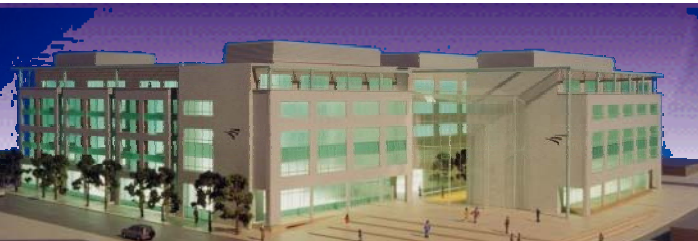


Problems and Pitfalls in Evaluating Adaptive Systems

Stephan Weibelzahl

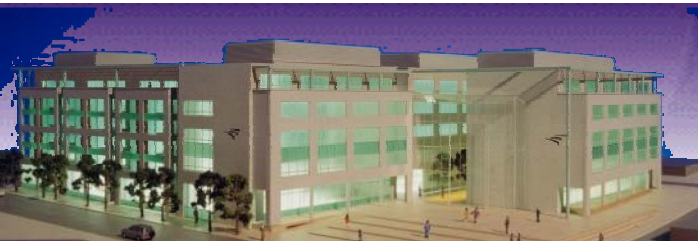
National College of Ireland

sweibelzahl@ncirl.ie



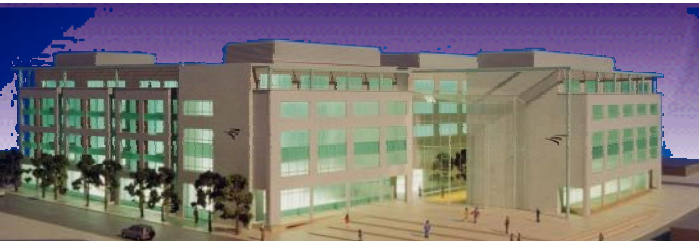
Overview

- Why is evaluation important?
- Which pitfalls should I try to avoid?
- Where can I find help?



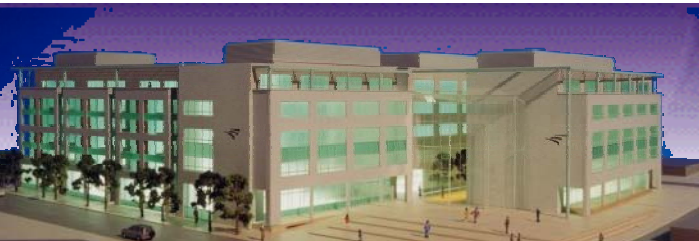
Why is evaluation important?

- **Find out whether it really works**
 - Effectiveness
 - Efficiency
 - Usability, user satisfaction
- **Detect inaccuracies and invalid assumptions**
- **Convince users, customers, investors**
- **Scientific advancement**



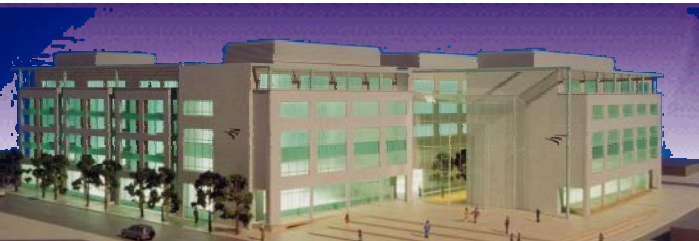
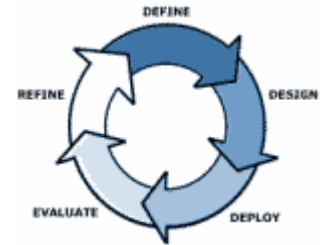
Which pitfalls should I try to avoid?

- **Pitfalls and problems often observed**
 1. Big evaluation study planned for the end of a project
 2. Not enough resources left
 3. Wrong control condition selected
 4. Too much variance in data
 5. Confusion which criterion to choose
 6. Users are unable to tell about adaptivity effects
 7. Evaluation results are reported incomplete or anecdotally



Pitfall 1: Big evaluation study planned for the end of a project

- **Summative evaluation cannot recover failures in earlier stages**
 - Conduct several formative studies
 - Distributed across the development cycle

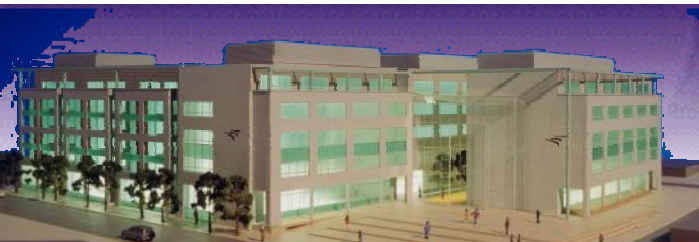


Pitfall 2: Not enough resources left



National
College of
Ireland

- **Empirical studies require personnel, organizational, and financial resources**
 - Spread studies across the development cycle
 - Expert evaluation
 - Evaluate inference mechanism with simulated users and empirical data
 - Use cognitive models



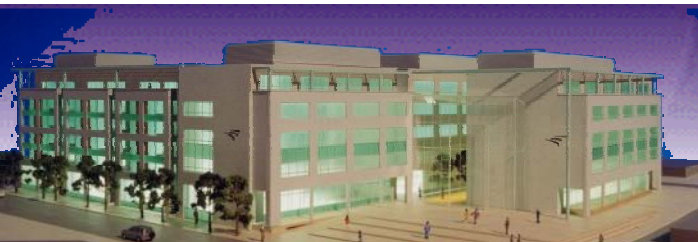
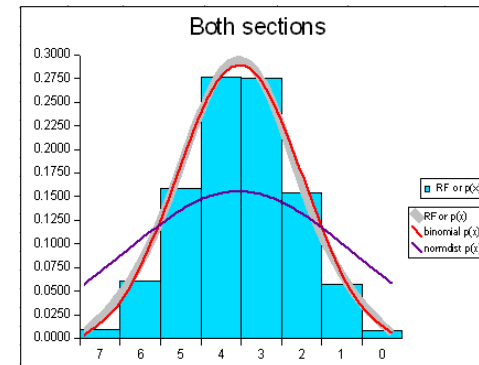
Pitfall 3: Wrong control condition selected

- **Switching off the adaptivity might result in an incomplete or even useless system**
 - Compare various adaptation decision conditions that are based on the same user characteristics



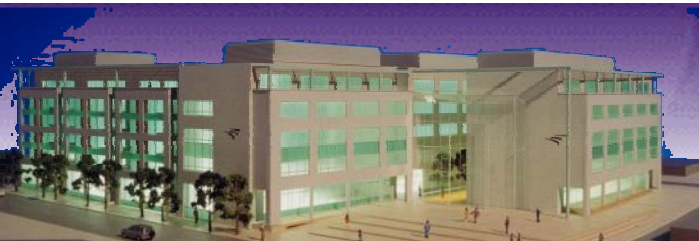
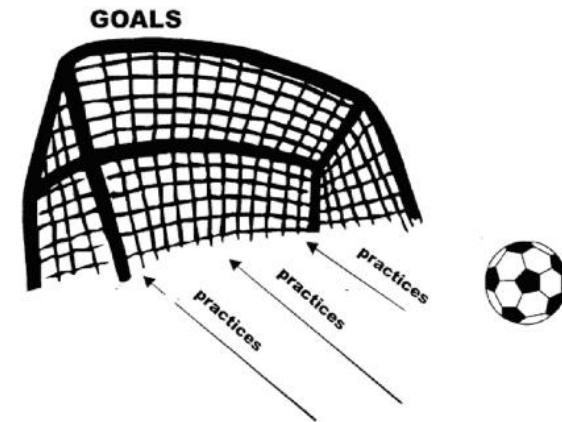
Pitfall 4: Too much variance in data

- **High variance corrupts statistical analysis**
 - Try to find a sample that is
 - heterogeneous in terms of the modeled user characteristics,
 - but homogeneous in terms of other factors
 - Use repeated measurement
 - Control variables that might have an impact on the results



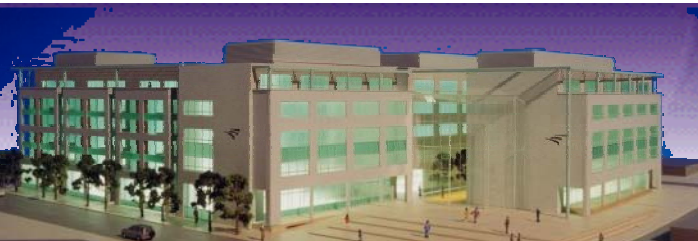
Pitfall 5: Confusion which criterion to choose

- There is no single evaluation criterion for adaptivity
 - Define goals of adaptivity precisely
 - Derive criteria from these goals



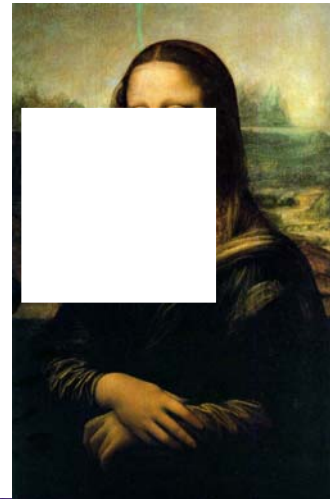
Pitfall 6: Users are unable to tell about adaptivity effects

- Users might not have noticed adaptivity at all
 - Use user feedback in combination with objective measures



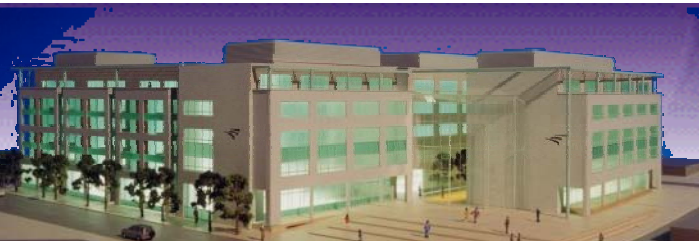
Pitfall 7: Results are reported incomplete or anecdotally

- **Incomplete report of results corrupts interpretation of study**
 - Guidelines on reporting statistical data
 - Include important information for adaptivity (e.g., empirically identified user characteristics, effect size)



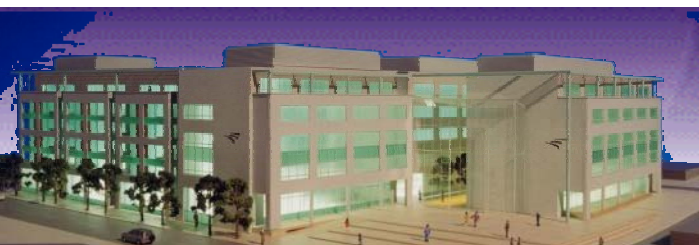
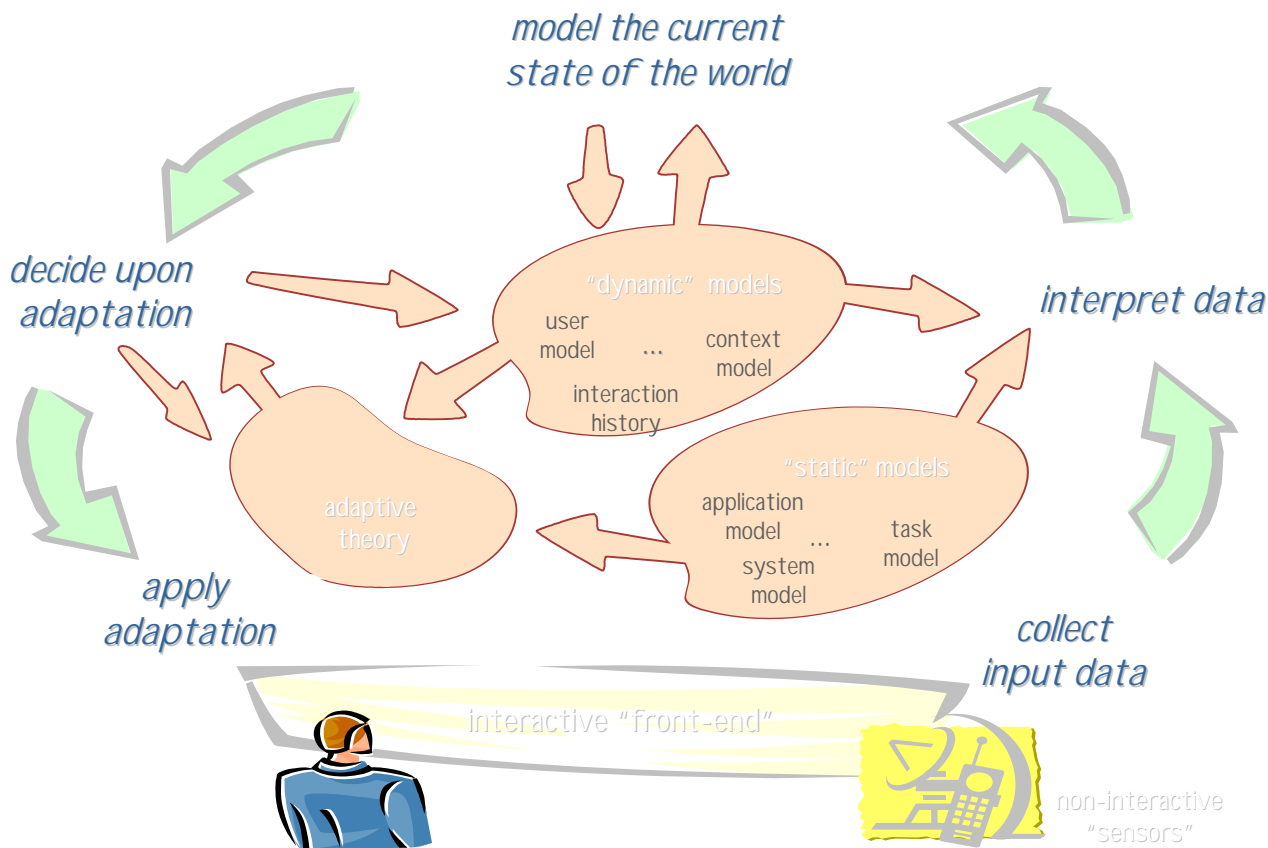
Recommendations

- **Plan carefully and in advance**
 - Sample
 - Control condition
 - Criteria
- **Slice (or dice) system: Layered Evaluation**
- **Publish results**





Layered Evaluation



Where can I find help?

- Ask an expert in empirical research methods before you conduct your study
- <http://www.easy-hub.org>
 - Previous workshops
 - Guidelines
 - Literature references

