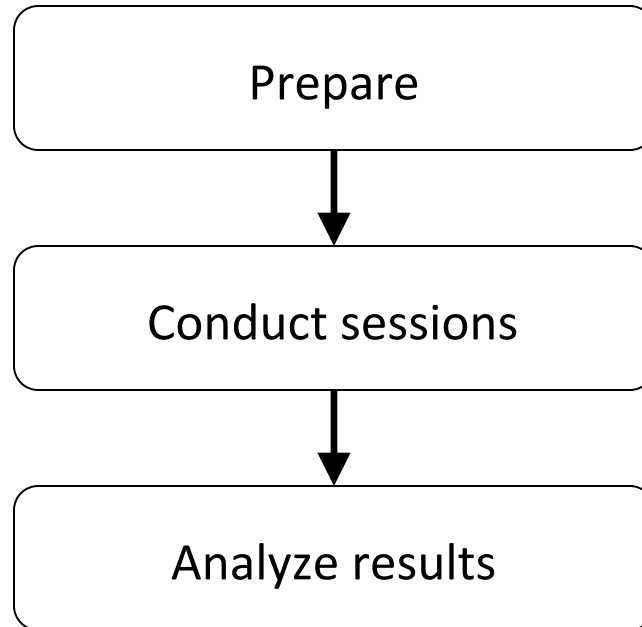
absolute magnitude (H_0)

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>

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**

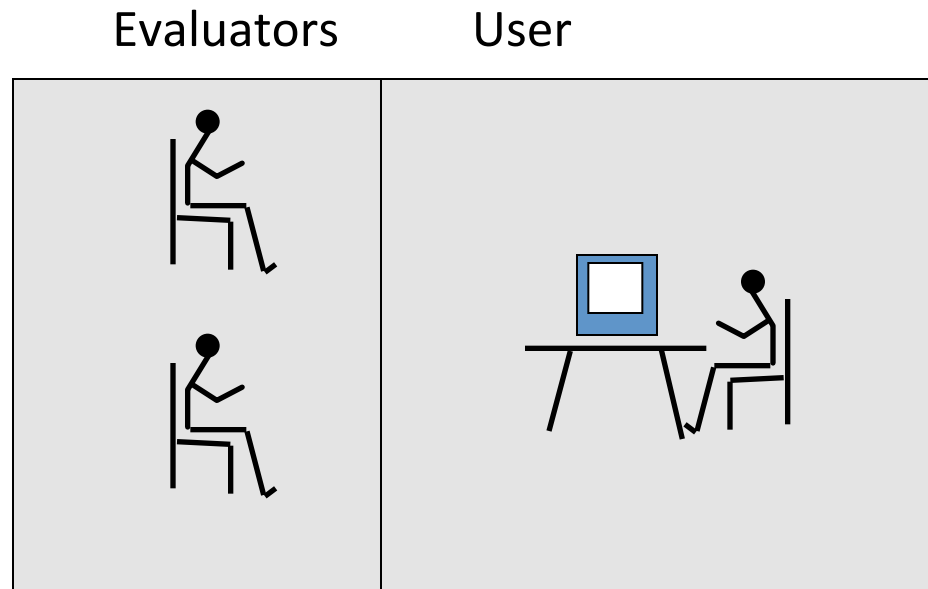
“Given a new customer application form, add a new customer to the customer database.”

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 with Users: 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 for "children's unicycle". The search bar contains the text "children's unicycle" and a "Search" button. Below the search bar, there are navigation links for "Web", "Images", "Video", "News", "Maps", and "more". The search results are displayed under the "Web" heading. There are six search results, each with a red circle containing a number (1-6) and a red arrow pointing to a specific part of the result. The results are:

- 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?Category=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=Selecting_a_unicycle](#) - 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](#)

Evaluation Example: Eye Tracking

Welcome to MSN.com - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media History Mail Print Edit Discuss

Address [http://www.msn.com/](#)

msn.com MSN Home My MSN Sign In .net

Help Friday, Apr 05

MSN Home | My MSN | Hotmail | Search | Shopping | Money | People & Chat

Search Web: Search

Ben Dinklage, Britney & more: Fashion Police 'spice' it up [Click Here!](#)

Which city has the most unmarried men?

Free DVD player with any phone purchase

Buy **7** for only \$4 and **10** mums

Explore **8** **9** **10** **11** **12** **13** **14** **15** **16** **17** **18** **19** **20** **21** **22** **23** **24** **25** **26** **27** **28** **29** **30** **31** **32** **33** **34** **35** **36** **37** **38** **39** **40** **41** **42** **43** **44** **45** **46** **47** **48** **49** **50** **51** **52** **53** **54** **55** **56** **57** **58** **59** **60** **61** **62** **63** **64** **65** **66** **67** **68** **69** **70** **71** **72** **73** **74** **75** **76** **77** **78** **79** **80** **81** **82** **83** **84** **85** **86** **87** **88** **89** **90** **91** **92** **93** **94** **95** **96** **97** **98** **99** **100**

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

Friday Fun

- Brad Pitt 'Spy' scenes
- Find your Chinese sign
- New lip-reading cell phones?
- Rap with NHL experts

Sick of AOL?

50 new reasons to switch to MSN

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 [more...](#)

MSNBC News

- U.S. envoy meets with Arafat
- W.C. Bush's first plan
- pus: tons
- Surviving the sneezin' season

Slate: Opinion & Commentary

CNBC on MSN Money [more...](#)

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 (second milestone)

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

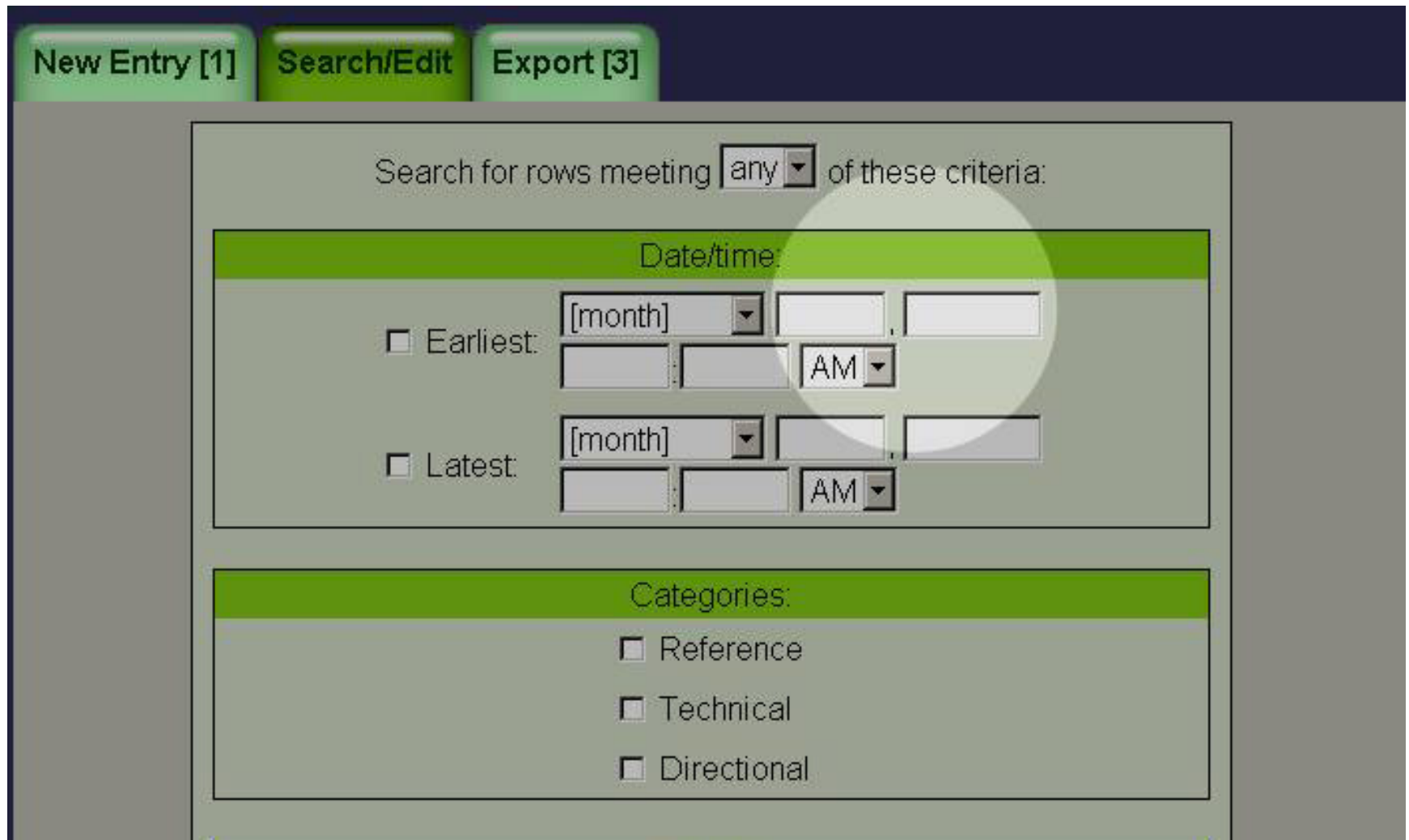
How we're user testing:

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 #1, Search Form Confusion!



The image shows a screenshot of a web application interface. At the top, there are three buttons: "New Entry [1]", "Search/Edit", and "Export [3]". Below these buttons is a search form. The form has a header that says "Search for rows meeting [any] of these criteria:". The word "any" is in a dropdown menu, which is highlighted with a green circle. Below the header, there are two sections: "Date/time:" and "Categories:". The "Date/time:" section has two rows, each with a checkbox and a date/time input field. The "Categories:" section has three checkboxes: "Reference", "Technical", and "Directional".

New Entry [1] Search/Edit Export [3]

Search for rows meeting **any** of these criteria:

Date/time:

Earliest: [month] [] [] [] AM

Latest: [month] [] [] [] AM

Categories:

Reference

Technical

Directional

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

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

New Entry

New Entry

Category: Reference

Label: 1 to 5 minutes

Medium: In Person

Notes:

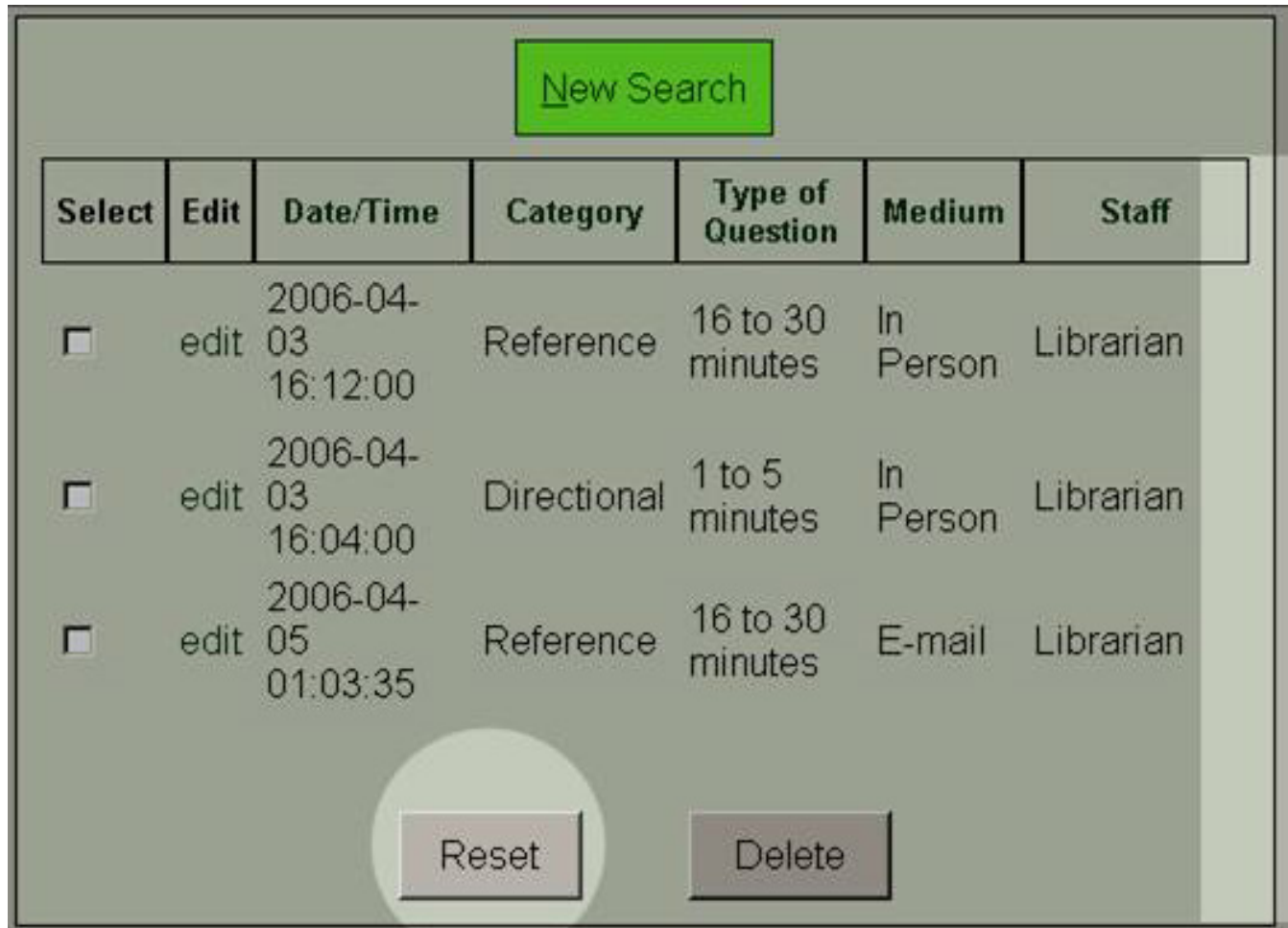
Referred to a Librarian/Unit for this question?

Record this question how many times? 1

Record Entry

A CS 5150 Project: Results

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



The screenshot shows a web interface for search results. At the top center is a green button labeled "New Search". Below it is a table with seven columns: "Select", "Edit", "Date/Time", "Category", "Type of Question", "Medium", and "Staff". The table contains three rows of results. Each row has a checkbox in the "Select" column and an "edit" link in the "Edit" column. At the bottom of the interface are two buttons: "Reset" and "Delete".

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

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