

Homework 1

Measurement and Evaluation of HCC Systems

How to hand in this homework

- Please email the homework to me as a PDF.
- Late assignments get a penalty of 20% when submitted after the deadline, plus an additional 10% per hour late.
- Make sure you include the R input you used to get to your answer, but do not “dump” the resulting R output on the paper. Copy from the output selectively, and explain it in your own words.
- You may collaborate on this homework, but not copy from others... again, please write your answers in your own words.
- Please include a collaboration statement that says: “I collaborated on this homework with [name].” or “I worked alone on this homework”

Dataset

You are going to use the dataset from an experiment conducted on the TasteWeights recommender system. The TasteWeights system uses the overlap between your Facebook “likes” and the “likes” of your Facebook friends to give you music recommendations. This works as follows:

- Your friends are given a “weight” based on the overlap between your music likes and your friends’ music likes: a friend with 2 overlapping likes gets a weight of 2.
- Then, the friends’ *other* music likes—the ones that are not among are user’s likes—are tallied by weight: if three friends with weights 2, 5 and 7 all like Coldplay, then Coldplay gets a score of $2+5+7 = 14$.
- This tallied list is then sorted by score, and the Top 10 is presented to the user.

Manipulations

We tested two features of the TasteWeights system. These are translated into two manipulations: “Control” (3 levels) and “Inspectability” (2 levels). All combinations are tested, so there are 6 conditions. This is a between-subjects experiment.

- **Control (see Figure 1)**

- **None:** The user has no control; the system calculates the recommendation with the default method (see above).
- **Item:** In this condition, participants can weigh their likes before the system calculates their recommendations. The friend weights are now the sum of the weights of the overlapping items (rather than just the number of overlapping items).
- **Friend:** in this condition, participants can adjust the friend weights before the system calculates their recommendations.

This manipulation is represented by two dummy variables, **citem** and **cfriend**.

- **Inspectability (see Figure 2)**

- **List view:** The recommendations are presented as a sorted list.
- **Graph view:** The recommendations are presented as a graph that shows their music likes, their friends, and their recommendations. By hovering over the lists, they can see how these things are connected to each other.

This manipulation is represented by the dummy variable **cgraph**.

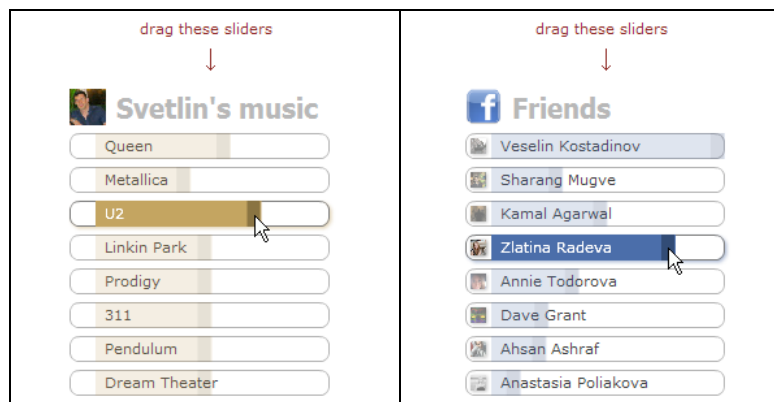


Figure 1: Item control (left) and Friend control (right)



Figure 2: List view (left) and Graph view (right)

Procedure and measurements

At the start of the experiment, the participants get a questionnaire measuring the following personal characteristics:

- **Expertise:** Participants' music expertise, measured with four 5-point scale questions (ranging from -2 to +2), translated into a sum score.
- **Trust:** Participants' trusting propensity, measured with three summed 5-point scale questions.
- **Familiarity:** Participants' familiarity with recommender systems, measured with two summed 5-point scale questions.

Participants are then taken to the experiment where they are asked to do the following:

- Control the items/friends (or not), depending on the condition they are in
- Inspect the recommendations (either in graph or list, depending on the condition), the time they take to inspect the recommendation is captured in the variable **time** (measured in seconds)
- Rate the recommendations on a 5-star rating scale (these ratings are averaged as the variable **rating**)
- Indicate which of the recommendations they already knew (the number of known recommendations is captured as the variable **known**)

At the end of the experiment, the participants get another questionnaire measuring the following subjective evaluations:

- **Satisfaction:** Participants' overall satisfaction with the TasteWeights system, measured with seven summed 5-point scale questions.
- **Quality:** Participants' perception of the overall quality of the recommendations, measured with six summed 5-point scale questions.
- **Perceived_control:** Participants' perception of their control over the recommendations, measured with four summed 5-point scale questions.
- **understandability:** Participants' understanding of how the recommendations were calculated, measured with three summed 5-point scale questions.

If you want to learn more about this experiment, you can read Knijnenburg et al. (2013) "Inspectability and control in social recommenders". DOI: 10.1145/2365952.2365966

In this homework assignment, you are going to specify, identify, fit, and trim the model presented in Figure 3. Please remember that the thick arrows are the hypothesized effects, while the thin arrows are added for saturation.

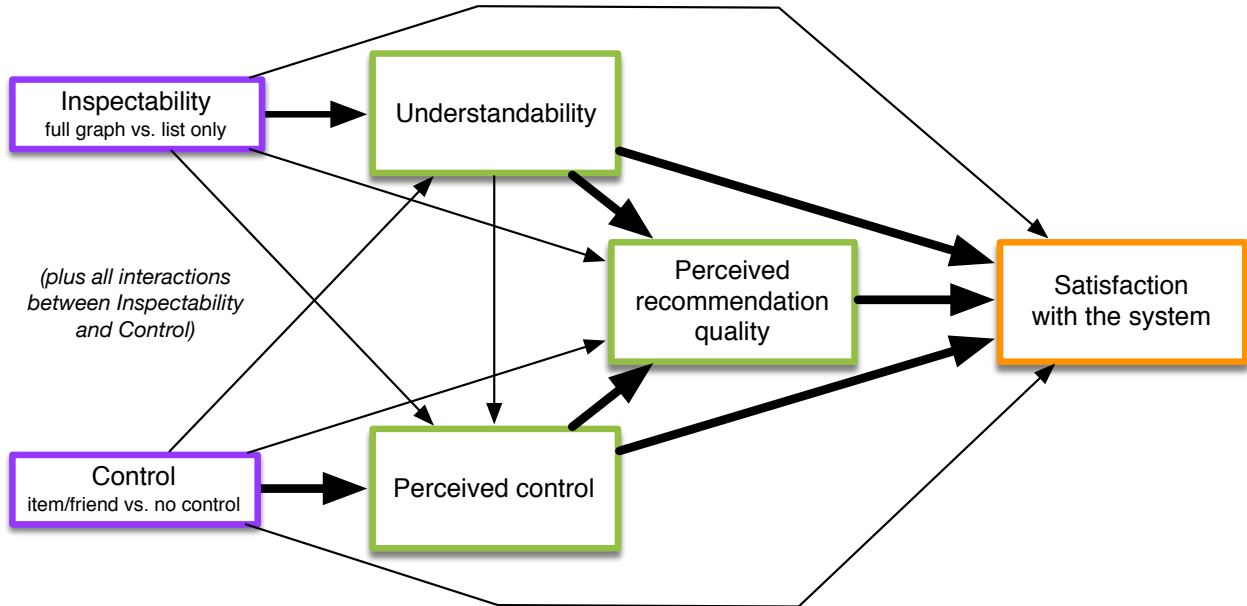


Figure 3: Saturated path model

Question 1. Specify, identify, and run the model

- Write the model specification in lavaan syntax. Make sure to include all dummies for main and interaction effects of inspectability and control. Hint: the regression for satisfaction will look like this: `satisfaction ~ quality + perceived_control + understandability + cgraph + citem + cfriend + cig + cfg`
- Is this model recursive? Why (not)?
- How many observations does this model have? How many parameters? How many degrees of freedom?
- Is this model under-identified, just-identified, or over-identified? Explain your answer.

Question 2. Run the model

- Fit the model using a robust maximum likelihood estimator. Get a summary of the output of the model.
- Interpret the effect of quality and perceived control on satisfaction. Make sure to mention the regression coefficient, its standard error, the test statistic, and the p-value. Hint: don't forget that you are controlling for the other variables in the regression!
- Interpret the effects of the experimental conditions and their interaction on understandability. Make sure to mention the regression coefficients, their standard errors, the test statistics, and the p-values.

Question 3. Trim the model (part 1)

Since none of the interaction effects seem to be significant, we will trim them from the model.

- a. Edit your model specification by removing all interaction effects between inspectability and control (cig and cfg). Make sure you keep the main effects of inspectability and control (cgraph, citem and cfriend)!
- b. How many degrees of freedom does this new model have?
- c. Fit the model using a robust maximum likelihood estimator. Get a summary of the output of the model, including R-square values.
- d. What percentage of the variance in satisfaction is explained by the combined effects of quality, perceived control, understandability, and the experimental conditions?

Question 4. Trim the model (part 2)

The next candidates for removal are the main effects of inspectability and control on satisfaction.

- a. Why don't we first remove the effect of control on perceived control?
- b. Edit your model specification by removing the main effects of inspectability and control on satisfaction.
- c. How many degrees of freedom does this new model have?
- d. Fit the model using a robust maximum likelihood estimator. Get a summary of the output of the model.
- e. What is the chi-square model fit of this model? What does this test mean?

Question 5. Trim the model (part 3)

Next, we remove the main effects of inspectability and control on perceived control.

- a. Edit your model specification by removing the main effects of inspectability and control on perceived control.
- b. Fit the model using a robust maximum likelihood estimator. Get a summary of the output of the model.
- c. Does the removal of these effects mean that inspectability and control have no effect whatsoever on perceived control? Why (not)?
- d. If not, then how can we characterize the effects of inspectability and control on perceived control?

Question 6. Final model

Finally, we remove the main effects of inspectability and control on quality, and the effect of understandability on quality and on satisfaction

- Edit your model specification removing the mentioned effects. Also, add labels to “citem” and “cfriend” in the “understandability” regression.
- If we remove “citem” from the “quality” regression, we must also remove “cfriend”. Why?
- Fit the model using a robust maximum likelihood estimator. Get a summary of the output of the model, including R-square values and fit measures.
- Conduct the omnibus test on the effect of the control manipulation on understandability. Report and interpret the results.

Question 7. Reporting

- What is the chi-square model fit of the model? What does this test mean?
- What is the outcome of the model test against the baseline model? What does this test mean?
- What are the values of CFI and TLI? Are these values adequate?
- Interpret the RMSEA, its confidence interval, and the p -value of $RMSEA \leq 0.05$.
- Interpret the regression coefficients of the final model. Make sure to mention the regression coefficients, their standard errors, the test statistics, and the p -values. Write your interpretation like a “story” that explains what influences what.
- Finally, interpret the r -squares of the final model.

Bonus question: create a graphical representation of the final model like the one below, filling in the values for the x 'es. Note: we usually put regression coefficients on the arrows with standard errors in parentheses, and use stars to represent p -values: * $p < .05$, ** $p < .01$, and *** $p < .001$.

