

Appendix B

Checklist for the Empirical Cycle

This checklist can be used to design your research, write a report about it (internal report, published paper, or thesis), and read a research report written by others. The questions are precisely that: questions. They are not instructions to do something. It is up to you to decide how to design and execute your research, how to write a report, and how to read one. Related to this, the checklist is not a table of contents for a report. But you can use it to get inspiration for a report.

The first ten items can be used to design your research and is also useful when considering what to include in a report. They are stated from a design point of view, in the future tense. The remaining part of the list can be used for reporting and not for designing. The items are written in the past tense.

Research Context

1. Knowledge goal(s)
 - What do you want to know? Is this part of an implementation evaluation, a problem investigation, a survey of existing treatments, or a new technology validation?
2. Improvement goal(s)?
 - If there is a higher-level engineering cycle, what is the goal of that cycle?
 - If this is a curiosity-driven project, are there credible application scenarios for the project results?
3. Current knowledge
 - State of the knowledge in published scientific, technical, and professional literature?
 - Available expert knowledge?
 - Why is your research needed? Do you want to add anything, e.g., confirm or falsify something?
 - Theoretical framework that you will use?

Research Problem

4. Conceptual framework
 - Conceptual structures? Architectural structures, statistical structures?
 - Chance models of random variables: Semantics of variables?

- Validity of the conceptual framework? Clarity of definitions, unambiguous application, avoidance of mono-operation and mono-method bias?
5. Knowledge questions
- Open (exploratory) or closed (hypothesis-testing) questions?
 - Effect, satisfaction, trade-off, or sensitivity questions?
 - Descriptive or explanatory questions?
6. Population
- Population predicate? What is the architecture of the elements of the population? In which ways are all population elements similar to each other and dissimilar to other elements?
 - Chance models of random variables: Assumptions about distributions of variables?

Research Design and Validation

7. Object(s) of study

7.1 Acquisition of objects of study

- * If OoSs are selected, how do you know that a selected entity is a population element?
- * If OoSs are constructed, how do you construct a population element?
- * Validity of OoS
 - *Inference support*. Which inferences would be valid with respect to this design? See checklists for validity of descriptive statistics and abductive and analogic inferences.
 - *Repeatability*. Could other researchers use your report to construct or select a similar OoS?
 - *Ethics*. Are people informed that they will be studied, and do they consent to this? Are they free to stop at any time without giving reasons, and do they know this?

7.2 Construction of a sample

- * Case-based research: What is the analytical induction strategy? Confirming cases, disconfirming cases, extreme cases?
- * Sample-based research: What is the sampling frame and probability sampling strategy? Random with or without replacement, stratified, cluster? What should the size of the sample be?
- * Validity of sampling procedure
 - *Inference support*. Which inferences would be valid with respect to this design? See the applicable parts of the checklists for validity of statistical, abductive, and analogic inferences.
 - *Repeatability*. Can the sampling procedure be replicated by other researchers?
 - *Ethics*. No new issues.

8. Treatment design

- Which treatment(s) will be applied?
- Which treatment instruments will be used? Instruction sheets, videos, lessons, software, computers, actuators, rooms, etc.
- How are treatments allocated to OoSs?
 - * In sample-based research: Blocking, factorial designs, crossover designs? Between-subjects or within-subject designs?
 - * In case-based research: Are treatments scaled up in successive cases?
- What is the treatment schedule?
- Validity of treatment design:

- * *Inference support.* Which inferences would be valid with respect to this design? See the applicable parts of the checklists for validity of statistical, abductive, and analogic inferences.
- * *Repeatability.* Is the specification of the treatment and the allocation to OoSs clear enough so that others could repeat it?
- * *Ethics.* Is no harm done, and is everyone treated fairly? Will they be informed about the treatment before or after the study?

9. Measurement design

- Variables and constructs to be measured? Scales, chance models.
- Data sources? People (e.g., software engineers, maintainers, users, project managers, politically responsible persons, etc.), primary data (e.g., source code, log files, bug tracking data, version management data, email logs), primary documents (e.g., project reports, meeting minutes, organization charts, mission statements), etc.
- Measurement instruments? Interview protocols, questionnaires, video recorders, sound recorders, clocks, sensors, database queries, log analyzers, etc.
- What is the measurement schedule? Pretests, posttests? Cross-sectional or longitudinal?
- How will measured data be stored and managed? Provenance, availability to other researchers?
- Validity of measurement specification:
 - * *Inference support.* Which inferences would be valid with respect to this design? See the applicable parts of the checklists for validity of abductive and analogic inferences.
 - * *Repeatability.* Is the measurement specification clear enough so that others could repeat it?
 - * *Ethics.* Which company data must be kept confidential? How is privacy of persons respected?

Inference Design and Validation

10. Inference design

10.1 Descriptive inference design

- * How are words and images to be interpreted? (Content analysis, conversation analysis, discourse analysis, analysis software, etc.)
- * What descriptive summaries of data are planned? Illustrative data, graphical summaries, descriptive statistics, etc.
- * Validity of description design
 - *Support for data preparation*
 - Will the prepared data represent the same phenomena as the unprepared data?
 - If data may be removed, would this be defensible beyond reasonable doubt?
 - Would your scientific opponents produce the same descriptions from the data?
 - *Support for data interpretation*
 - Will the interpretations that you produce be facts in your conceptual research framework? Would your scientific peers produce the same interpretations?
 - Will the interpretations that you produce be facts in the conceptual framework of the subjects? Would subjects accept them as facts?
 - *Support for descriptive statistics*
 - Is the chance model of the variables of interest defined in terms of the population elements?
 - *Repeatability:* Will the analysis repeatable by others?
 - *Ethics:* No new issues.

10.2 Statistical inference design

- * What statistical inferences are you planning to do? What data do they need? What assumptions do they make?
- * Statistical conclusion validity
 - Assumptions of confidence interval estimation
 - *Stable distribution.* Does X have a stable distribution, with fixed parameters?
 - *Scale.* Does X have an interval or ratio scale?
 - *Sampling.* Is sample selection random or does it contain a known or unknown systematic selection mechanism?
 - *Sample size.* If the z distribution is used, is the sample sufficiently large for the normal approximation to be used?
 - *Normality.* If the t -distribution is used, is the distribution of X normal, or is the sample size larger than 100?
 - *Treatment allocation.* Are the treatments allocated randomly to sample elements?
 - Avoid the following omissions in a report about difference-making experiments:
 - *Effect size.* Seeing a very small difference, but not telling that it is small
 - *Fishing.* Seeing no difference most of the time, but not telling this
 - *Very high power.* Not telling about a reason why you can see a difference (very large sample size makes very small differences visible)
 - *Sample homogeneity.* Not telling about another reason why you can see a difference (groups are selected to be homogeneous, so that any intergroup difference stands out)

10.3 Abductive inference design

- * What possible explanations can you foresee? What data do you need to give those explanations? What theoretical framework?
- * Internal validity
 - *Causal inference*
 - *Ambiguous relationship.* Ambiguous covariation, ambiguous temporal ordering, ambiguous spatial connection?
 - *OoS dynamics.* Could there be interaction among OoSs? Could there be historical events, maturation, and dropout of OoSs?
 - *Sampling influence.* Could the selection mechanism influence the OoSs? Could there be a regression effect?
 - *Treatment control.* What other factors than the treatment could influence the OoSs? The treatment allocation mechanism, the experimental setup, the experimenters and their expectations, the novelty of the treatment, compensation by the researcher, and rivalry or demoralization about the allocation?
 - *Treatment instrument validity.* Do the treatment instruments have the effect on the OoS that you claim they have?
 - *Measurement influence.* Will measurement influence the OoSs?
 - *Architectural inference*
 - *Analysis.* The analysis of the architecture may not support its conclusions with mathematical certainty. Components fully specified? Interactions fully specified?
 - *Variation.* Do the real-world case components match the architectural components? Do they have the same capabilities? Are all architectural components present in the real-world case?

- *Abstraction.* Does the architectural model used for explanation omit relevant elements of real-world cases? Are the mechanisms in the architectural model interfered with by other mechanisms, absent from the model but present in the real-world case?
- *Rational inference:*
 - *Goals.* An actor may not have the goals assumed by an explanation. Can you get information about the true goals of actors?
 - *Motivation.* A goal may not motivate an actor as much as assumed by an explanation. Can you get information about the true motivations of actors?

10.4 Analogic inference design

- * What is the intended scope of your generalization?
- * External validity
 - *Object of study similarity*
 - *Population predicate.* Will the OoS satisfy the population predicate? In which way will it be similar to the population elements? In which way will it be dissimilar?
 - *Ambiguity.* Will the OoS satisfy other population predicates too? What could be the target of analogic generalization?
 - *Representative sampling*
 - Sample-based research. Will the study population, described by the sampling frame, be representative of the theoretical population?
 - Case-based research. In what way will the selected/constructed sample of cases be representative of the population?
 - *Treatment*
 - *Treatment similarity.* Is the specified treatment in the experiment similar to treatments in the population?
 - *Compliance.* Is the treatment implemented as specified?
 - *Treatment control.* What other factors than the treatment could influence the OoSs? Could the implemented treatment be interpreted as another treatment?
 - *Measurement*
 - *Construct validity.* Are the definitions of constructs to be measured valid? Clarity of definitions, unambiguous application, and avoidance of mono-operation and mono-method bias?
 - *Measurement instrument validity.* Do the measurement instruments measure what you claim that they measure?
 - *Construct levels.* Will the measured range of values be representative of the population range of values?

At this point, the checklist for research design ends. From this point on, the checklist describes research execution and analysis. We switch from the future tense to the past tense because this part of the checklists asks questions about what has happened, not about what you plan to do.

Research Execution

11. What has happened?
 - What has happened when the OoSs were selected or constructed? Did they have the architecture that was planned during research design? Unexpected events for OoSs during the study?
 - What has happened during sampling? Did the sample have the size you planned? Participant flow, dropouts?
 - What has happened when the treatment(s) were applied? Mistakes, unexpected events?
 - What has happened during measurement? Data sources actually used, response rates?

Data Analysis

12. Descriptions
 - Data preparations applied? Data transformations, missing values, removal of outliers? Data management, data availability.
 - Data interpretations? Coding procedures, interpretation methods?
 - Descriptive statistics. Demographics, sample mean and variance? Graphics, tables.
 - Validity of the descriptions: See checklist for the validity of descriptive inference.
13. Statistical conclusions
 - Statistical inferences from the observations. Confidence interval estimations, hypothesis tests.
 - Statistical conclusion validity: See checklist for the validity of statistical inference.
14. Explanations
 - What explanations (causal, architectural, rational) exist for the observations?
 - Internal validity: See checklist for the validity of abductive inference.
15. Generalizations
 - Would the explanations be valid in similar cases or populations too?
 - External validity: See checklist for the validity of analogic inference
16. Answers
 - What are the answers to the research questions? Summary of conclusions, support for, and limitations of conclusions.

Research Context

17. Contribution to knowledge goal(s). Refer back to items 1 and 3.
18. Contribution to improvement goal(s)? Refer back to item 2.
 - If there is no improvement goal, is there a potential contribution to practice?