

Review Report for the Ph.D. thesis
submitted to Charles University
Faculty of Physical Education and Sport

Title: Computerized Adaptive Testing in Kinanthropology: Monte Carlo Simulations Using the Physical Self Description Questionnaire

PhD Candidate: Martin Komarc

Supervisor: Doc. PhDr. Jan Štochl, M.Phil., Ph.D.

General comments

The submitted thesis addresses a very up to date topic – computerized adaptive testing (CAT) and its application to measurement in kinanthropology. As the main contribution, the presented thesis uses a simulation study to confirm applicability of the CAT approach to the administration of Physical Self Description Questionnaire (PSDQ). In contrast to education and psychology where CAT has been used for decades, the use of CAT in kinanthropology has been minimal and for this reason the presented thesis makes a strong contribution to the subject field.

The thesis is 125 pages long and it is built on 12 chapters. It includes 4 Appendices, 7 tables and 24 figures. Reference list contains 179 references demonstrating author's familiarity with key literature in the field. Thesis is written in English, wording of the thesis is clear with minimum of misprints which I provide on additional sheet.

The thesis is clearly structured. Chapters 1-6 provide thorough literature review and methodological grounds of various aspects of CAT. It starts from "Brief introduction to measurement", moves through "Historical paths to modern test theory", forward to "Groundwork for Item Response Theory". In Chapter 4 devoted to IRT, author presents various dichotomous and polytomous IRT models and discusses assumptions and parameter estimation. Chapter 5 provides historical and conceptual origins for CAT, whereas chapter 6 provides detailed description of testing algorithms in unidimensional IRT-based model including settings for the phase of starting, continuing and stopping the CAT, as well as practical issues such as size of the item pool and psychometric quality of its items, content balancing or exposure control to prevent overexposure of some items.

Empirical part starts with Chapter 7 on page 68. Chapter 8 provides clearly stated aims and more detailed hypotheses related to the simulation study. Chapter 9 provides detailed description of selected methods. Results are clearly presented in Chapter 10 with focus on length of the test and bias of latent trait estimate under different conditions of the CAT. Results are further discussed in Chapter 11, conclusion is provided in Chapter 12. Appendices contain item wording and IRT parameters of PSDQ previously published in Fletcher & Hattie, 2004, and used in this thesis, together with selected R code and Figure of Test information and standard error for the PSDQ.

Overall the thesis is clearly written, theoretical part provides strong background for the empirical part. Empirical part uses adequate methods in the simulation study as well as appropriate methods to present the results. Discussion covers most of the issues one may think of while reading the thesis.

Specific comments and discussion points

1. On page 72, author discusses difference between Monte-Carlo study (which only needs item parameter estimates) and a Post-hoc study (which uses actual data). In the thesis, a Monte-Carlo study is used. I would appreciate a discussion comparing the two type of studies and commenting on when each of them should be used. E.g., is Post-hoc study preferable in case we have the actual data, or should it be also accompanied by Monte-Carlo study?
2. On page 75₁, the relationship between SE and reliability should be referenced.
3. I very much appreciate that the analyses were performed using freely available software R. I value R highly as it provides possibility for reproducible research. To fully use advances of the software, I would recommend the author to provide a full code and item-parameters on a CD attached to the thesis or online. In the code, I would recommend using `set.seed()` function at the beginning of the simulation which allows for full replication of the results.
4. Please compare the algorithm for simulation of CAT administration described in Methods on page 76 with algorithm described in the reference manual for `simlirt()` function of the `catlirt` package.
5. Package `catlirt` was used for the analyses. Other two relevant R packages are available – `catR` and `mirtCAT`. Please discuss the advantages which provides `catlirt` and which might be provided by `mirtCAT` and `catR`.
6. In Results section, on page 79, dimensionality analysis is provided which is performed on data generated using a unidimensional model. Please provide justification for this analysis.
7. What limitations may bring the fact that the PSDQ was calibrated on Australian sample of high school students involved in sports activities?
8. As mentioned in the Discussion (p. 100), the fact that content balancing was not considered in the simulation study is a potential limitation of this work, given the fact that PSDQ was intended to measure 11 different specific sub-domains. I would recommend to check that the results of the simulation study don't change substantively when content balancing is applied.
9. I fully agree with the author (p. 100) that a natural continuation of this work would be creating an on-line tool for adaptive test of PSDQ. Could you mention freely available tools and packages which could be used?

Conclusion

Provided comments are meant as points for discussion or suggestions for improvement. Overall, thesis submitted by Mr. Martin Komarc represents an impressive amount of work worthy of the doctoral distinction.

Patrícia Martinková, PhD
Researcher, Psychometrics and Statistics
Institute of Computer Science, Czech Academy of Sciences
Pod Vodárenskou věží 2, 182 07, Prague, Czech Republic
E-mail: martinkova@cs.cas.cz