



An Evaluation of the Inference Mechanism of an Adaptive Help System

Dorothea Iglezakis

Department of Applied Computer Science

Cath. University of Eichstätt-Ingolstadt

CHeetah – An extensible adaptive help system for web-based applications



```
GET page1?para=1&para2=2  
GET page2?para=2&para2=2
```

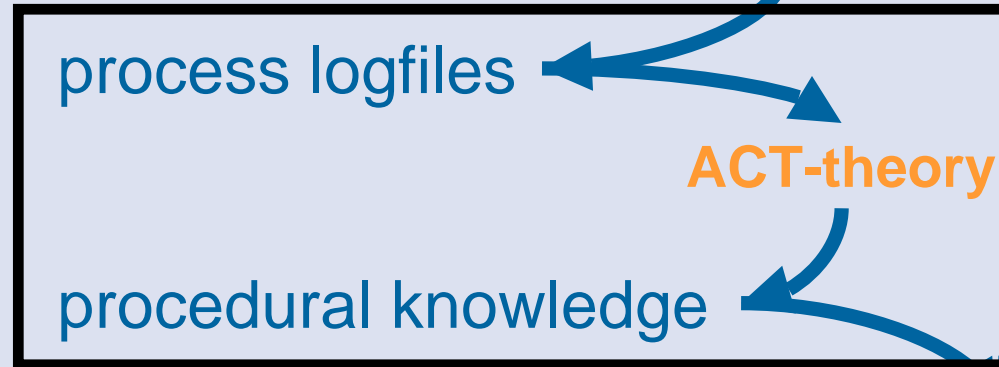
UID	PID	start	end
1	1	20040302	20040302

UID	task	activation
1	1	0.35



standard logfiles

process recognition



frame-based adaptation

adapted help contents

Evaluation of Adaptive Systems

CHEetah – ACT-theory (Anderson and Lebiere, 1998)



- One of the best empirically founded theories about knowledge and forgetting
- The *strength accumulation formula* models the knowledge activation using the practice of knowledge items:

$$a_i^z(T) = \sum_{j=1}^n t_j^{-d}$$

- a_i^z : activation of knowledge item i at time z
- t_j : time passed since j -th practice of item i
- d : domain-dependent parameter $]0;1[$

Evaluation – Framework



- Framework for evaluations of adaptive systems by Weibelzahl (2003):
 - Evaluation of the input data
 - Evaluation of the inference mechanism
 - Evaluation of adaptation decisions
 - Evaluation of total interaction

Evaluation – Framework



- Framework for evaluations of adaptive systems by Weibelzahl (2003):
 - Evaluation of the input data
 - **Evaluation of the inference mechanism**
 - Evaluation of adaptation decisions
 - Evaluation of total interaction

Evaluation



Questions:

- Is the ACT-value a reliable and valid **estimate** of procedural knowledge?
- Does the ACT-value correspond with **subjective and objective measures** of the knowledge?
- Is the ACT-value a better estimate of procedural knowledge than **general computer experience**?

Evaluation – Participants



- 16 employees of the hmd-software AG (6 m, 10 f)
- Usage of the target software WebTime logged over 11 months

Measure	<i>M</i>	<i>SD</i>	Min	Max
subj. performance NA	3.23	1.54	0	5
obj. performance NA	3.00	0.61	1	4
ACT value task NA	0.89	1.19	0.02	6.12
subj. performance NT	2.20	1.62	0	4
obj. performance NT	2.89	0.80	1	4
ACT value NT	0.02	0.04	0	0.17
$comp_h$	0.51	1.02	-1.25	3
$comp_t$	4.26	1.91	1	6

Evaluation – Procedure



Web-Questionnaire



Evaluation – Procedure



Web-Questionnaire



Evaluation – Operationalization



Procedural Knowledge

Subjective self-assessment on a six-step rating scale

Objective rating of quality and time of performance

General computer experience

Scale of helplessness with the computer

Gutmann-scale of activities with the computer

ACT-value

Observation of the actions in the target software over 11 months

Computation of ACT-values for two different tasks

Dependent variables

Independent variables

Evaluation of Adaptive Systems

Evaluation – Results for task NewAppointment



Correlation between	r	Significance
ACT — subjective expertise	0.19	n.s.
ACT — objective performance	0.09	n.s.
ACT — time	-0.06	n.s.
<i>Comp_h</i> — subjective expertise	-0.32	n.s.
<i>Comp_h</i> — objective performance	-0.59	$p < 0.01$
<i>Comp_h</i> — time	-0.27	n.s.
<i>Comp_t</i> — subjective expertise	-0.27	n.s.
<i>Comp_t</i> — objective performance	-0.48	$p < 0.05$
<i>Comp_t</i> — time	-0.27	n.s.

Evaluation – Results for task NewTodo



Correlation between	r	significance
ACT — subjective expertise	0.66	$p < 0.05$
ACT — objective expertise	0.30	n.s.
ACT — time	-0.23	n.s.
<i>Comp_h</i> — subjective expertise	-0.25	n.s.
<i>Comp_h</i> — objective expertise	-0.15	n.s.
<i>Comp_h</i> — time	0.34	n.s.
<i>Comp_t</i> — subjective expertise	-0.36	n.s.
<i>Comp_t</i> — objective expertise	-0.16	n.s.
<i>Comp_t</i> — time	0.30	n.s.

Evaluation – Conclusions



- all correlations between ACT-values and performance measures are in the right direction
- ACT-measure not reliable for very easy tasks
⇒ ceiling-effect
- ACT-measure better for subjective than for objective measures
- general computer experience is no valid estimate for procedural knowledge

Further Work



- not necessary • Evaluation of the input data ✗
- topic of this talk • **Evaluation of the inference mechanism** ✓
- already done • Evaluation of adaptation decisions ✓
- in progress • Evaluation of total interaction ✓

Evaluation – Discussion



- long-term participants not easy to find
 - participants used only a very small subset of tasks ⇒ restricted choice of tasks
 - general computer experience and procedural knowledge were negatively correlated
- ⇒ Controlled study with selected participants