ADAPTIVE LEARNING ENVIRONMENTS:
Summative Evaluation
Contents

1. Using Experiments for System Design and Evaluation
2. Evaluating the Design and Effectiveness of a Maths Tutoring System
3. Summative evaluation of Standup
4. Writing up Experiments and Empirical Studies
5. References

Some material based on Ainsworth’s AIED 2003 tutorial on Evaluation Methods for Learning Environments, see AILE course web page and link:
http://www.psychology.nottingham.ac.uk/staff/sea/Evaluationtutorial.ppt
1. Design of Experiments
Role of Experiment in Design

Often experiments are used to guide new designs or the help understand existing design

**Programs are not themselves experiments** but are part of the basis for conducting experiments (on an algorithm or a system or a group of people)

Three types of activity:

**Exploratory:** where we are wondering what to design
**Formative Evaluation:** we experiment with a preliminary design with the aim of building a better one
**Summative Evaluation:** where a final design is analysed definitively
Formative v. Summative Evaluation

Formative Evaluation:
- iterative, throughout design and implementation
- test preliminary designs for usability etc
- assessing impact of changes
- make decisions about later project stages
- frequently qualitative

Summative Evaluation:
- on completion of each stage
- assessing effectiveness
- frequently quantitative
Qualitative v. Quantitative Data

**Qualitative**

Descriptive data
Subjective

Based on system behaviour or user experience

Obtained from observation, questionnaires, interviews, protocol analysis, heuristic evaluation, cognitive and post task walkthrough

**Quantitative**

Numerical data
Objective

Based on measures of variables relevant to performance or user experience

Obtained from empirical studies, e.g. experiments, also questionnaires, interviews

Amenable to statistical analysis
Systems and Experiments

When we talk about **experiments**, generally talking about...
- stating specific hypotheses
- identifying and manipulating variables
- systematic procedures to TEST our hypotheses
- some degree of control (often limited in “real world” settings)

**Not all studies we do in ALE are “experiments” in a strict sense**

- May do a survey about how system was used in class
- May observe participants using a system
- May mine data afterwards doing post hoc analysis, looking for general patterns

Also, difference between a **true experiment** with randomised group assignment and so forth, versus a **quasi-experiment**, where less control, may not be able to randomly assign groups, etc.
Typical Questions

Having gone through a number of iterations of formative evaluation, you think that the system is finally ready. You need to see now how well it works....

 Does it do what it was claimed it would do?
 Is it effective?

Such questions need to be made more precise.

A number of methods can be used, e.g.
• an experimental set-up with alternative versions of the tool - perhaps without a crucial feature
• a control group for comparison.

Methodology has to be tight for strong claims to be made.
Common Measures (Dependent Variables) *(from Ainsworth, 2003)*

Learning gains  
Post-test – Pre-test

Learning efficiency  
i.e. does it reduce time spent learning

How the system is used in practice (and by whom)  
ILEs cannot help if learners do not use them!  
What features are used

User attitudes

Cost savings

Teachbacks  
How well can learners now teach what they have learnt
Prototypical designs *(Ainsworth, 2003)*

1. (intervention) post-test
2. Pre – (intervention) - post-test
3. Pre – (intervention) - post-test – delayed post-test
4.Interrupted time-series
5. Cross-over

*Look at Ainsworth (2003) tutorial for examples of these (see web page)*
Nature of Comparison *(Ainsworth, 2003)*

1. ILE alone
2. ILE v non-interventional control
3. ILE v Classroom
4. $\text{ILE}_{(a)}$ v $\text{ILE}_{(b)}$ (within system)
5. ILE v Ablated ILE
6. Mixed models

*Again, see Ainsworth (2003) tutorial for examples of these (see web page)*
ILE alone *(Ainsworth, 2003)*

**Examples**
- Smithtown — Shute & Glaser (1990)
- Cox & Brna (1995) SWITCHER
- Van Labeke & Ainsworth (2002) DEMIST

**Uses**
- Does something about the learner or the system predict learning outcomes? e.g.
  - Do learners with high or low prior knowledge benefit more?
  - Does reading help messages lead to better performance?

**Disadvantages**
- No comparative data — is this is good way of teaching??
- Identifying key variables to measure
ILE v non-interventional control
(Ainsworth, 2003)

Examples

Uses
– Is this a better way of teaching something than not teaching it at all?
– Rules out improvement due to repeated testing

Disadvantages
– Often a no-brainer!
– Does not answer what features of the system lead to learning
– Ethical?
ILE v Classroom (Ainsworth, 2003)

Examples
- LISPITS (Anderson & Corbett)
- Smithtown (Shute & Glaser, 1990)
- Sherlock (Lesgold et al, 1993)
- PAT (Koedinger et al, 1997)
- ISIS (Meyer et al, 1999)

Uses
- Proof of concept
- Real world validity

Disadvantages
- Classrooms and ILEs differ in some many ways, what can we truly conclude?
ILE \textsuperscript{(a)} v ILE \textsuperscript{(b)} \textit{(within system)} \textit{(Ainsworth, 2003)}

Examples
\begin{itemize}
  \item PACT – Aleven et al (1999)
  \item CENTS – Ainsworth et al (2002)
  \item Galapagos – Lucken et al (2001)
  \item Animal Watch – Arroyo et al (1999, 2000)
\end{itemize}

Uses
\begin{itemize}
  \item Much tauter design, e.g. nullifies Hawthorne effect
  \item Identifies what key system components add to learning
  \item Aptitude by treatment interactions
\end{itemize}

Disadvantages
\begin{itemize}
  \item Identifying key features to vary – could be very time consuming!
\end{itemize}
ILE v Ablated ILE *(Ainsworth, 2003)*

Ablation experiments remove particular design features and performance of the systems compared

**Examples**
- VCR Tutor – Mark & Greer (1995)
- Luckin & du Boulay (1999)

**Uses**
- What is the added benefit of AI

**Disadvantages**
- System may not be modular
**Context (Ainsworth, 2003)**

- (a) Expt in Laboratory with experimental subjects
- (b) Expt in Laboratory with ‘real’ subjects
- (c) Expt in ‘real’ environment with ‘real’ subjects
- (d) Quasi-experiment in ‘real’ environment with ‘real’ subjects
- (e) For Real!
## Learning Gains: Effect Size *(Ainsworth, 2003)*

\[(\text{Gain in Exp Condtn} - \text{Gain in Control})/\text{St Dev in Control}\]

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Ratio</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom teaching v Expert Tutoring</td>
<td>1:30 v 1:1</td>
<td>2 sd</td>
</tr>
<tr>
<td>Classroom teaching v Non Expert Tutoring</td>
<td>1:30 v 1:1</td>
<td>0.4 sd</td>
</tr>
<tr>
<td>Classroom teaching v Computer Tutoring</td>
<td>1:30 v C:1</td>
<td>?</td>
</tr>
</tbody>
</table>

A 2 sigma effect means that 98% of students receiving expert tutoring are likely to do better than students receiving classroom instruction.
Choosing Between Designs
(Ainsworth, 2003)

Validity

Construct validity
  Is it measuring what it’s supposed to?

External validity
  Is it valid for this population?

Ecological validity
  Is it representative of the context?

Reliability

Would the same test produce the same results if:
  – Tested by someone else?
  – Tested in a different context?
  – Tested at a different time?
Some issues and problems

Natural environment v ability to control variables
  e.g. test in classroom v. bring into laboratory

Interference with participants - ethical issues
  * Should you use a method of teaching that you do not think is going to work on your participants?
  * Should everyone get the opportunity to use the best approach?
  * Will getting poor scores on a test that is not relevant to the curriculum affect student's morale and consequently their other work?
  * Should you use teaching time to do experiments?

Problems of measurement:
  * What is improvement?
  * How long does it last?
  * Does it generalise?
2. Evaluating the Design and Effectiveness of a Maths Tutoring System
Maths Tutoring System Example

Goal: *intelligent computer tutor for university maths students to practice calculus*

- How do human tutors teach calculus?
- What can we infer from human tutors behaviour to inform tutor design?
- What is feasible to incorporate in system and what not?

Questions we might consider to inform design:
1. What errors do students typically make?
2. What should the system do when students make errors?
Methods for collecting maths errors

<table>
<thead>
<tr>
<th>Task analysis</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Walkthrough</td>
<td>Mock-ups</td>
</tr>
<tr>
<td>Protocol analysis</td>
<td>Wizard of Oz</td>
</tr>
<tr>
<td>Video Recording</td>
<td>Interview</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Focus groups</td>
</tr>
<tr>
<td>Sensitivity Analysis</td>
<td>Expert evaluation</td>
</tr>
<tr>
<td>Post-hoc analysis</td>
<td>Logging use</td>
</tr>
<tr>
<td>Dialogue mark-up and analysis</td>
<td>Sentient analysis</td>
</tr>
<tr>
<td>Manipulation experiment</td>
<td></td>
</tr>
<tr>
<td>Self Report</td>
<td></td>
</tr>
</tbody>
</table>
What errors do students typically make?

1. **Interview** teachers about errors that target users frequently make *(error types and examples)*

2. Devise a **set of test calculus examples**

3. Give target user group test set and observe, collect log of their **interaction** *(example errors)*

4. **Analyse** results to see most frequent errors

5. Give **questionnaire** to teachers with example errors and ask what feedback they would give *(feedback types in relation to each error)*

6. **Observe** tutor teaching student through chat interface + **record interaction** *(example errors)*

7. **Analyse interaction** in relation to student errors and actions taken by teacher *(feedback types)*

8. **Cognitive walkthrough** by tutor *(when feedback type given and general feedback strategies)*
What should the system do when students make errors?

Using these methods you find that human tutors usually use one of the following feedback options:

1. *give feedback immediately*
2. *just flag to the student that they have made an error*
3. *let the student realise they have made a mistake and ask for help*

You want to see which works best...

*Do some experiments with the tutoring system, with some students.....*

*[Based loosely on a experimental study described in Corbett, A.T. and Anderson, J.R., 1990]*
Other Evaluation Questions...

Does interface A to the Maths tutor work better than interface B?

Does student enjoyment correlate with learning?
Does student enjoyment correlate with learning?

Assessing student enjoyment - affective measures:
- Observe facial expressions
- Self-report of enjoyment: sliders
- Questionnaire
- Verbal Protocol
- Expert observation

Assessing Learning - performance measures:
- Number of errors
- Time to learn to mastery
- Amount of materials covered in set time
Does interface A to the Maths tutor work better than interface B?

Could use **various methods:**

– Questionnaire
– Observation
– Interviews
– Logging use
– ...

but considering **experimental methods** here.....
General Experimental Design: Overview

1. Testing Hypotheses
2. Experimental Design
3. Method
   - Participants
   - Materials
   - Procedure
4. Results
5. Discussion and Conclusions
Testing Hypotheses

"Immediate Feedback is best!"

*Hard to test - we need to be more specific*

"Differences in performance on a specific test will be shown between students given no feedback and students given immediate feedback."

= *the experimental hypothesis*

"There will be no difference in performance shown by students given immediate feedback or no feedback."

= *the null hypothesis*
Possible Variables

* Whether or not feedback is given
* **When it is given** -- immediately? after 3 errors of same type? after certain types of errors? at the end of session?
* **What is given as feedback** -- correct or incorrect; detailed explanation; further examples
* **How much control** does student have over feedback?
* **How long does the student take** to complete a task?
* What is the student's **level of performance**?
* **How does the student feel** about different types of feedback -- which do they prefer? Which do they feel they learn most from? Which helps them learn most quickly?
* **How good are students at estimating** their performance on a task?
Experimental Design

Experimental conditions:
1. immediate error feedback and correction
2. immediate error flagging but no correction
3. feedback on demand

Control condition: to eliminate alternative explanations of the data obtained
- no feedback
Experimental Variables

Independent Variable - manipulated by experimenter
Dependent Variable - not manipulated, but look to see if manipulating the independent variable has an effect on it (but not necessarily a causal relationship)

Independent Variable: *type of feedback*
Dependent variable: *time to complete the exercises; post-test performance*

*Keep what is taught constant, so all learners cover the same material*

Other factors are *Extraneous Variables* - things that vary without our wanting them to...
### Results: Test Scores and Completion Time
*(from Corbett and Anderson, 1990)*

Mean post-test scores (% correct) and Mean Exercise Completion Times (minutes) for 4 versions of the tutor.

<table>
<thead>
<tr>
<th></th>
<th>Immediate feedback</th>
<th>Error flagging</th>
<th>Demand feedback</th>
<th>No tutor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-test Scores</td>
<td>55%</td>
<td>75%</td>
<td>75%</td>
<td>70%</td>
</tr>
<tr>
<td>Exercise Times</td>
<td>4.6</td>
<td>3.9</td>
<td>4.5</td>
<td>4.5</td>
</tr>
</tbody>
</table>

- We could then compare the sets of scores across the different feedback conditions.
# Results: Table 3 from Corbett and Anderson, 1990

## Questionnaire 1 Mean Ratings

<table>
<thead>
<tr>
<th>Question</th>
<th>Immediate Feedback</th>
<th>Error Flag</th>
<th>Demand Feedback</th>
<th>No Tutor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How difficult were the exercises? (1 = easy, 7 = challenging)</td>
<td>4.1</td>
<td>3.9</td>
<td>3.4</td>
<td>2.8</td>
</tr>
<tr>
<td>2. How well did you learn the material? (1 = not well, 7 = very well)</td>
<td>5.4</td>
<td>4.6</td>
<td>5.4</td>
<td>5.8</td>
</tr>
<tr>
<td>3. How much did you like the tutors? (1 = disliked, 7 = liked)</td>
<td>5.2</td>
<td>4.5</td>
<td>4.8</td>
<td>4.9</td>
</tr>
<tr>
<td>4. Did the tutor help you finish more quickly? (1 = slower, 7 = faster)</td>
<td>5.1</td>
<td>4.6</td>
<td>4.7</td>
<td>4.5</td>
</tr>
<tr>
<td>5. Did the tutor help you understand better? (1 = interfered, 7 = helped)</td>
<td>5.3</td>
<td>4.9</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>6. Did you like the tutor’s assistance? (1 = disliked, 7 = liked)</td>
<td>5.3</td>
<td>5.0</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>7. Would you like more or less assistance? (1 = less, 7 = more)</td>
<td>4.3</td>
<td>4.9</td>
<td>4.5</td>
<td>4.6</td>
</tr>
</tbody>
</table>
Discussion and Conclusions

The effect of tutor type, as measured by post-test scores and mean exercise completion times, is not statistically significant.

- So there would be no evidence in this case that feedback manipulation affected learning.

There were no significant differences among the four groups in rating:

* how much they liked working with the tutor
* how much help the tutor was in completing the exercises
* how well they liked the tutor's assistance
* whether they would prefer more or less assistance
Correlational design

If this study had showed that immediate feedback was best, we might want to follow it up by looking at the relationship between:

* performance on later maths tests
* the amount of time spent using the tutor over the year

Does spending more time on the tutor correlate well with best performance on later tests?

*Warning: correlation is not causation*

e.g. if it doesn't rain, reservoirs dry out
if it doesn't rain, people stop using umbrellas

..... So using umbrellas stops reservoirs drying out? (NO)

*A correlation between use of umbrellas and dry reservoirs is likely, but one does not cause the other.*
3. Summative evaluation of Standup
Evaluation with children with CCN

Week

1  2  3  4  5  6  7  8  9

Formal testing

Baseline Phase

Introduction Phase

Intervention phase

Evaluation Phase

Post intervention

Video observation

Base

Intro

Intervention

Evaluation

Post
The evaluation study

1. 9 participants from independent special school
2. 14 sessions c. 30 minutes over 9 weeks (April/May/June),
3. Consent obtained from parents and children
4. Pre-testing with standardised tests
5. Children shown how to use the software weeks 1 and 2
6. Intervention period exploring software weeks 3 to 6
7. Level of support and guidance reduced, and task complexity increased, as sessions went on
8. Use of system video-recorded for study
9. Favourite jokes stored in paper folder and on AAC devices
10. Evaluation period weeks 7 and 8
11. Further standardised testing
12. Structured interviews and questionnaires for feedback from staff and parents
13. Talking mats to collect feedback from children

Use with typically-developing children March/April 2007
<table>
<thead>
<tr>
<th>Level</th>
<th>Participant</th>
<th>Communication</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early primary</td>
<td>S1, female; age: 8y4m</td>
<td>Dynavox DV4 user: PCS</td>
<td>Head switch</td>
</tr>
<tr>
<td>Middle primary</td>
<td>S2, female; age: 10y10m</td>
<td>Intelligible speech: poor articulation</td>
<td>Direct</td>
</tr>
<tr>
<td>Middle primary</td>
<td>S3, female; age: 10y9m</td>
<td>Communication book: gross fist &amp; eye gaze</td>
<td>Head switch</td>
</tr>
<tr>
<td>Middle primary</td>
<td>S4, male; age: 10y3m</td>
<td>Communication Board: PCS, TechSpeak</td>
<td>Direct</td>
</tr>
<tr>
<td>Middle primary</td>
<td>S5, male; age: 10y3m</td>
<td>Clear speech</td>
<td>Direct</td>
</tr>
<tr>
<td>Senior primary</td>
<td>S6, male; age: 11y3m</td>
<td>Dynavox DV4 user: PCS</td>
<td>Head switch</td>
</tr>
<tr>
<td>Senior primary</td>
<td>S7, male; age: 12y9m</td>
<td>Speech: poor intelligibility uses PCS</td>
<td>Head switch</td>
</tr>
<tr>
<td>Senior primary</td>
<td>S8, male; age: 11y10m</td>
<td>Dynavox DV4 user: PCS</td>
<td>Direct</td>
</tr>
<tr>
<td>Senior primary</td>
<td>S9, female; age: 11y3m</td>
<td>Intelligible speech</td>
<td>Direct</td>
</tr>
</tbody>
</table>
Evaluation Instruments

CELF Clinical Evaluation of Language Fundamentals (Semel, Wiig, Secord, 1995)

- **CELF Linguistic concepts** (participants are asked to point to...: “the blue line”, “the line that is not yellow”; participants must point to a stop sign if they think they cannot do what they are asked to do.)
- **CELF Sentence structure** (e.g. show me...: “The girl is not climbing”, “The dog that is wearing a collar is eating a bone”)
- **CELF Oral directions** (e.g. point to...: “The black circle”, “The last white triangle and the first black square”)
- **CELF Word classes** (participants choose two related items from a set of four, e.g. “girl boy car table”, “slow nurse doctor rain”)

PIPA Preschool and primary inventory of phonological awareness (Frederickson, Frith and Reason, 1997)
EM tells AL one of ‘her’ jokes Week 3 (intervention)
NI exploring to get ‘any joke’ Week 8 (evaluation)
Results

Videos transcribed, annotated and analysed:
- Determine task achievement, degree of participant’s initiation, response and anticipation
- Good inter-rater reliability
- Transcripts and interview also coded by SLTs

All children benefited
- nearly all able to locate name; exit program; generate and tell, and store and retrieve jokes by end of study
- some participants in exploring system discovered different ways to accomplish tasks and worked out shortcuts
- all gave feedback using talking mats
- reported increase in self-confidence and maturity in all
- carry-over to day-to-day use of AAC
- participants distinguished between generating and telling joke
- joke folders used to tell jokes to others
- jokes liked even when poor
## Task Difficulty: progress

<table>
<thead>
<tr>
<th>Description</th>
<th>Train</th>
<th>Inter</th>
<th>Eval</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 A2</td>
<td>Find name (log onto the system)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>End program (log off from the system)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 B2</td>
<td>Generate any joke from new jokes</td>
<td>P1,3,7,8,9, P5</td>
<td>P5</td>
</tr>
<tr>
<td></td>
<td>Speak a joke using speech synthesis</td>
<td>P2,4,6</td>
<td>P7,8</td>
</tr>
<tr>
<td>B3 B4</td>
<td>Save a joke to favourites</td>
<td></td>
<td>P8</td>
</tr>
<tr>
<td></td>
<td>Choose a joke from favourite s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>Generate a joke on specified topic (e.g. about an animal)</td>
<td></td>
<td>P3</td>
</tr>
<tr>
<td>C2</td>
<td>Generate a joke on a specified sub topic (e.g. about a wild animal)</td>
<td></td>
<td>P9</td>
</tr>
<tr>
<td>C3</td>
<td>Choose a joke from old joke collection not saved to favourites.</td>
<td>P1,2,4,5,9</td>
<td>P2,7</td>
</tr>
<tr>
<td>C4 C5</td>
<td>Generate a joke of a particular Joke Class</td>
<td></td>
<td>P6</td>
</tr>
<tr>
<td></td>
<td>Generate a joke by keyword, from topics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1 D2</td>
<td>Generate a joke by keyword, using alphabet</td>
<td></td>
<td>P4</td>
</tr>
<tr>
<td></td>
<td>Generate a joke by keyword, typing in word</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>Generate a joke appropriate to a current conversation topic.</td>
<td>P1,3,6</td>
<td></td>
</tr>
</tbody>
</table>
### Pre-post test results

CELF WC: choose 2 related items from 4, e.g. “girl boy car table”

PIPA Rhyme: Phonological awareness

<table>
<thead>
<tr>
<th>Level</th>
<th>Participant</th>
<th>CELF word (max. 27)</th>
<th>CELF classes</th>
<th>PIPA Rhyme (max. 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early primary</td>
<td>S1, female; age: 8y4m</td>
<td>19</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Middle primary</td>
<td>S2, female; age: 10y10m</td>
<td>11</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>Middle primary</td>
<td>S3, female; age: 10y9m</td>
<td>23</td>
<td>26</td>
<td>11</td>
</tr>
<tr>
<td>Middle primary</td>
<td>S4, male; age: 10y3m</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Middle primary</td>
<td>S5, male; age: 10y3m</td>
<td>17</td>
<td>26</td>
<td>11</td>
</tr>
<tr>
<td>Senior primary</td>
<td>S6, male; age: 11y3m</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Senior primary</td>
<td>S7, male; age: 12y9m</td>
<td>17</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>Senior primary</td>
<td>S8, male; age: 11y10m</td>
<td>9</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Senior primary</td>
<td>S9, female; age: 11y3m</td>
<td>12</td>
<td>13</td>
<td>10</td>
</tr>
</tbody>
</table>

CELF scores significantly higher on post-test (t-test, 8df, p < 0.01)
Results:

Feedback

Unexpected Outcomes impact on school curriculum

Questionnaires with parent, teachers and Classroom assistants (not significant issues raised but all positive)

Semi-structured interviews with SLTs
Participant Feedback using Talking Mats S1

**Good:**
- Jester character
- Way screen changes
- Way of telling jokes

**OK**
- Jokes
- Scanning

**Bad**
- Voice
### Participant Feedback using Talking Mats S8

**Good:**
- Jester character

**OK**
- Touchscreen

**OK/Bad**
- Way screen changes
- Way of telling jokes
- Voice

**Bad**
- Jokes

<table>
<thead>
<tr>
<th>Bad</th>
<th>OK</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Bad]</td>
<td>![OK]</td>
<td>![Good]</td>
</tr>
</tbody>
</table>

*Image: Talking Mats S8 with 'jokes' indicated as a bad feature.*
STANDUP: some initial conclusions

Interfaces CAN be designed which provide children with CCN with successful access to complex underlying technology

Using STANDUP:
- the generative capabilities allows opportunity for natural language development, cf DA choosing punchline first
- the generative capabilities allows novel explorative learning, cf NI searching subjects

All children benefited
- enhanced desire to communicate
- knock on effect on other AAC usage
- illustrated children’s abilities and potential of AAC

Illustrated use of technology within a wider environment
**STANDUP: some initial conclusions**

**Issues with interface design**
- scanning
- voice output
- improved appropriateness of vocabulary

The telling of the joke is important - what is the impact of **STANDUP:**
- on interactive conversation?
- on joke comprehension and vocabulary acquisition?

Do we want better jokes? (yes)

Use with speaking children with language impairment and other user groups
4. Writing up Experiments and Empirical Studies
Writing-up empirical studies 1

Abstract:
短 summary of the problem, the results and the conclusion.

Introduction:
What is the problem? What related work have other people done?
[Should go from general statement of the problem to a succinct and testable statement of the hypothesis].

Method:

Participants: state number, background and any other relevant details of participants

Materials: exactly what test materials, teaching materials, etc. were used, giving examples

Procedure: clear and detailed description of what happened at each stage in the experiment
[Someone reading should be able to duplicate it from this information alone. Should also clearly indicate what data was collected and how.]
Writing-up empirical studies 2

Results:
Give actual data, or a summary of it.
Provide an analysis of data, using statistical tests if appropriate.
Use tables and graphs to display data clearly.
[Interpretation of results goes in discussion section, NOT here].

Discussion:
Interpretation of results; restating of hypothesis and the implications of results; discussion of methodological problems such as weaknesses in design, unanticipated difficulties, confounding variables, etc.
Wider implications of the work should also be considered here, and perhaps further studies suggested.

Conclusion:
Statement of overall conclusion of the study.
5. References
References - Methodology


References – various studies


References


References


References


References


References

