



# Evaluation & Experiments







# Learning Goals

- Understanding why we need evaluations and experimental user studies
- Get an overview about research methods in HCI
- What we learn through evaluations and experiments with users

# What is an Evaluation?

*“Evaluation is a process that critically examines a program. It involves collecting and analyzing information about a program’s activities, characteristics, and outcomes. Its purpose is to make judgments about a program, to improve its effectiveness, and/or to inform programming decisions” [1]*

- Improves system design and implementation
- Demonstrate the system’s impact
- Often reveal something new about people
- But how can we evaluate?

[1] Patton, M.Q. (1987). Qualitative Research Evaluation Methods. Thousand Oaks, CA: Sage Publishers.

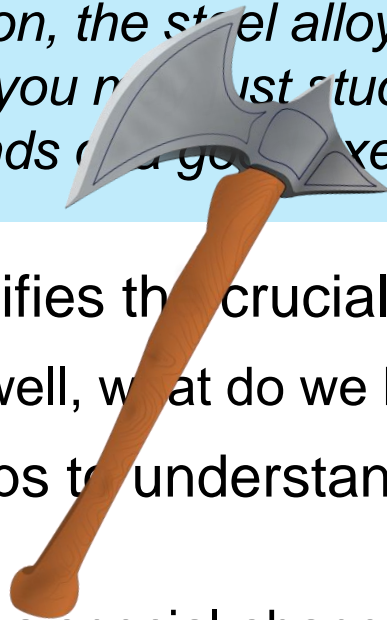
# Evaluation

## Criteria and Motivation

- Motivation (examples)
  - Developers are often not aware about the impact of their system
  - People are often not aware about how a system affects them
  - Developers see no need for evaluation when the system “works”, but there are always ways to improve it!
- Criteria (examples)
  - Informal comparison of two (or more) ideas against each other
  - Analysis of average performance using a representative user group (or in field usage)
  - Fulfillment of formal and informal usability metrics
  - Inform design

# Evaluating an Axe

*“If you want to evaluate a tool, say an axe, you might study the design of the bit, the weight distribution, the steel alloy used, the grade of hickory in the handle, etc., or you must study the kind and speed of the cuts it makes in the hands of a good axeman.” [1]*



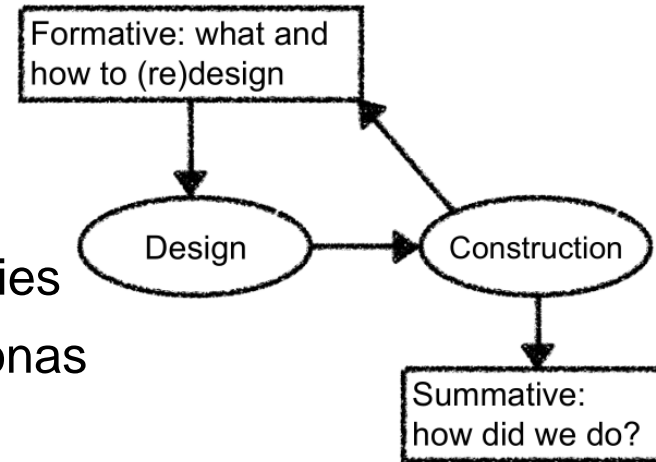
- Analytic evaluation identifies the crucial characteristics
  - “If the axe does not cut well, what do we have to change?”
- Empirical evaluation helps to understand the context for object properties
  - “Why does the axe have a special-shaped handle?”

[1] Scriven, M. (2013). Conceptual revolutions in evaluation. Evaluation roots, SAGE Publications, 167-179.

# Assessments

## Formative and Summative

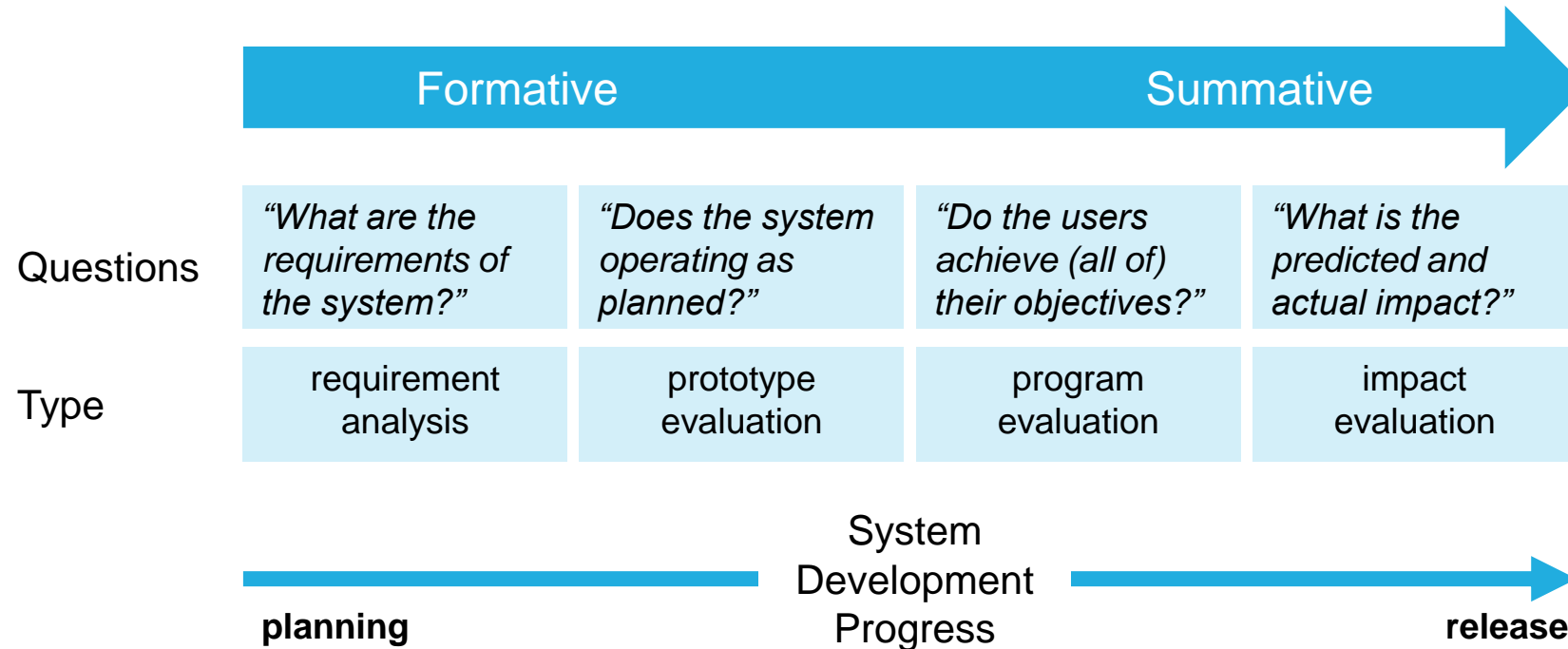
- Formative („How you design it?“)
  - Improvement-oriented
  - (Paper) prototyping, sketching
  - Interviews, surveys, diaries, case studies
  - Focus groups, user role playing, personas
- Summative („How good is it?“)
  - based on criteria
  - Ratings, questionnaires, scales
  - Efficiency measures
  - Physiological measures
- ...depends on the progress of the development!



[1] Scriven, M. (2012). Conceptual revolutions in evaluation. Evaluation roots, 167-179.

# System Evaluation Process

## Evaluations during Development





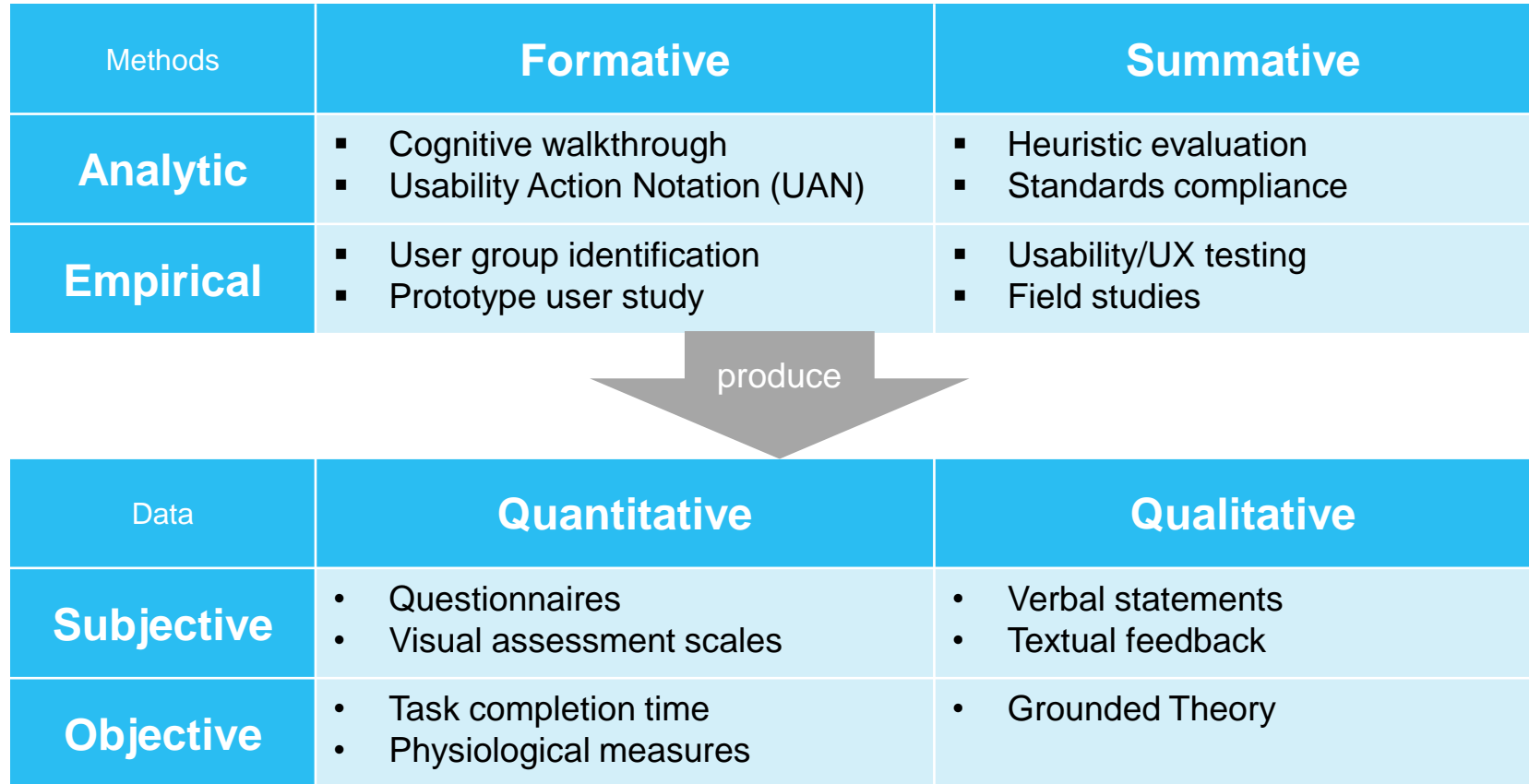
# Orthogonality of Approaches

Evaluation approaches include and exclude some others

Methods	Formative	Summative
<b>Analytic</b>	<ul style="list-style-type: none"><li>▪ Cognitive walkthrough</li><li>▪ Usability Action Notation (UAN)</li></ul>	<ul style="list-style-type: none"><li>▪ Heuristic evaluation</li><li>▪ Standards compliance</li></ul>
<b>Empirical</b>	<ul style="list-style-type: none"><li>▪ User group identification</li><li>▪ Prototype user study</li></ul>	<ul style="list-style-type: none"><li>▪ Usability/UX testing</li><li>▪ Field studies</li></ul>

# Orthogonality of Approaches

Evaluation approaches include and exclude some others

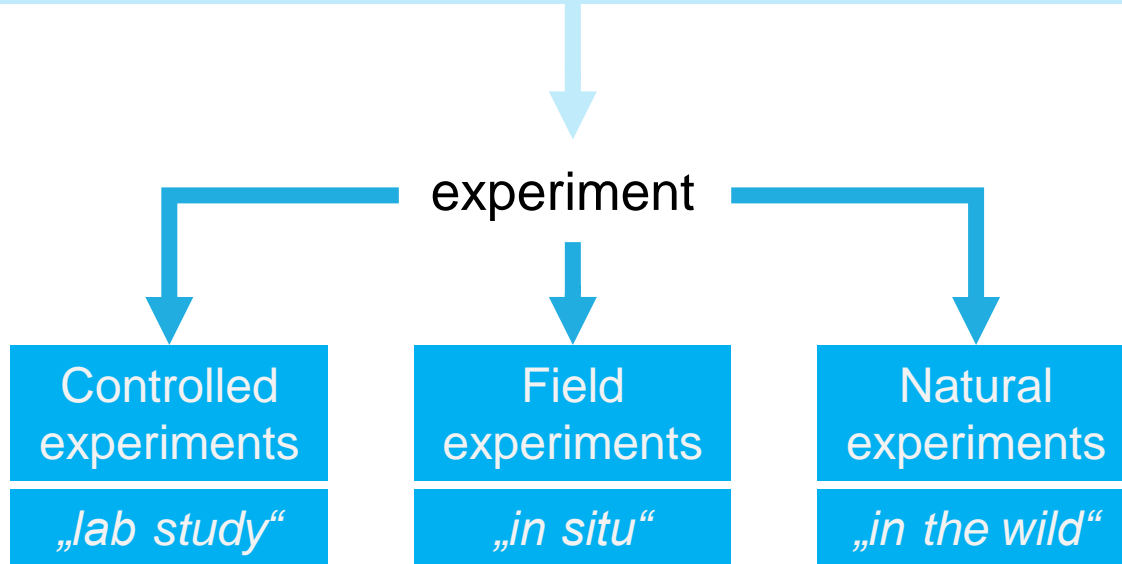


# Evaluation Types

- Process
  - Design (interfaces, accessibility, appeal, ...)
  - Implementation (functionality, usability, problems, ...)
- Data collection
  - Obstrusive (experiment, interview, ...)
  - Unobtrusive (published statistics, footprints, public domain, ...)

# Evaluation Types

*“an experiment is a methodical investigation for the empirical extraction of information (data)” [1] or “an ‘interrogation’ of nature” [2]*



[1] [en.wikipedia.org/wiki/Experiment](https://en.wikipedia.org/wiki/Experiment)

[2] Brockhaus Encyclopedia, 19th edition, Mannheim 1988, Keyword "experiment".







# Choosing Evaluation Types

- Empirical researchers found in experiments that greenish-yellow is the easiest color to see in all lighting conditions



Image: Darren C. from wikipedia.org (CC BY-SA 3.0)

...thus, perceptual research scientists recommended this one [1]



... but today mostly used are:

[1] Solomon, S. S. (1990). Lime-yellow color as related to reduction of serious fire apparatus accidents: The case for visibility in emergency vehicle accident avoidance. Journal of the American Optometric Association, 61, 827-831.

# Choosing Evaluation Types

- The preferred color for a firefighting vehicle depends highly on **emotions**, **individual needs**, and **traditions**
- Colors have a very subjective and emotional component
  - **Yellow** often evokes pleasant, cheerful feelings but men usually perceive yellow as a lighthearted or childish
  - **Red** is highly emotional and a subjective sign of „danger“, „energy“, „strength“, and „power“ – everything a firefighter stands for!
- Objective and subjective methods are required to understand which is the best color for a firefighting vehicle!



License to use Creative Commons Zero - CC0, pxfuel.com



# Outcome

- Evaluations...
  - ...are fundamental in HCI research
  - ...allow us to assess the impact of a system on users
  - ...allow us to improve a system
  - ...can be quantitative or qualitative (or both!)
  - ...must be carefully designed to draw no wrong conclusions!

# References

- Scriven, M. (2012). Conceptual revolutions in evaluation. *Evaluation roots*, 167-179.
- Field, A., & Hole, G. (2002). *How to design and report experiments*. Sage.
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.