

The Influence of Expectation and System Performance on User Satisfaction with Retrieval Systems

Katrin Lamm, Christa Womser-Hacker,
Thomas Mandl and Werner Greve
University of Hildesheim, Germany

Outline

- Interactive IR Studies
- Experimental Design
- Main Results
- Follow-up Study

Outline

- **Interactive IR Studies**
- Experimental Design
- Main Results
- Follow-up Study

User Studies in IR

Typical Questions

- How well do users perform with different systems?
- How satisfied are users with different systems?

Performance of Users

- Turpin & Hersh (2001)
 - TREC interactive track
 - User tests do not reflect system differences
- Scholer & Turpin (2008)
 - Relevance threshold in relation to system performance
 - Different users adopt different relevance criteria

Satisfaction of users

- Jansen et al. (2007)
 - Effect of branding
 - Correlation with perception
 - User expectations influence satisfaction
- Szajna & Scamell (1993)
 - User expectation of information systems
 - Correlation with perception
 - Effect wears off over time

Summary



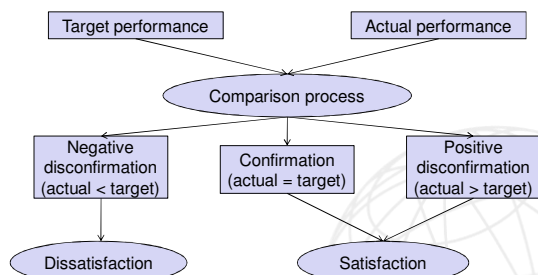
- Users compensate
- Relevance judgements depend on context
- Expectations affect satisfaction
- Expectations wear off over time

Outline



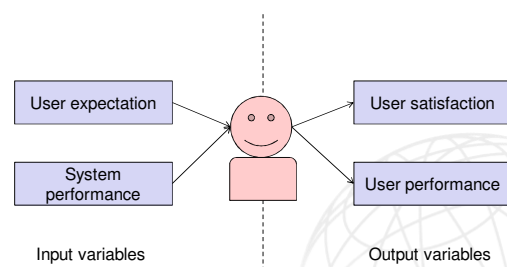
- Interactive IR Studies
- **Experimental Design**
- Main Results
- Follow-up Study

C/D Paradigm



[e.g. Homburg et al. 1999]

Research Model



Experimental Design



		System performance	
		good	bad
User expectation	low	Group 1	Group 2
	high	Group 3	Group 4

Experimental Procedure



- Instruction
 - Expectation manipulation
 - Test instructions
- Search
 - Three CLEF topics
 - 10 minutes per task
- Evaluation
 - User satisfaction questionnaire

Test System

The screenshot shows a search interface with a search bar containing 'Kühn: Kaffee-Schokolade'. Below the search bar, there are several search results displayed in a list format. The results include titles like 'Kühn: Kaffee-Schokolade', 'China: Weniger Kohle', 'Kernenergie: EIN GEWALTIGES FEUER', 'Heiß Kohl, wenn Sie in Berlin die erste solare Hauptstadt!', and 'Energie: Lächeln Sie die Zellen'. Each result includes a short summary and the source 'Quelle: Der Spiegel'.

EVIA 2010 13

Test System

The screenshot shows a search interface with a search bar containing 'China: Weniger Kohle'. Below the search bar, there are several search results displayed in a list format. The results include titles like 'Kühn: Kaffee-Schokolade', 'China: Weniger Kohle', 'Kernenergie: EIN GEWALTIGES FEUER', 'Heiß Kohl, wenn Sie in Berlin die erste solare Hauptstadt!', and 'Energie: Lächeln Sie die Zellen'. Each result includes a short summary and the source 'Quelle: Der Spiegel'.

EVIA 2010 14

Outline

- Interactive IR Studies
- Experimental Design
- **Main Results**
- Follow-up Study

EVIA 2010 15

Analysis

- Sample
 - 89 female students
 - Test language German
- Investigation of differences by ANOVA
 - User satisfaction questionnaire
 - Direct and indirect items
 - User performance measures
 - Completeness and accuracy of results

EVIA 2010 16

User Performance Measures

- User recall
 - Documents correctly identified as relevant
 - Relevant documents in result list
- User precision
 - Documents correctly identified as relevant
 - Documents saved as relevant by user

EVIA 2010 17

Overview of Results

- User expectation
 - No significant differences
 - Predictions of C/D paradigm apparent
- System performance
 - User satisfaction
 - Significant differences for precision items
 - User performance
 - Compensatory behavior for user recall
 - Adaptive behavior for user precision

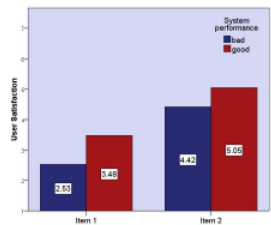
EVIA 2010 18

User Satisfaction



Significant differences for precision items (7-point scale)

- Item 1: The filtering of articles could have been better. ($p = 0.008$)
- Item 2: Most articles have been relevant with respect of the queries. ($p = 0.025$)

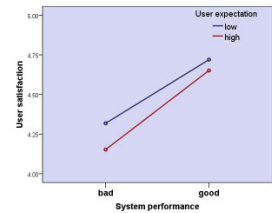


C/D Paradigm



Predictions of C/D paradigm apparent (combined scale, Cronbach's Alpha 0.69)

- No significant differences for user expectations ($p = 0.50$)
- Significant differences for system performance ($p = 0.01$)



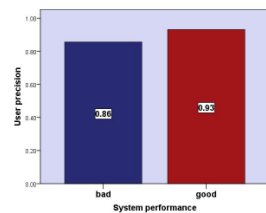
User Performance



Significant differences for user precision

- User precision on average higher for better system (0.86 vs. 0.93)
- 8% difference

No user compensation?

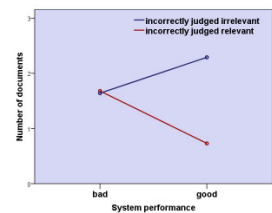


User Adaptation



Adaptive behavior for user precision Comparison of incorrectly judged documents

- Average number of documents incorrectly judged irrelevant
- Average number of documents incorrectly judged relevant



Outline



- Interactive IR Studies
- Experimental Design
- Main Results
- Follow-up Study

Follow-up Study



- Similarities
 - C/D paradigm as framework
 - Input and output variables
- Differences
 - Comparison of two systems
 - Server-based testing
 - Web corpus
 - Iterative search behavior

Selected Results



- User performance
 - Compensation for recall
 - Adaptation of relevance criteria for precision
- User satisfaction
 - Task 1 significant differences for expectation
 - Task 2 significant differences for system
 - C/D paradigm not apparent

Conclusion



- Relevance judgements are context dependent
- Users can compensate differences in system performance
- Expectations tend to wear off over time
- Results highlight need to consider expectations

Outlook



- Further elaborate the concept of user expectations
- Future research should establish reliable methods to measure user satisfaction
- Development of an instrument to measure user expectations

- Lamm, K., Mandl, T., Womser-Hacker, C. and Greve, W., "The Influence of Expectation and System Performance on User Satisfaction with Retrieval Systems", Proc. International Workshop on Evaluating Information Access (EVIA) '10 (to appear)
- Lamm, K., Mandl, T., Womser-Hacker, C. and Greve, W., "User Experiments with Search Services: Methodological Challenges for Measuring the Perceived Quality", Proc. International Workshop on Perceptual Quality of Systems (PQS) '10 (to appear)

Thank you for your attention!

References



- Homburg, C.; Giering, A. and Hentschel, F. (1999): Der Zusammenhang zwischen Kundenzufriedenheit und Kundenbindung. In: Bruhn, M.; Homburg, C. (Hrsg.): Handbuch Kundenbindungsmanagement: Grundlagen, Konzepte, Erfahrungen. Wiesbaden: Gabler, 81-112.
- Jansen, B. J.; Zhang, M. and Zhang Y. (2007): The effect of brand awareness on the evaluation of search engine results. In: Proc. CHI '07, 2471-2476.
- F. Scholer and A. Turpin (2008): Relevance Thresholds in System Evaluations. In: Proc. SIGIR '08, 693-694.
- Szajna, B. and Scamell, R. W. (1993): The Effects of Information System User Expectations on Their Performance and Perceptions. MIS Quarterly, 17(4): 493-525.
- A. H. Turpin and W. Hersh (2001): Why Batch and User Evaluations Do Not Give the Same Results. In: Proc. SIGIR '01, 225-231.