Final project assignment

The final project comprises validation and analysis of CERMAT measurement data or other data of your choice. Your final project is to be submitted electronically via e-mail in PDF form. As the Supplementary Materials, you should include .R (or .Rmd) file with commented R code, including data preparation and code for relevant analyses, dataset(s) needed to run the code, and/or a PDF or HTML ShinyItemAnalysis report.

Your project should be written in Czech or English and should contain the following sections and content specified below:

Data (10 pts)

Describe your data: Which dataset did you choose? Provide the links for your dataset. Full item wording should be provided as Supplementary Material, or a link/reference for the document with item wording should be included. How many respondents were in your data? How many items and of what type are present in your data? Describe data preparation: How did you recode the data? Did you merge some items? What other changes to the data were necessary?

Provide descriptive statistics and a histogram for the total score, descriptives for item scores, and any additional variables (grouping, criterion, etc.). Compute the Z-scores, T-scores, percentiles, and success rate for all respondents, and interpret the results for the first respondent.

Test validity (10 pts)

What construct is your test measuring? What is the intended use of test scores? Is there another use you can think of? How would you obtain evidence of the validity of test scores based on test content? How would you obtain evidence of criterion validity? What data would you need? Use terms concurrent, predictive, incremental, convergent, and discriminant. Provide proof of criterion validity if the criterion is available. Provide evidence of the validity of test scores based on internal structure: Include corrplot, and discuss which items correlate the most. Try cluster analysis, and interpret the dendrogram. Try exploratory factor analysis, and discuss the path diagram.

Test reliability (10 pts)

Provide reliability estimate(s) regarding internal consistency, include confidence interval, and interpret with respect to Cicchetti's cut-off values. Implement the Spearman-Brown formula. What would be the estimated reliability if you doubled the number of items? What is the number of items needed for reliability of 0.9? What kind of data would you need to collect to provide other proofs of reliability? Discuss at least two other reliability estimates.

Item analysis (10 pts)

Include a table of traditional item indices (difficulty, RIR, RIT, ULI, alphaDrop). Comment on items with the highest and lowest difficulty, items of low discrimination, or items with inappropriate distractor plot. Provide an explanation taking into account the wording of those items.

Choose one or two items and include empirical item characteristic curve these items, or distractor plots in case of multiple-choice items. Select and fit an appropriate regression model with respect to the standardized total score using intercept-slope and/or IRT parametrization. Discuss your model decision with respect to item type. Interpret item parameters.

Item response theory models (10 pts)

Select optimal IRT model: Provide reasoning for the selected IRT model (based on data type or comparing more IRT models). Provide model equation(s) and interpretation of parameters. Which method was used for the estimation of parameters?

Plot and discuss a Wright map (preferably on the 1PL model or other Rasch-type models). Plot item characteristic curves, item information curves, and test information curves. Provide a table of model parameter estimates and their standard errors, or confidence intervals. Which item is the most informative for average ability level, and ability levels 1SD above and 1SD below the average?

Provide an ability estimate for the first respondent, including the standard error or confidence interval. Plot relationship between ability estimates in IRT models and traditional ability estimates based on (standardized) total scores.

Differential item functioning (10 pts)

Explain why DIF analysis is important in test validation. Select one grouping variable, and provide an analysis of DIF. Provide reasoning for the selected DIF detection method, or try more methods and compare results. Discuss any items detected as DIF, provide wording or explanation for DIF: Which items favor one and which the other group?What other grouping variables should be considered for DIF analysis? If no grouping variable is present in your data, discuss what grouping variables should be collected and used for DIF detection, and which DIF detection methods could be used.

Discussion (10 pts)

Provide a discussion of your results. What are the results, and are they expected? What recommendations would you provide for the analyzed assessment instrument (regarding removing/rewording items, increasing reliability or validity, data to be collected for further analyses, etc.)? For what other situations/data is your analysis relevant? What are the limitations of your study/analysis? Could you think of a situation when range restriction may occur in your data? How would you adjust your analysis? What is your conclusion?

Supplement – Commented R code (10 pts)

Attach commented .R or .Rmd code for your analysis.

Supplement –ShinyItemAnalysis report and datasets (10 pts)

Attach automatically generated PDF or HTML report, including available analyses. Attach datasets used to generate the report. In this section, describe the settings you chose and discuss the results. How do the models or results differ from the results above? (Based on limitations of the ShinyItemAnalysis, or due to reduced dataset).

Report form (10 pts)

- Make the project one fluent text. Section instructions should not be repeated in your project.
- Use your own words to describe methods. When citing longer text, a reference to the original source needs to be given.
- When using `RMarkdown` with `knitr`, see https://bookdown.org/yihui/rmarkdown-cookbook/, hide code and output messages, see options here https://yihui.org/knitr/options/

Additional analyses (Bonus - up to 10 pts)

Provide any additional relevant analysis that is not included in the list above. You may include these analyses in the relevant section above; in such case, mention and discuss these analyses in this section. Which additional analyses did you consider? What were the results? Here are some possibilities: Testing item and person fit, estimating reliability with McDonald's omega, and/or running post-hoc CAT analysis. DIF: Provide a table of item parameters and interpret for at least one DIF item; choose some DIF items and provide item characteristic curves; provide a prediction for the correct answer from members of one and of the other group.

Policy on the use of generative artificial intelligence (AI). We will follow the APA policy[1]. For this policy, AI refers to generative LLM AI tools and does not include grammar-checking software, citation software, or plagiarism detectors.

- When a generative AI model is used in drafting the project, the use of AI must be disclosed and cited.
- When AI is cited, employ the software citation template, which specifies how, when, and to what extent AI was used. You are also required to upload the full output of the AI as supplemental material.

Due date: The Final project (PDF file and supplementary materials) is due at least 48 hours before the oral exam and needs to be sent to martinkova@cs.cas.cz. One feedback will be provided to drafts sent at least one week before the oral exam.

[1] https://www.apa.org/pubs/journals/resources/publishing-policies.