Improving Computer-Adaptive Psychological Tests

automatic item generation & alternative response formats

Immo Köster

Thesis advisor:
Prof. Dr. Lutz F. Hornke

Second examiner:
Prof. Dr. Jan O. Borchers

Daily advisor:
Dipl.-Psych. Jonas W.B. Lang

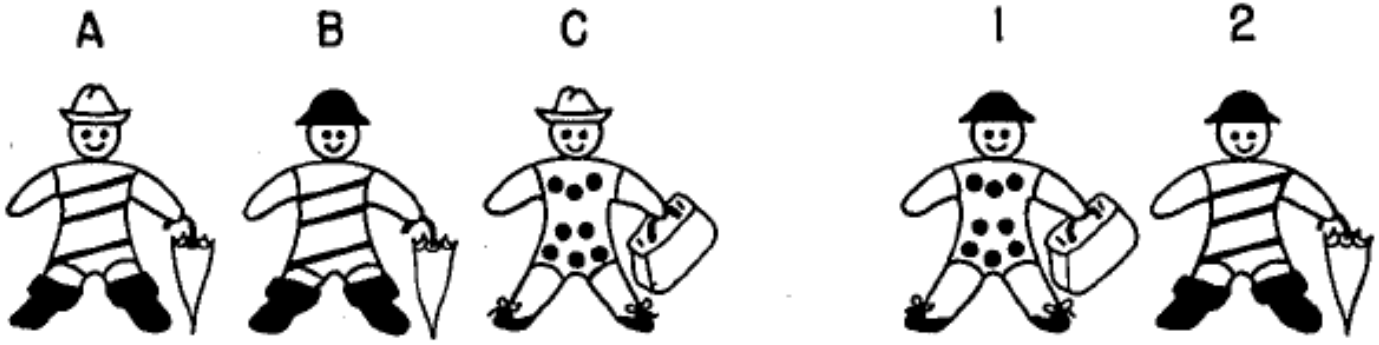
Outline

- objectives
- software prototype
- response formats
- demonstration
- psychological study
- results
  - Psychometrics
  - Usability
- discussion

further results in the thesis

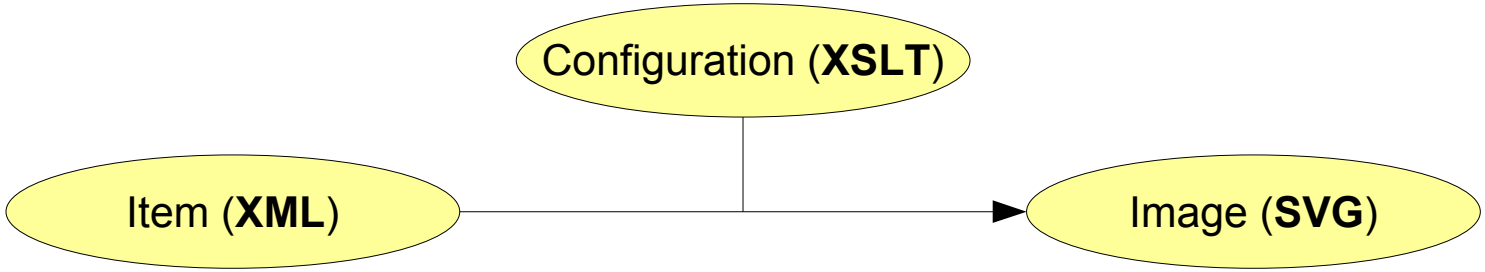
Objectives

- develop a **software prototype** of a **computer-adaptive test** for **analogical reasoning**
- inspired by **figural analogies** used by Sternberg (1977) and Sternberg & Rifkin (1979)
- **automatic item generation**
  - empirical models of item difficulty (Mulholland et al., 1980; Bethell-Fox et al., 1984; Leon & Revelle, 1985)
  - Linear Logistic Test Model (LLTM; Fischer 1973)
- optimize **response format** and **usability**



Realization: Software prototype

- Java
- rendering:



- adaptive testing
  - Expected a Posteriori (EAP) estimation (Bock & Aitkin, 1982)
- automatic generation of items & distractors
  - LLTM based on elements and transformations of the analogy
- high flexibility:
  - customizable structure & rendering of figures
  - XML User Interface Language: SwiXml
    - → minor UI changes do not require to recompile
  - i18n & l10n using Java property files
    - → currently German & English,
    - → allows to focus on the psychological aspects of translation
  - configuration through Java property files
    - → e.g., computer-adaptive test or static fixed-length test ?

Optimize response format & usability

- optimization criteria for psychological tests:
  - Flow experiences (Rheinberg, 2004)
  - State anxiety: Worry & Emotionality (Zeidner, 1998)
  - Performance (percentage correct)
    - → maximize *unbiased* individual performance
- usability optimization criteria:
  - typical usability questions
  - individual feedback

Literature search

- limited research on response formats in psychological tests available
- previously studied response formats in psychological research (Martinez, 1999):
  - Multiple-Choice (MC)
    - very common & familiar, easy to explain
    - easy scoring allows for immediate feedback
    - **but** performance biased: guessing chance, response elimination strategy
  - Computerized Modified Multiple-Choice Testing (Park, 2005)
    - reasoning set apart from response recognition and selection (user has to actively request response options)
    - time limit (few seconds) hinders response elimination strategy
    - **but** hard to find an appropriate time limit due to individual differences
  - Non-Computer-based Constructed-Response (CR)
    - no response options, answer has to be constructed
    - eliminates MC bias
    - **but** scoring is difficult, subject to interpretation

New response format

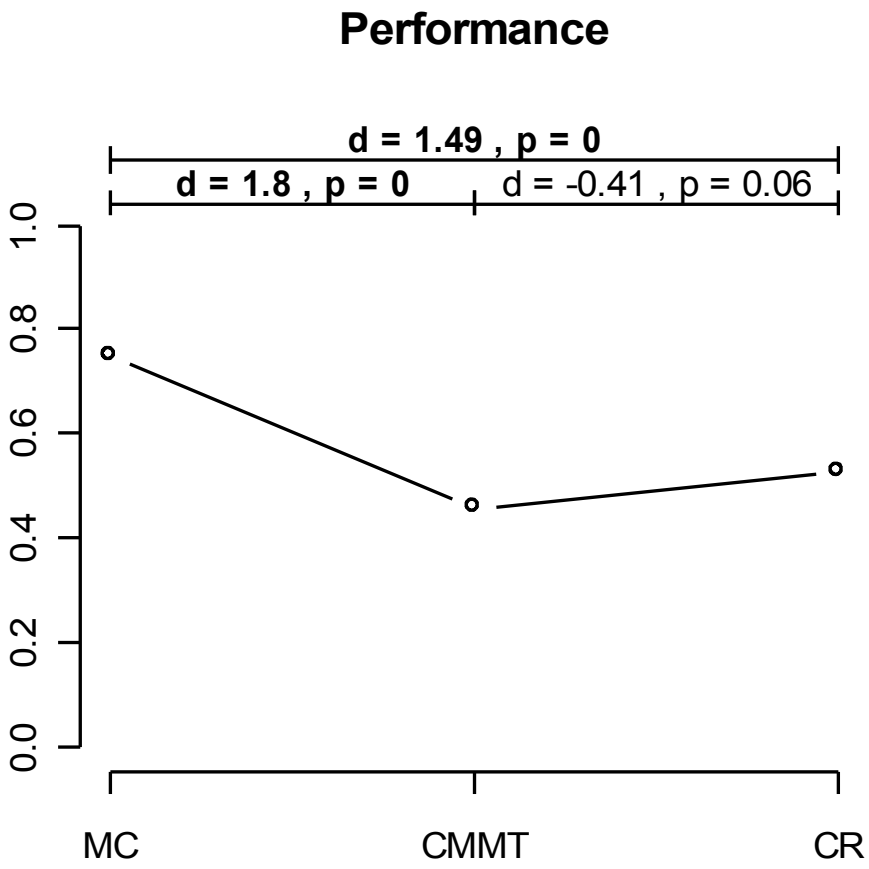
- newly developed response format in this thesis:
  - Computer-based Constructed-Response (CCR)
    - automatic scoring (compare MC)
    - high subjective user control (beneficial for people with high test anxiety)
    - inspired by usability research & design patterns

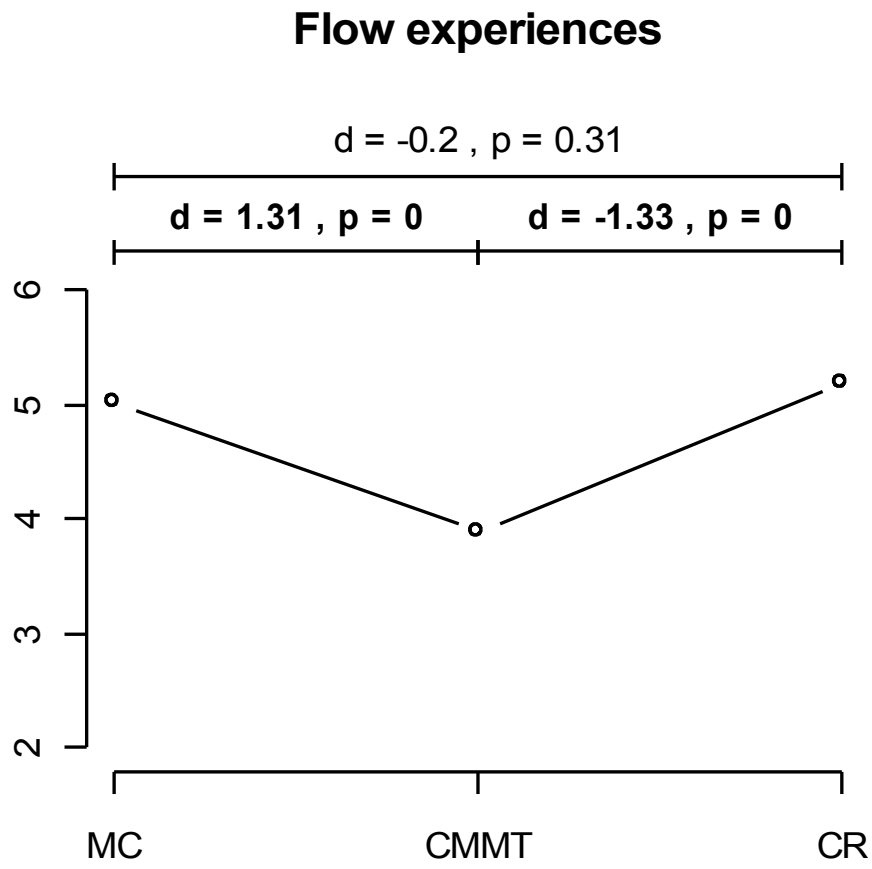
Time for a short demonstration...

Study: Objective, Method, Design, & Analysis

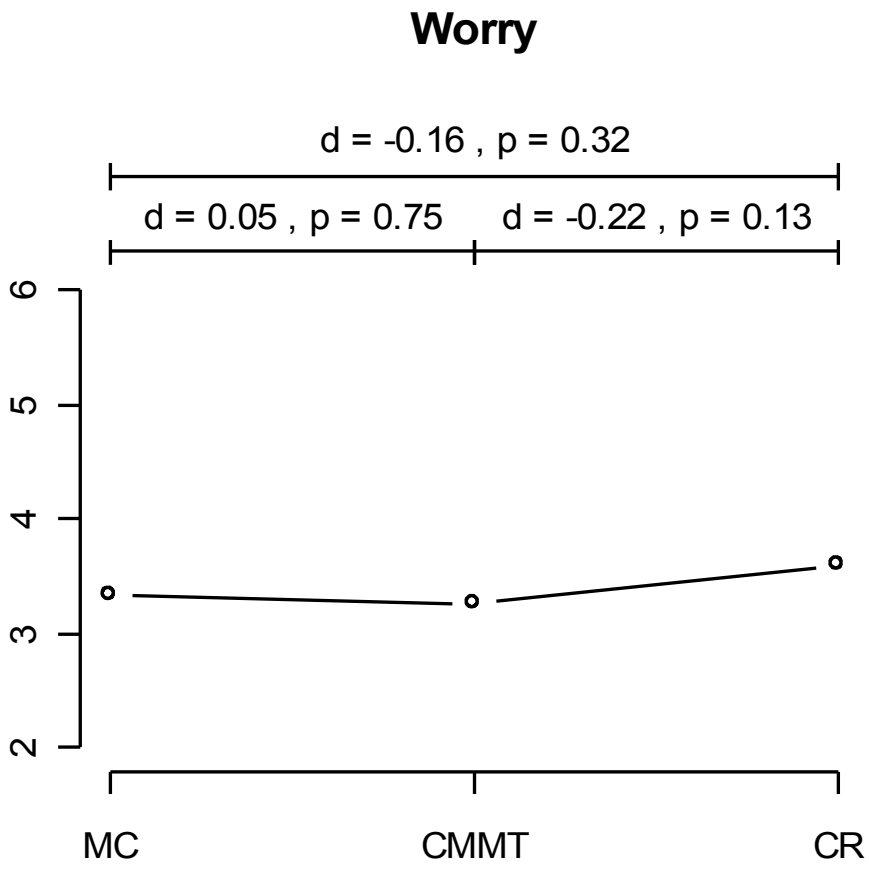
- **Objective:** comparison of three response formats with respect to outlined criteria
- **Method/Design:** repeated measures design, 27 subjects
  - 3 different response formats
    - order: MC-CMMT-CR, CMMT-CR-MC, CR-MC-CMMT
  - 3 different item sets
  - 3 subjects per condition
  - psychological questionnaires
- **Data Analysis:**
  - paired t-tests
  - Cohen's d (effect size measure)
    - $d \geq 0.2$: small effect size
    - $d \geq 0.5$: medium effect size
    - $d \geq 0.8$: large effect size

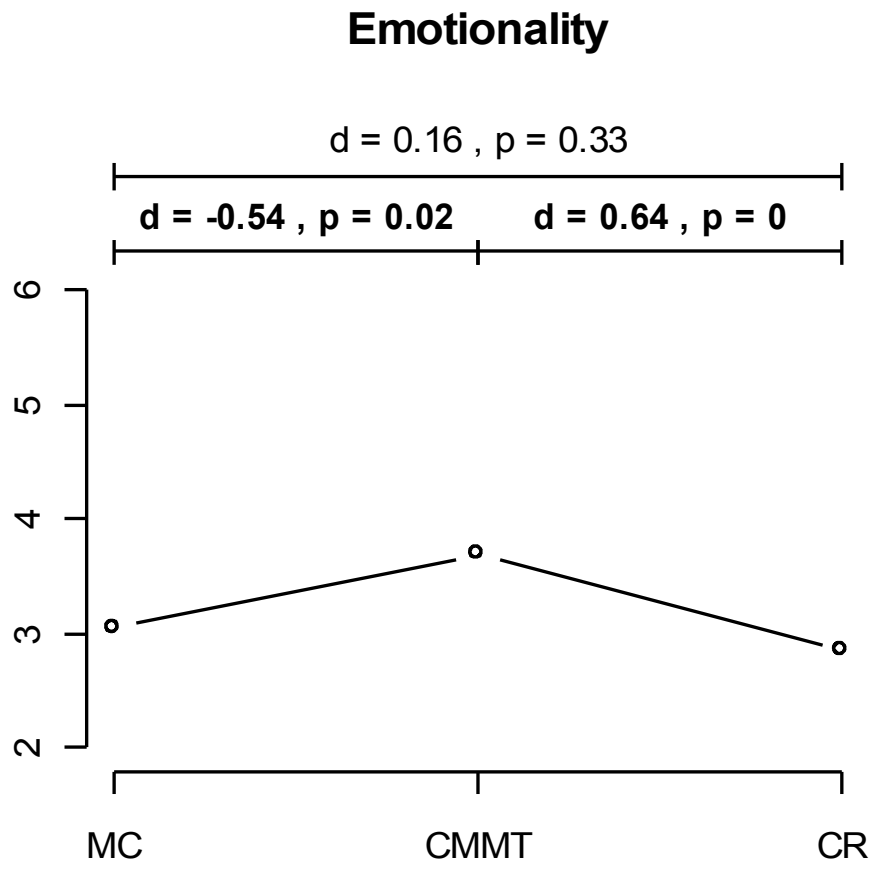
Study: Results – Psychometrics (1/2)





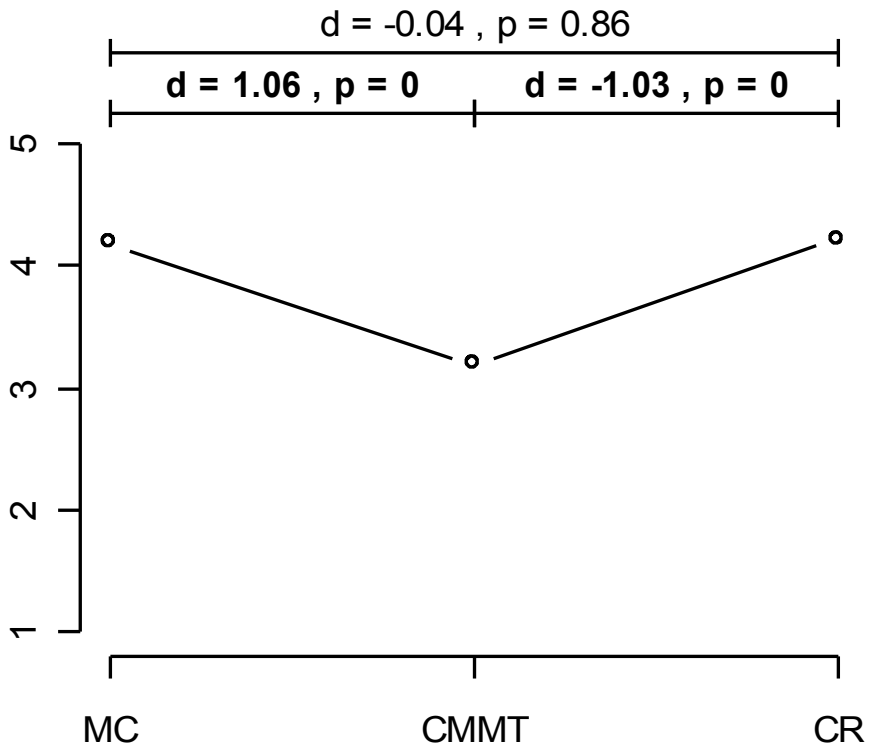
Study: Results – Psychometrics (2/2)



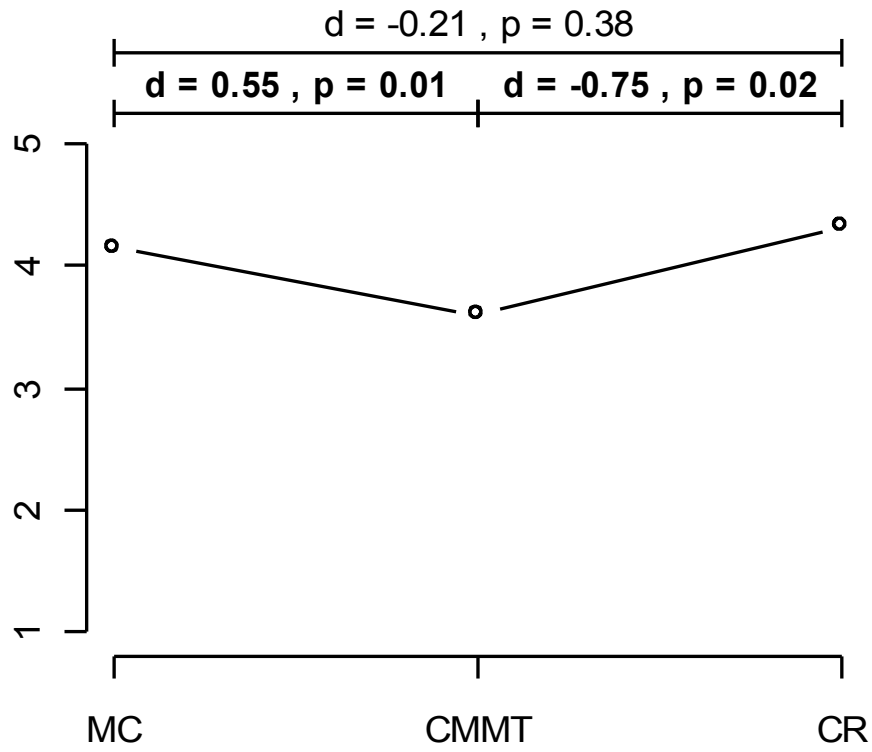


Study: Results – Usability (1/3)

Everything worked as I expected it



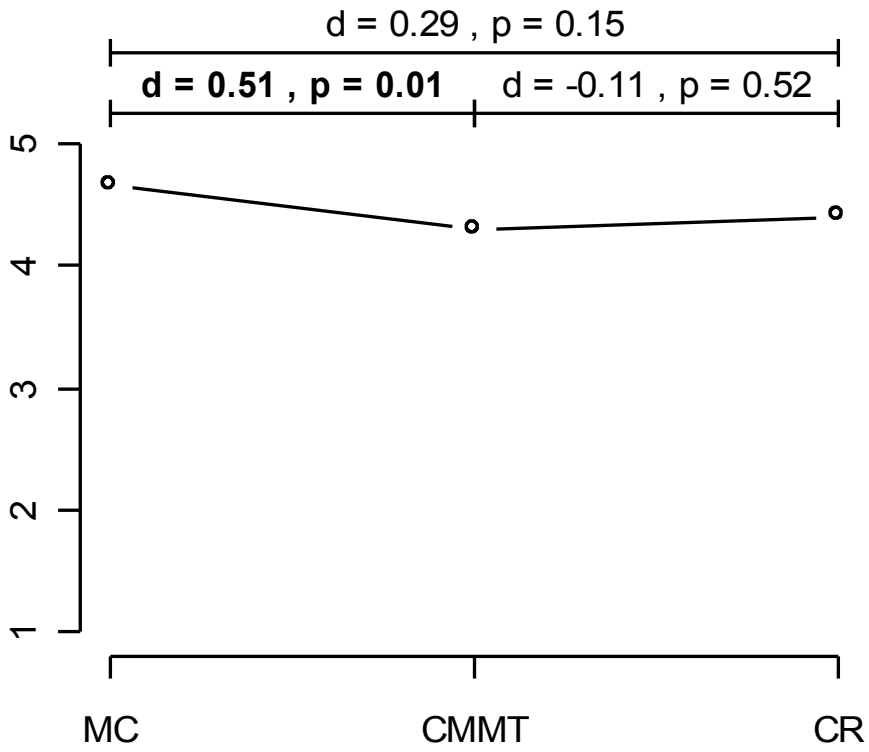
It was always clear to me what to do next



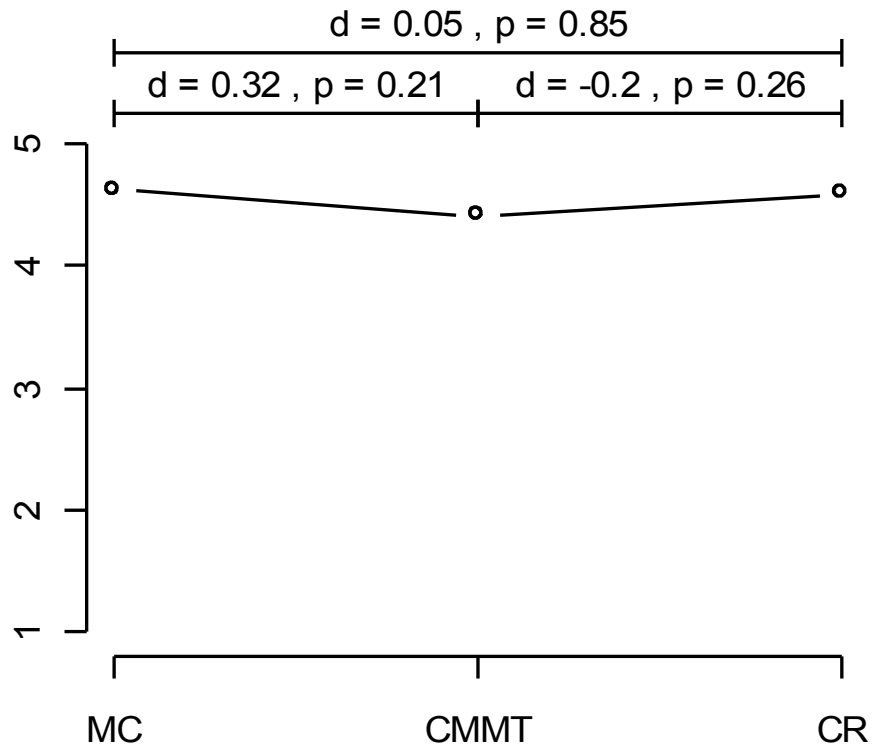
- Overall user feedback was positive
  - “good user interface”
  - “enjoyed it”
  - “the diversity of response options was impressive”

Study: Results – Usability (2/3)

Arrangement of control elements was concise

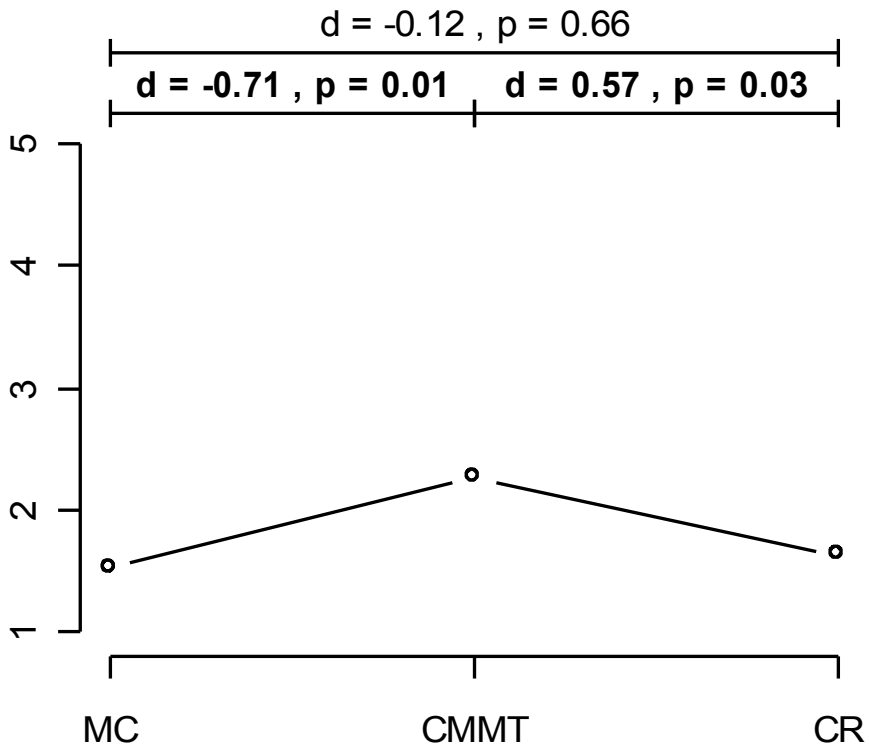


Usage of the program was easy to learn

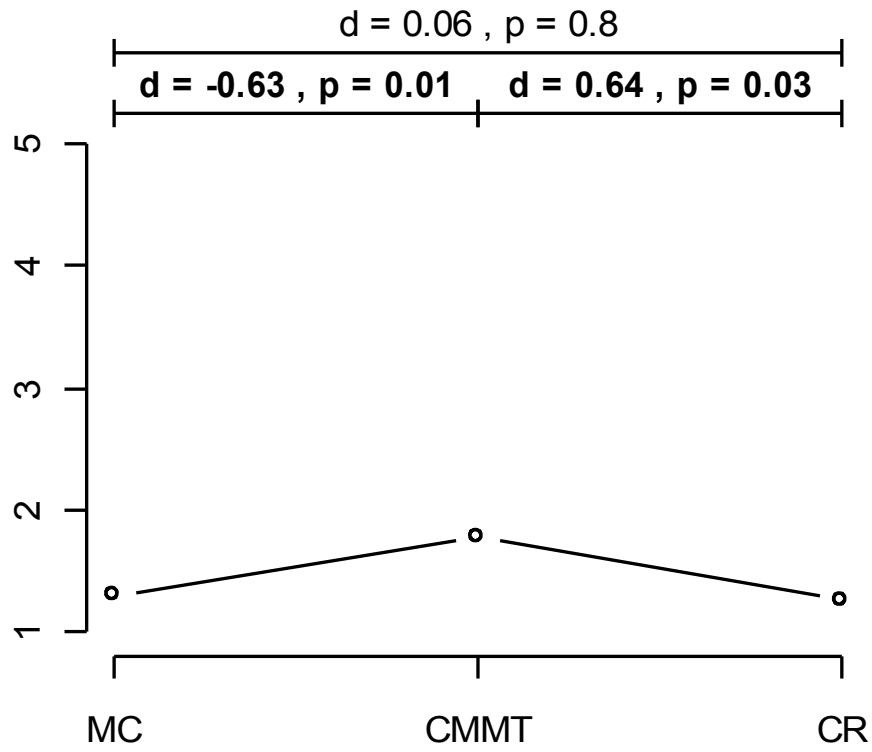


Study: Results – Usability (3/3)

The representation of the interface confused me



Working with the interface was a problem for me



- some problems:
  - time limit of CMMT format was too short [6 participants]
  - color contrasts (green and black) too low [3 participants]
  - figures (a bit) too small [2 participants]
  - contrast of small body shapes too low [1 participant]

Discussion

- CMMT did not fulfill the expectations
  - chosen time limit too short
  - probably UI issues as well
  - main problem: forced separation of reasoning and answering
- CCR performs very well
  - no significant differences to MC
  - exception: performance, which is known to be biased for MC
  - “it was great fun to assemble the figures myself”
- adaptive testing requires large LLTM calibration study

References

- Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive Reasoning: Componential and Eye Movement Analysis of Geometric Analogy Performance. Intelligence, 8 (3), 205-238.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP Estimation of Ability in a Microcomputer Environment. Applied Psychological Measurement, 6 (4), 431-444.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. Acta Psychologica, 37, 359-374.
- Leon, M.R., & Revelle, W. (1985). Effects of Anxiety on Analogical Reasoning: A Test of Three Theoretical Models. Journal of Personality and Social Psychology, 49 (5), 1302-1315.
- Martinez, M. E. (1999). Cognition and the question of test item format. Educational Psychologist, 34 (4), 207-218.
- Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. Cognitive Psychology, 12, 252-284.
- Park, J. (2005). Learning in a new computerized testing system. Journal of Educational Psychology, 97 (3), 436-443.
- Rheinberg, F. (2004). Motivationsdiagnostik. Göttingen: Hogrefe.
- Sternberg, R. J. (1977). The people piece analogy experiment. In Intelligence, information processing and analogical reasoning: The componential analysis of human abilities (pp. 173-221). Hillsdale, NJ: Lawrence Erlbaum.
- Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. Journal of Experimental Child Psychology, 27, 195-232.
- Zeidner, M. (1998). Test Anxiety: The State of the Art. New York: Plenum.