Fundamental empirical research principles

(version 1.0, 1/4/05)

Code: intro-methods

Daniel K. Schneider, TECFA, University of Geneva

Menu

1. The logic of empirical research 2
2. Objectives 4
3. Conceptualizations 5
4. The measure 16
5. Interpretation: validity (truth) and causality 21
6. Conclusion 26
1. The logic of empirical research

1.1 Elements of a typical research cycle

- Details may considerably change within a given approach
1.2 Key elements of empirical research

For a given research question, you usually do:

- **Conceptualisations**: make questions explicit, identify major concepts (variables), define terms and their dimensions, find analysis grids, define hypothesis, etc.
- **Artifacts**: develop research materials (experiments, surveys), implement software, etc.
- **Measures**: Observe (measure) in the field or through experiments (use your artifacts).
- **Analyses & conclusions**: Analyze the measures (statistic or qualitative) and link to theoretical statements (e.g. operational research questions and hypothesis).
2. Objectives

Research questions are the result of:

- your initial objectives (which you may have to revise)
- a (first) review of the literature

Everything you plan to do, must be formulated as a research question!

- See slides on "Finding a research subject"
3. Conceptualizations

- elaborate and "massage" concepts so that they can be used to study observable phenomena.
3.1 The usefulness of analysis frameworks

E.g. activity theory

Quote: The Activity Triangle Model or activity system representationally outlines the various components of an activity system into a unified whole. Participants in an activity are portrayed as subjects interacting with objects to achieve desired outcomes. In the meanwhile, human interactions with each other and with objects of the environment are mediated through the use of tools, rules and division of labour. Mediators represent the nature of relationships that exist within and between participants of an activity in a given community of practices. This approach to modelling various aspects of human activity draws the researcher's attention to factors to consider when developing a learning system. However, activity theory does not include a theory of learning, (Daisy Mwanza & Yrjö Engeström)

- Translation: It helps us thinking about a phenomenon to study.
- A framework is not true or false, just useful or useless for a given intellectual task!
3.2 Models and hypothesis

- These constructions link concepts and **postulate causalities**
- causalities between concepts (theoretical variables) do not "exist" per se, they only **can be observed indirectly**
- Typical statements: "More X leads to more Y", "an increase in X leads to a decrease in Y"

**Exemple 3-1: Causality between teacher training and quality**

Hypothesis (often heard): Continuous teacher training (cause X) improves teaching (Y)

- x= teacher training
do amount of training
- y= quality of teaching

- x= amount of training
- y= grades of pupils

- x= days of teacher training / year
correlation between cases
- y= average of grades / class / year

X= explanatory independent variable
Y= explained dependant variable

postulate
observed dimension
indicators: observations of N classes + statistical analysis

conclusions
3.3 The importance of difference (variance) for explanations

Without variance, no differences .... no explanatory science
    we’d like to know why things exist, why we can observe "more" and "less" ....

Without co-variance, no correlations / causalities ... no explanation

A. Quantitative example:

- We got different grade averages and different training days
  - therefore variance for both variables
- According to these data: increased training days lead to lower averages
  - (consider this hypothetical example false please !)
B. Qualitative example

Imagine that we wish to know why certain private schools introduce technology faster than others. One hypothesis to test could be: "Reforms need external pressure".

<table>
<thead>
<tr>
<th>Type of pressure</th>
<th>Strategies of a school</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strategy 1: no reaction</td>
</tr>
<tr>
<td></td>
<td>strategy 2: a task force is created</td>
</tr>
<tr>
<td></td>
<td>strategy 3: internal training programs are created</td>
</tr>
<tr>
<td></td>
<td>strategy 4: resources are reallocated</td>
</tr>
<tr>
<td>Letters written by parents</td>
<td>(N=4)</td>
</tr>
<tr>
<td></td>
<td>(p=0.8)</td>
</tr>
<tr>
<td>Letters written by supervisory</td>
<td>(N=2)</td>
</tr>
<tr>
<td>boards</td>
<td>(p=0.4)</td>
</tr>
<tr>
<td>newspaper articles</td>
<td>(N=1)</td>
</tr>
<tr>
<td></td>
<td>(p=100%)</td>
</tr>
</tbody>
</table>

N = number of observations, p = probability

• Result (imaginary): increased pressure leads to increased action
3.4 How can we measure general concepts?

A scientific proposition contains concepts (theoretical variables)
- Examples: “the learner”, “performance”, “efficiency”, “interactivity”

An academic piece links concepts
- ... empirical research requires that you work with data, find indicators, build indices, ..
- because of observed correlations we can make statements at the theory level

Exemple 3-2: “Collaborative learning improves pedagogical effect”:

- We got a real problem here! How could we measure "pedagogical effect" or "collaborative learning"?
A. The bridge/gap between theoretical concept and measure:

- There are 2 issues you must address:

(1) Going from “abstract” to “concrete” (theoretical concept - observables)

Examples:

- measure of “student participation” with “number of forum messages posted”
- measure of “pedagogical success” with “grade average of a class in exams”

(2) “whole - part” (dimensions):

Examples from educational design, i.e. dimensions you might consider when you plan to measure the socio-constructiveness of some teaching:

- Decomposition of “socio-constructivist design” in (1) active or constructive learning, (2) self-directed learning, (3) contextual learning and (4) collaborative learning, (5) teacher’s interpersonal behavior (Dolmans et. al)
- The Five Es socio-constructivist teaching model: Engagement, Exploration, Explanation, Elaboration and Evaluation (Boddy & al)

Example from public policy analysis:

- Decomposition of “economic development” in industrialization, urbanization, transports, communications and education.

Example from HCI:

- Decomposition of usability in "cognitive usability" (what you can achieve with the software) and "simple usability" (can you navigate, find buttons, etc.)
Exemple 3-3: COLLES Constructivist On-Line Learning Environment Survey (Taylor and Maor)

Dimensions (from teacher education over the Internet survey studies survey):

- **Relevance** How relevant is on-line learning to students' professional practices?
- **Reflection** Does on-line learning stimulate students' critical reflective thinking?
- **Interactivity** To what extent do students engage on-line in rich educative dialogue?
- **Tutor Support** How well do tutors enable students to participate in on-line learning?
- **Peer Support** Is sensitive and encouraging support provided on-line by fellow students?
- **Interpretation** Do students and tutors make good sense of each other's on-line communications?

Each of these dimensions is then measured with a few survey questions (items), e.g.:

<table>
<thead>
<tr>
<th>Statements</th>
<th>Almost Never</th>
<th>Seldom</th>
<th>Sometimes</th>
<th>Often</th>
<th>Almost Always</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Items concerning relevance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>my learning focuses on issues that interest me.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>what I learn is important for my professional practice as a trainer.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>I learn how to improve my professional practice as a trainer.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>what I learn connects well with my prof. practice as a trainer.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td><strong>Items concerning reflection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… I think critically about how I learn.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>… I think critically about my own ideas.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>… I think critically about other students' ideas.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>… I think critically about ideas in the readings.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Exemple 3-4: measure of economic development

- usage of official statistics
- (only part of the diagram is shown)
Exemple 3-5: measure of the strategic efficiency of a private distance teaching agency

- example taken from a french methodology text book (Thiétard, 1999)
B. Dangers and problems of concept operationalization

1. Gap between data and theory
   • Example: measure communication within a community of practice (e.g. an e-learning group) by the quantity of exchanged forum messages
   • (students may use other channels to communicate !)

2. You forgot a dimension
   • Example: measure classroom usage of technology only by looking at the technology the teacher uses e.g. powerpoint, demonstrations with simulation software or math. software
   • (you don’t take into account technology enhanced student activities )

3. Concept overloading
   • Example: Include “education” in the definition of development
     (it could be done, but at the same you lose an important explanatory variable for development, e.g. consider India’s strategy that "overinvested" in education with the goal to influence on development)
   • Therefore: never ever collapse explanatory and explainable variables into one concept !!

4. Bad measures
   • (see later)
4. The measure

- observe properties, attributes, behaviors, etc.
- select the cases you study (sampling)
4.1 Sampling

As a general rule:

- Make sure that "operative" variables have good variance, otherwise you can’t make any statements on causality or difference .....  
- operative variables = dependant (to explain) and independent (explaining) variables

Overview on sampling strategies

<table>
<thead>
<tr>
<th>Type of selected cases</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>maximal variation</td>
<td>will give better scope to your result (but needs more complex models, you have to control more intervening variables, etc. !!)</td>
</tr>
<tr>
<td>homogeneous</td>
<td>provides better focus and conclusions; will be &quot;safer&quot; since it will be easier to identify explaining variables and to test relations</td>
</tr>
<tr>
<td>critical</td>
<td>exemplify a theory with a &quot;natural&quot; example</td>
</tr>
<tr>
<td>according to theory, i.e. your research questions</td>
<td>will give you better guarantees that you will be able to answer your questions ....</td>
</tr>
<tr>
<td>extremes and deviant cases</td>
<td>test the boundaries of your explanations, seek new adventures</td>
</tr>
<tr>
<td>intense</td>
<td>complete a quantitative study with an in-depth study</td>
</tr>
</tbody>
</table>

- sampling strategies depend a lot on your research design!
4.2 Measurement techniques

- There are not only numbers, but also text, photos and videos!
- Not treated here, see the modules «ix Quantitative data acquisition methods (e.g. surveys and tests)» and «x Qualitative data acquisition methods (e.g. Interviews and observations)»!

Principal forms of data collection

<table>
<thead>
<tr>
<th>Situation</th>
<th>non-verbal and verbal</th>
<th>Articulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>informal</td>
<td>participatory observation</td>
<td>information interview</td>
</tr>
<tr>
<td>formal and unstructured</td>
<td>systematic observation</td>
<td>open interviews, semi-structured interviews, thinking aloud protocols, etc.</td>
</tr>
<tr>
<td>formal and structured</td>
<td>experiment simulation</td>
<td>standardized interview, standardized questionnaire, log files of structured user interactions,</td>
</tr>
</tbody>
</table>
4.3 Reliability of measure

\[ \text{reliability} = \text{degree of measurement consistency for the same object} \]

1. by different observers
2. by the same observer at different moments
3. by the same observer with (moderately) different tools

**example: measure of boiling water**

- A thermometer always shows 92 C. => it is reliable (but not construction valid)
- The other gives between 99 and 101 C.: => not too reliable (but valid)

**Sub-types de reliability (Kirk & Miller):**

1. circumstantial reliability: even if you always get the same result, it does not means that answers are reliable (e.g. people may lie)
2. diachronic reliability: the same kinds of measures still work after time
3. synchronic reliability: we obtain similar results by using different techniques, e.g. survey questions and item matching and in depth interviews

**In short:** can we reproduce and replicate, can we trust data?
The “3 Cs” of an indicator

**Are your data complete?**
- Sometimes you lack data ....
- Try to find other indicators

**Are your data correct?**
- The reliability of indicators can be bad.
  - Example: Software ratings may not mean the same
  - according to cultures (sub-cultures, organizations, countries) people are more or less outspoken.

**Are your data comparable?**
- The meaning of certain data are not comparable.
  - examples:
    (a) School budgets don’t mean the same thing in different countries (different living costs)
    (b) Percentage of student activities in the classroom don’t measure "socio-constructive" sensitivity of a teacher (since there a huge cultural differences between various school systems)
5. Interpretation: validity (truth) and causality

• Can you really trust your conclusions
• Did you misinterpret statistical evidence for causality?
5.1 The role of validity

- Validity (as well reliability) determine the formal quality of your research
- More specifically, validity of your work (e.g. your theory or model) is determined by the validity of its components.

In other words:
- can you justify your interpretations ??
- are you sure that you are not a victim of your confirmation bias ?
- can you really talk about causality (or should you be more careful) ?

Note: Validity is not the only quality factor

<table>
<thead>
<tr>
<th>Judgements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theories</td>
</tr>
<tr>
<td>Models (“frameworks”)</td>
</tr>
<tr>
<td>Hypotheses and models</td>
</tr>
<tr>
<td>Methodology (&quot;approach&quot;)</td>
</tr>
<tr>
<td>methods</td>
</tr>
<tr>
<td>Data</td>
</tr>
</tbody>
</table>

A good piece of work satisfies first of all an objective, but it also must be valid
The most important usefulness criteria is: "does it increase our knowledge"

The most important formal criteria are validity and reliability

Somewhere in between: "Is your work coherent and well constructed"?
5.2 Some reflections on causality

- A correlation between 2 variables (measures) does not prove causality
  - Co-occurrence between 2 events does not prove that one leads to the other
  - The best protection against such errors is theoretical and practical reasoning!

example:
- “We introduced ICT in our school and student satisfaction is much higher”
- (It’s maybe not ICT, but just a reorganization effect that had impact on various other variables such as teacher-student relationship, teacher investment, etc.)

- If you observe correlations in your data and you are not sure, talk about association and not cause!

- Even if can provide sound theoretical evidence for your conclusion, you have the duty to look a rival explanations!

  - There are methods to test rival explanations (see modules on data-analysis)
Some examples of bad inference

- Simple hidden causalities

**Situation 1: Badly interpreted correlation (no real link between A and B)**

Example from European folklore:
“Là où on voit des cigognes, il y a plus de naissances” naissances (A) et cigognes (B) sont un effet (négatif) de l’urbanisation (C)

**Situation 2: Double effect**

A and B are explained by C and/or the relation between A and B is function of C

Example:
“more parent meetings increase spending for ICT”
School’s organizational culture (C) influences participation (A) and spending (B)

**Situation 3: Causality chain**

Example: “Young teachers hate ICT”
Young teachers (A) use less ICT (C) because they have less resources (time) left to spend on it (B)

- Think!
6. Conclusion

6.1 Some advice

At every stage of research you have to think and refer to theory:

• Good analytical frameworks (e.g. instructional design theory or activity theory) will provide structure to your investigation and will allow you to focus on essential things.

• You can’t answer your research question without a serious operationalization effort.
  • Identify major dimensions of concepts involved, use good analysis grids!

Watch out for validity problems

• You can’t prove a hypothesis (you only can test, reinforce, corroborate, etc.).
  • Therefore, also look at anti-hypotheses!

• Good informal knowledge of a domain will also help
  • Don’t hesitate to talk about your conclusions with a domain expert

• Purely inductive reasoning approaches are difficult and dangerous.
  • ... unless you master an adapted (costly) methodology, e.g. "grounded theory"

You have a “confirmation bias”!

• humans tend to look for facts that confirm their reasoning and ignore contradictory elements

• It’s your duty to test rival hypothesis (or at least to think about them)!

Attempt some (but not too much) generalization

• show the others what they can learn from your piece of work, confront your work to other’s!
6.2 Choice and complementarity of methods

A. Triangulation of methods

- **Different viewpoints (and measures) can consolidate or even refine results**
  
  - E.g. imagine that you (a) led a quantitative study about teacher’s motivation to use ICT in school or (b) that you administered an evaluation survey form to measure user satisfaction of a piece of software.
  
  - You then can run a cluster analysis through your data and identify major types of users
    - (e.g. 6 types of teachers or 4 types of users).
  
  - Then you can do in-depth interviews with 2 representatives for each type and "dig" in their attitudes, subjective models, abilities, behaviors, etc. and confront these results with your quantitative study.

B. Theory creation v.s theory testing

- **Qualitative methods are better suited to create new theories**
  
  - (exploration / comprehension)

- **Quantitative methods are better suited to test / refine theories**
  
  - (explication / prediction)

... but:

- validity, causality, reliability issues ought to be addressed in any piece of research
- it is possible to use several methodological approaches in one piece of work