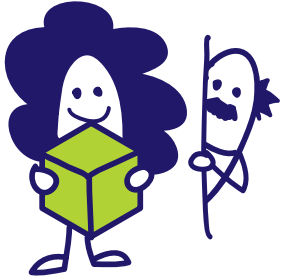# Data collection methods

Natalia Romero Herrera &
Pieter Jan Stappers &
Special guest: Jeff Love

Me: research through & design for big/small data interactions in daily life

Assistant professor in HCI/UCD/UX at ID-DCC

Research keywords
- #ExperienceSampling – active reporting
- #MixedMethods – connecting objective & subjective data
- #LivingLabs – user as co-collaborators
- #AppliedResearch – research through design in the wild

Areas of interest
- – Healthcare: eHealth, self-management
- – Energy: saving & wellbeing (home/work/school)

Example projects

- ENERGE (InterregNWE, 2019-2023)
  - How to activate secondary schools pupils, teachers and admin staff to manage their energy consumption and impact on wellbeing
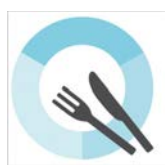
FoodSampler (ZonMw, 2018-2020)
  - How to activate people (Low SES) in gathering/using food knowledge & skills (general/own) ?

BOCS (Climate-KIC, 2015-2017)
  - How to engage office workers to provide/consume data on indoor climate and wellbeing ?

SUSLAB (INTERREG, 2012-2014)
  - How to engage social home residents to gather/use data on indoor climate, behaviours and wellbeing ?

Today's plan

What observation techniques are available?

What methods will you use and why?

PART 1

What are the requirements
for doing high quality research?

How do you ensure that you meet the requirements?

PART 2

How to publish your data?

Make a data management plan

PART 3

PART 1

What data collection techniques do exist?

Qualitative approaches

- Case study research

- Interview studies

- Ethnographic studies

Synopsis Watergebruik Thuis 2013

In opdracht van Vewin heeft TNS Nipo onderzoek gedaan naar het watergebruik thuis onder 1.349 respondenten en de resultaten weergegeven in het rapport 'Watergebruik Thuis 2013'.
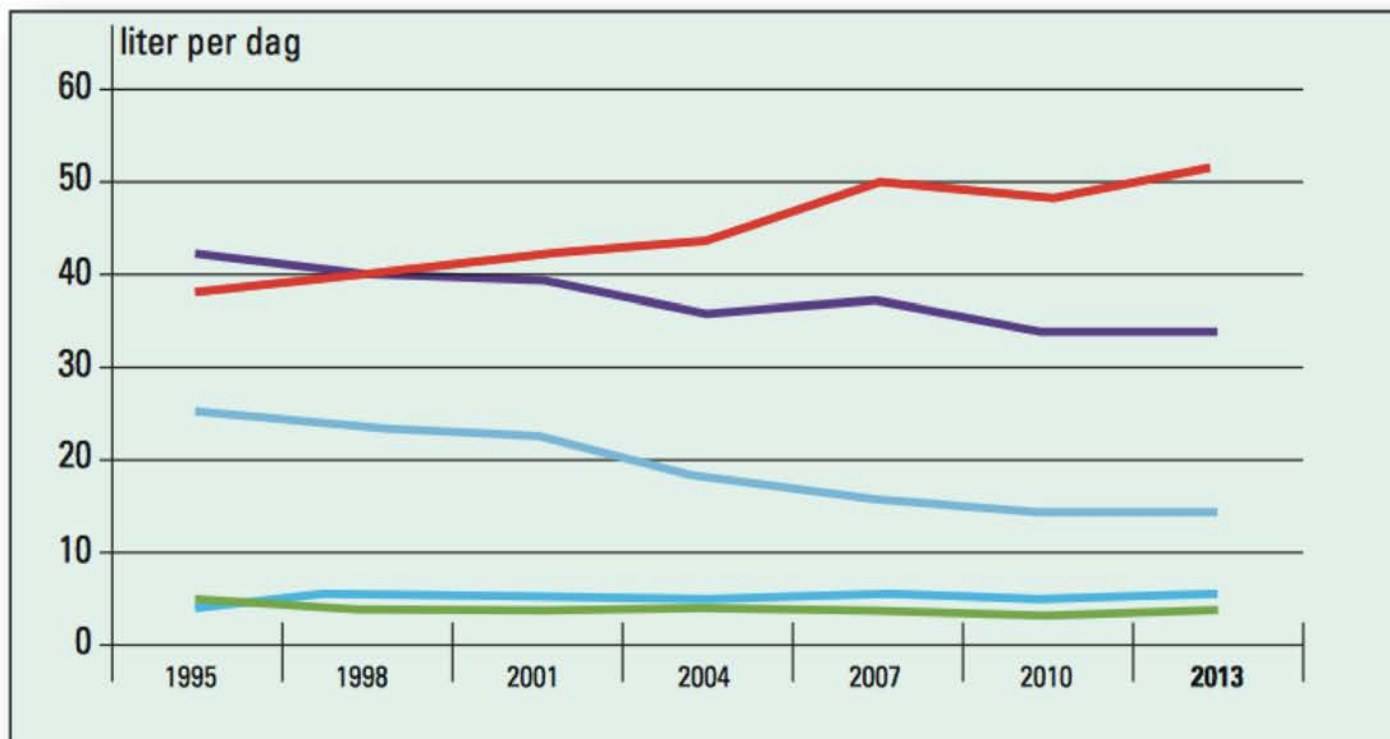
Quantitative survey

Synopsis Watergebruik Thuis 2013

Verschuivingen top 5

In 1995 en 1998 stond de toiletspoeling nog stipt op nummer 1 als grootste water-gebruiker. Vanaf 2001 is dat het watergebruik in de douche. Ook in 2013 is dit de grootste watergebruiker. Sinds 1995 is het douchewatergebruik met 34 procent toegenomen.

Figuur 2

Hoofdelijk gebruik in huis naar toepassing (liter/dag)

- Douche
- Toiletspoeling
- Wasmachine
- Wastafel
- Afwassen, hand

Design Assignment

- Develop design solutions (policies, product/service) to reduce the water consumption at home

Solutions

- National campaigns to shower for a max of 5 min
- Product: shower timers

Type of study

Study goal

- Add to knowledge base
- Have an impact
- Measure change
- Understand complex phenomena
- Test new ideas
- Generate new ideas
- Inform constituencies
- Examine the past

Research objective

- Exploration
- Description
- Explanation
- Prediction
- Influence

Onwuegbuzie, A., & Leech, N. (2006) Linking research Questions to Mixed Methods Data Analaysis Procedures. The Qualitative Report, 11 (3), pp.474-498

Type of study

Study goal
- Add to knowledge base
- Have an impact
- Measure change
- Understand complex phenomena
- Test new ideas
- Generate new ideas
- Inform constituencies
- Examine the past

Research objective
- Exploration
- Description
- Explanation
- Prediction
- Influence

Onwuegbuzie, A., & Leech, N. (2006) Linking research Questions to Mixed Methods Data Analaysis Procedures. The Qualitative Report, 11 (3), pp.474-498

Type of study

Study goal
– Add to knowledge base
– Have an impact
– Measure change
– Understand complex phenomena
– Test new ideas
– Generate new ideas
– Inform constituencies
– Examine the past

Research objective
– Exploration
– Description
– Explanation
– Prediction
– Influence

Qualitative approaches

- Case study research
  - A diary with 10 households to describe and explain the activity

- Interview studies
  - An interview to go through their showering activities and experiences of that morning (or the day before)

- Ethnographic studies
  - Video taping people showering (!)

Observing people in their natural surroundings

Ethnography provides researchers and their clients the opportunity to watch people in their 'natural habitat'. Why is that so useful? And is it possible to generalise from observing the behaviour of just a few people? By Hy Mariampolski

Is it possible?

The fact that we videotaped people while showering always attracts questions. Colleagues expect that it was difficult to recruit participants but that it was enjoyable to watch and analyze the tapes. Actually, the opposite was true: Participants were easily recruited among people comfortable with their bodies; however, the painstaking analysis of behavioral details, divided into 5-second intervals, made watching the tapes rather tiresome.

Engaging people in their natural habitats became a source of profound inspiration and excitement barely two decades ago among clients who felt alienated behind mirrored windows in viewing studios and remotely listening to phone interviews. Focus groups seemed like a scrum for recognition among

Observing people in their natural surroundings

Ethnography provides researchers and their clients the opportunity to watch people in their 'natural habitat'. Why is that so useful? And is it possible to generalise from observing the behaviour of just a few people? By Hy Mariampolski

Is it worth?

Ethnography in the shower

I don't think there will ever be another study as meaningful and controversial as the one we conducted for the Moen Corporation that led to the development of their "Revolution Shower" – the winner of a 2005 IDEA Award for Innovation from *Business Week* and Moen's most successful new product introduction in its history.

Engaging people in their natural habitats became a source of profound inspiration and excitement barely two decades ago among clients who felt alienated behind mirrored windows in viewing studios and remotely listening to phone interviews. Focus groups seemed like a scrum for recognition among

Observing people in their natural surroundings

Ethnography provides researchers and their clients the opportunity to watch people in their 'natural habitat'. Why is that so useful? And is it possible to generalise from observing the behaviour of just a few people? By Hy Mariampolski

Is it ethical?

We went from the behavioral observation that people spend most of their time in the shower interacting with water in a playful and relaxing way to the cultural understanding that the shower stall had gone beyond the utilitarian functionality of cleanliness to a place for indulgence, escape and sensual pleasure. We used this idea to plan and develop a showerhead positioned as the perfect tool for luxurious interaction with water.

Engaging people in their natural habitats became a source of profound inspiration and excitement barely two decades ago among clients who felt ienated behind mirrored windows in viewing studios d remotely listening to phone interviews. Focus groups seemed like a scrum for recognition among

Qualitative methods

Qualitative methods enable gathering an in-depth understanding of human behavior and the reasons for that behavior.

Qualitative studies enable a researcher to study contemporary phenomena in a real-life setting, where boundaries between context and phenomenon tend to be blurred (Gibbert and Ruigrok, 2010)

Qualitative research provide rich insights about the *what, why, where*, *when* and *how*.

Main challenge: How to deal with subjectivity; assessing the rigor of qualitative work is not standardized (Gibbert and Ruigrok, 2010)

Recommended books on doing qualitative research
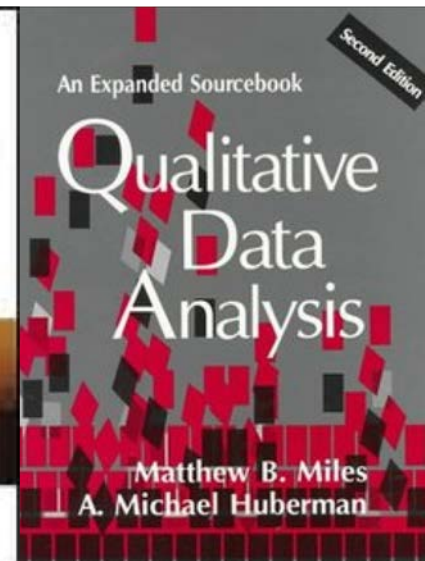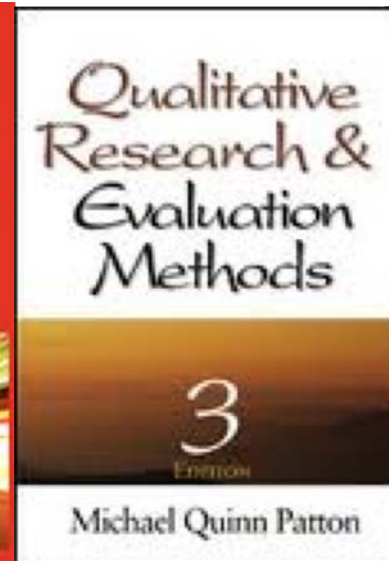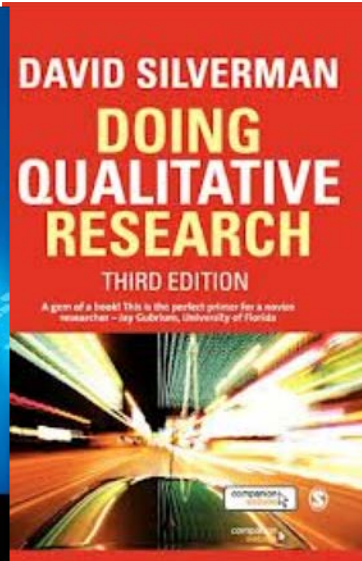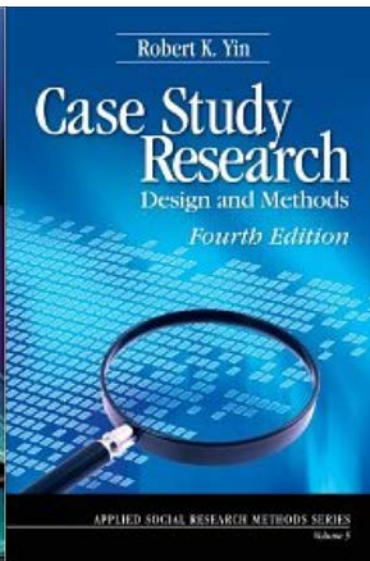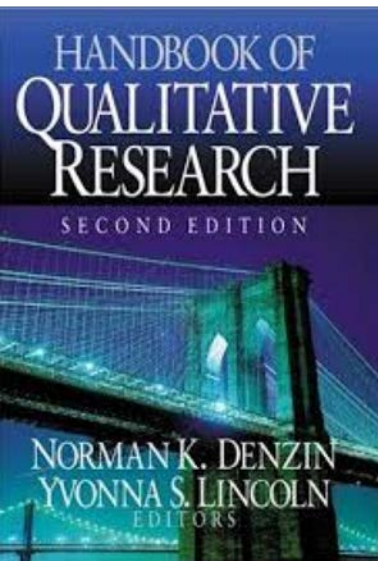
Quantitative approaches

- Surveys
  - A questionnaire with demographics, since when they have a shower, type of shower, etc.
  - A one week diary to annotate frequency and duration of shower

- Experiments
  - Condition 1: use an egg timer to set a 5 minutes shower. Condition 2: family member notifies that 5 minutes has passed.
  - Hypothesis: family member is more effective than egg timer

- Big data (bases)
  - Measure water use (sensor) during a full year
  - Find correlation between water use and outdoor temperature

Quantitative methods

The objective of quantitative research is to develop and employ mathematical models, theories and/or hypotheses pertaining to phenomena.

Quantitative studies enable quantifying data and generalize results from a sample to the population

Quantitative studies provides insights the incidence of various views and opinions in a chosen sample

Main challenge: to include all relevant variables

Trap1: Correlation and causation

- Variables that co-vary need not be causally linked
- For causation you need manipulation/intervention

- (it's tricky!)

Trap2: Sampling and generalization

- E.g., interviewing not all people
- E.g., selecting variables that have 'good' outcomes

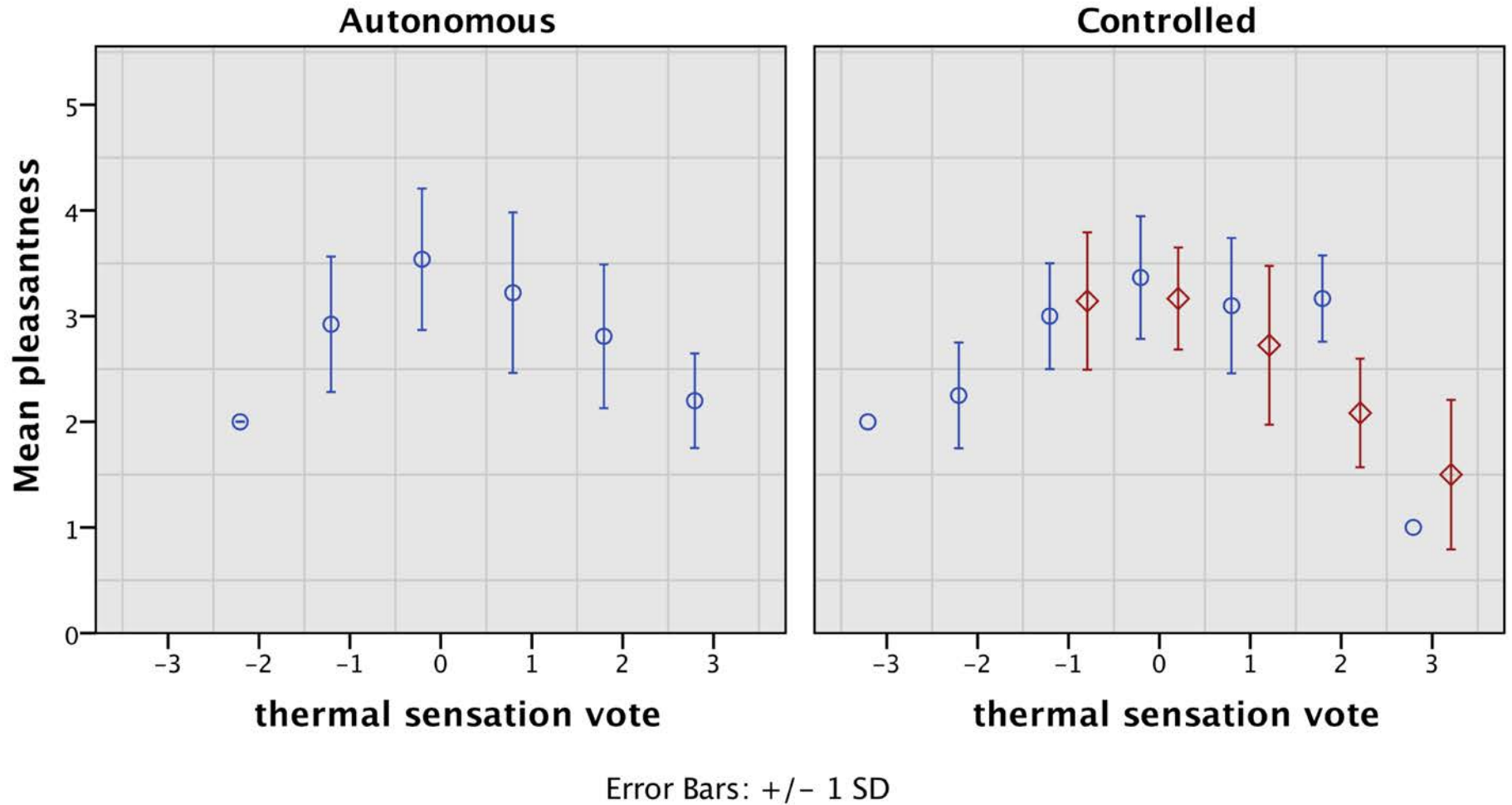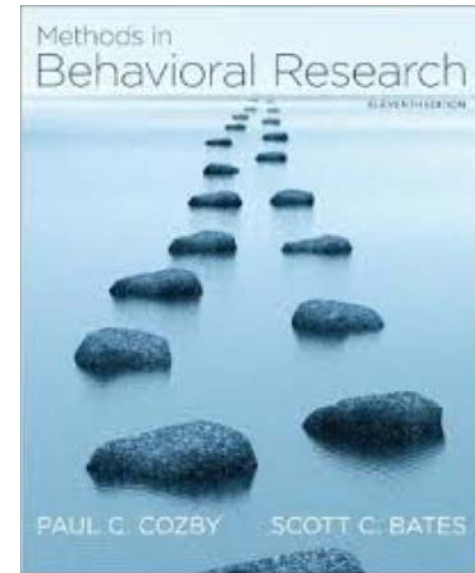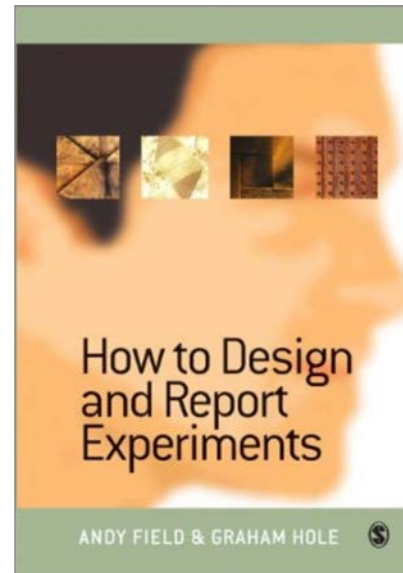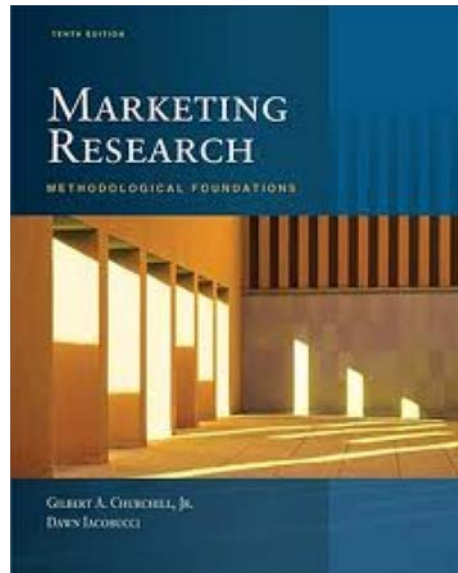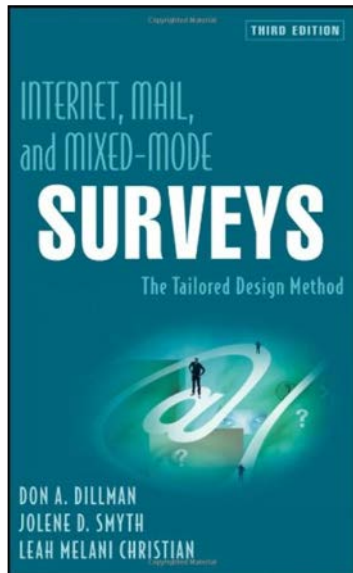Fig. 8: Means of pleasant votes calculated for each thermal sensation vote; diamonds represent warm rooms
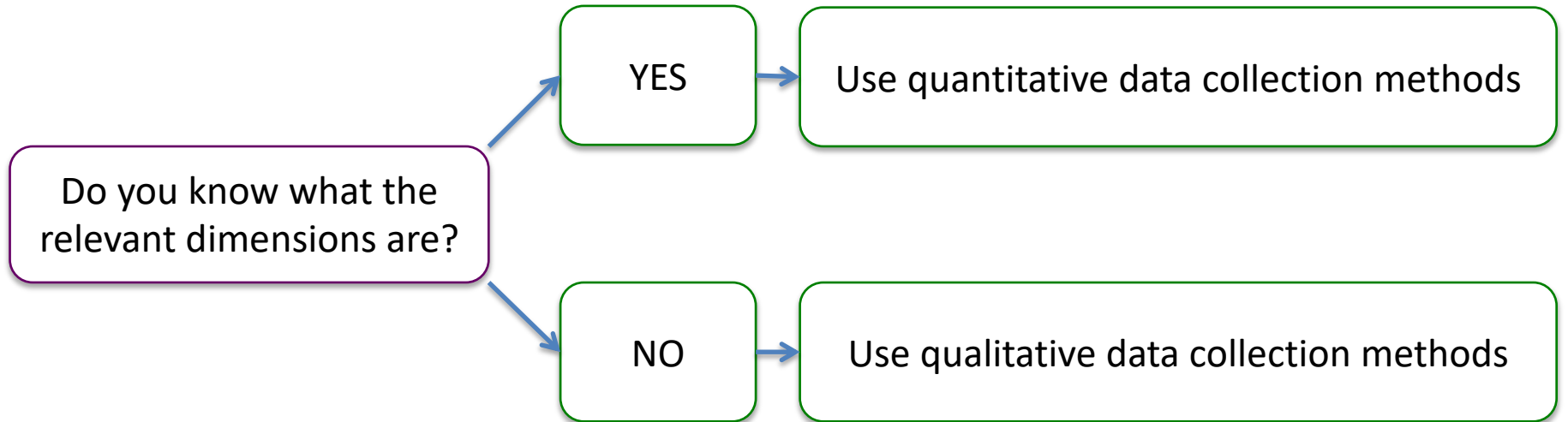
Recommended books quantitative research

Do you know what the relevant dimensions are?

YES → Use quantitative data collection methods

NO → Use qualitative data collection methods
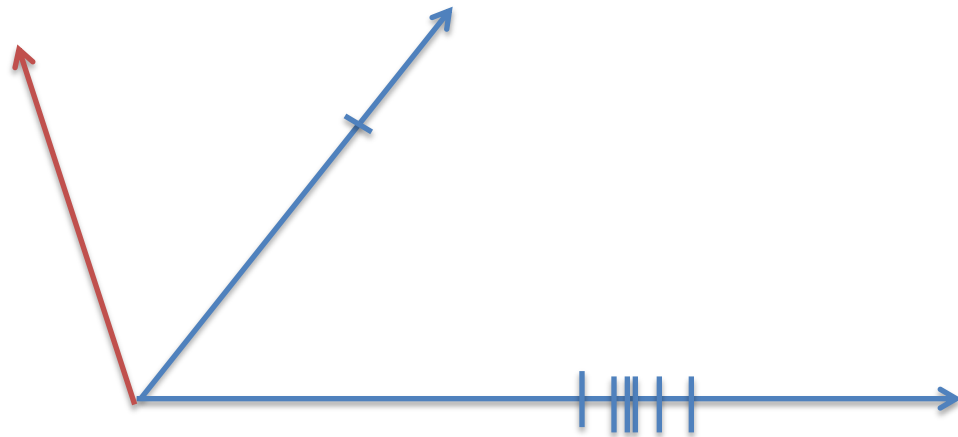
Alternating Methods

- Quantitative, then Qualitative

- Qualitative, then Quantitative

e.g., 'Freshness'

Alternating Methods

- Quantitative gives values along a known dimension

- Qualitative helps find relevant dimensions

e.g., 'Freshness'

Either/or vs. and/with

QUALITATIVE (diverging)	MIXED METHODS (combination)
MESS	QUANTITATIVE (converging)

Either/or vs. and/with

QUALITATIVE (diverging)	MIXED METHODS (combination)
MESS	QUANTITATIVE (converging)

Doing qualitative IS NOT not doing quantitative
Doing mixed methods means satisfying BOTH traditions

Mixed Methods

Why to use Mixed Methods?

When to use Mixed Methods?

What is needed to use Mixed Methods?

How to use Mixed Methods?

Mixed Methods: WHY?

Unites both quantitative and qualitative research

Provides more comprehensive evidence for studying a research problem

Encourages the use of multiple worldviews or paradigms.

Is "practical" in the sense that the researcher is free to use all methods possible to address a research problem.

TO CONTEXTUALISE KNOWLEDGE

- To contextualise energy savings
- To contextualise (THR) surgery recovery
- To contextualise food behaviour …
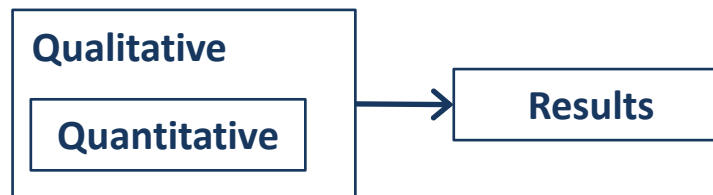
Mixed Methods: WHAT & HOW?

| Merge the data | Quantitative → Results ← Qualitative |

| Connect the data | Quantitative → Qualitative → Results |

| Connect the data | Qualitative → Quantitative → Results |

| Embed the data | Qualitative [Quantitative] → Results |

Mixed Methods: WHEN?

When one data source seems to be insufficient

Quantitative results need to be explained

Qualitative results need to be generalized

A second method is needed to enhance a primary method

An overall research objective can be best addressed with multiple phases or projects
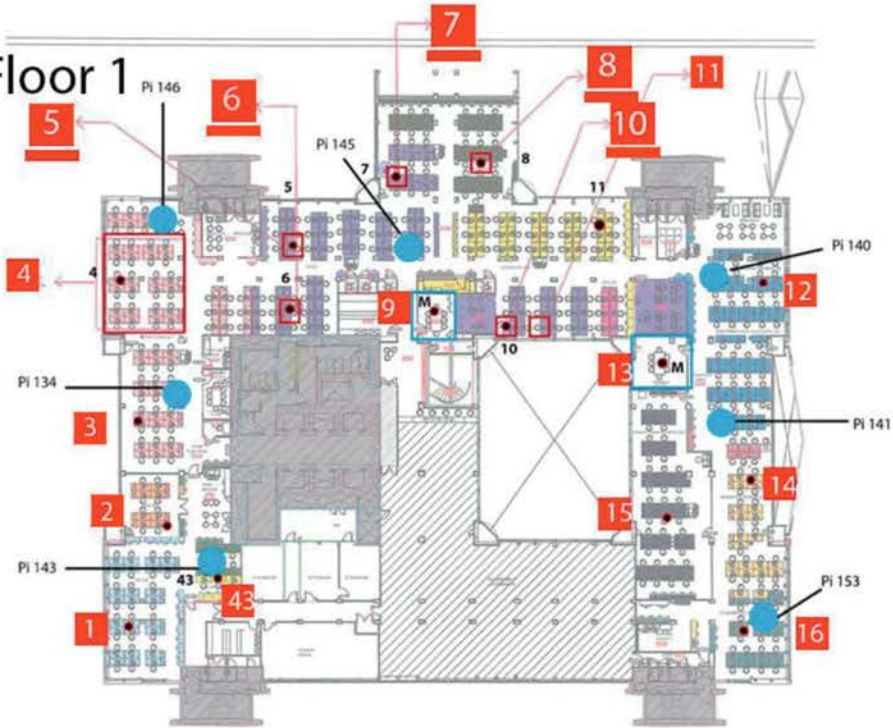
Office occupants as active actors in assessing and informing comfort: a context-embedded comfort assessment in indoor environmental quality investigations
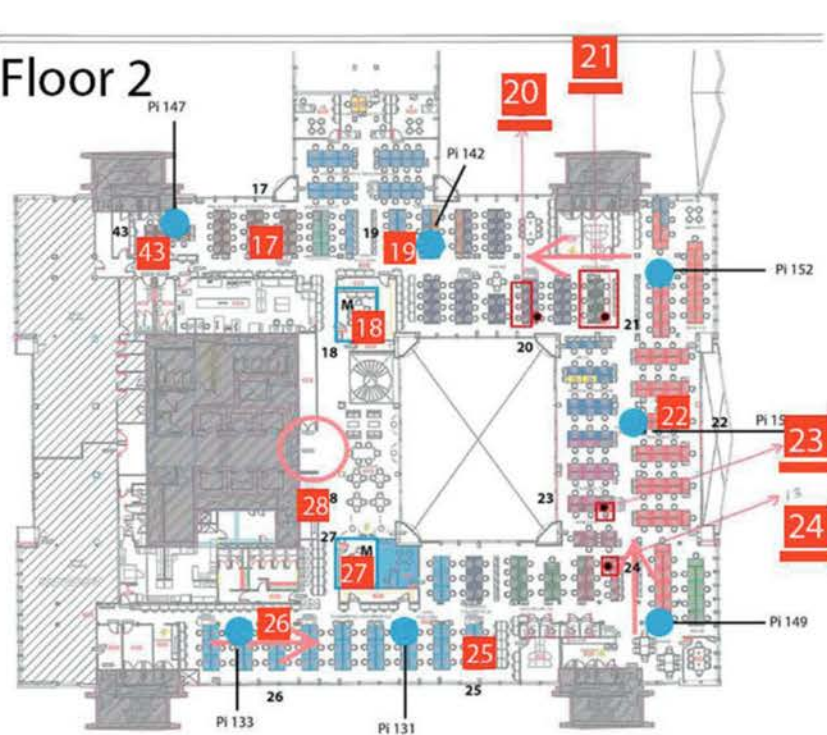
Natalia Romero Herrera, Jantien Doolaard, Olivia Guerra-Santin, Tomasz Jaskiewicz & David Keyson
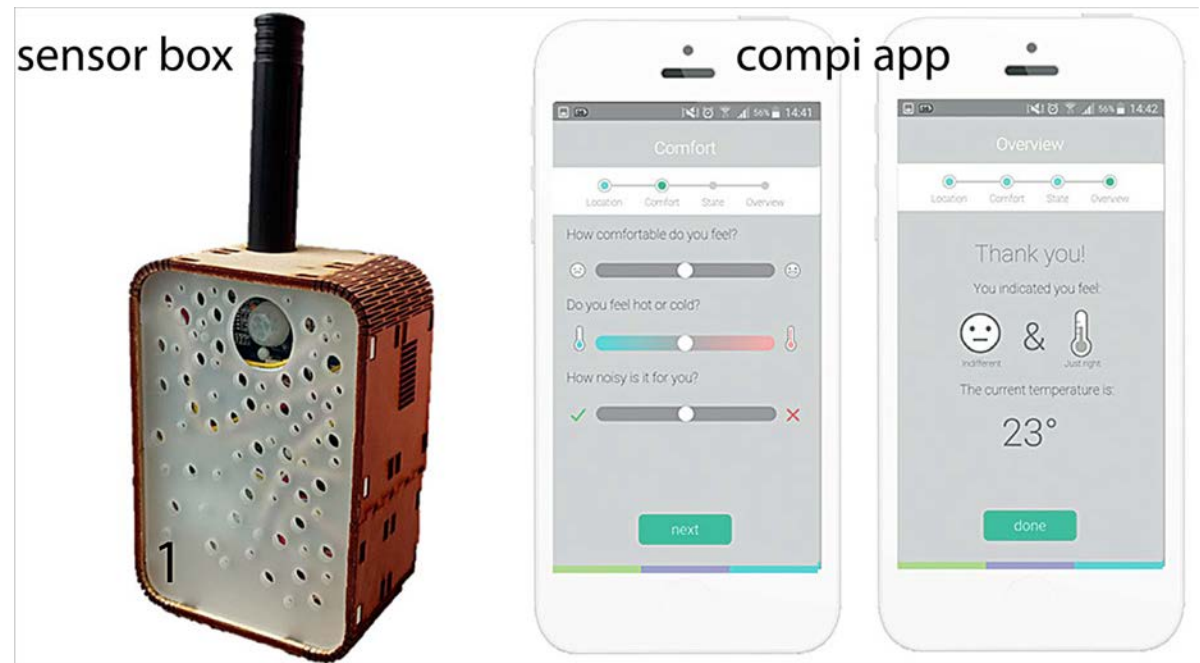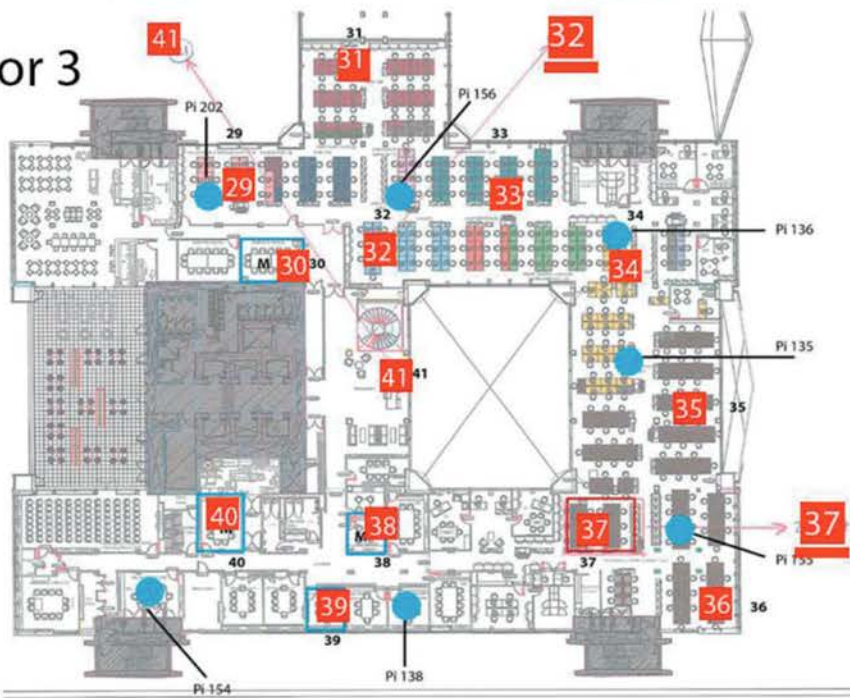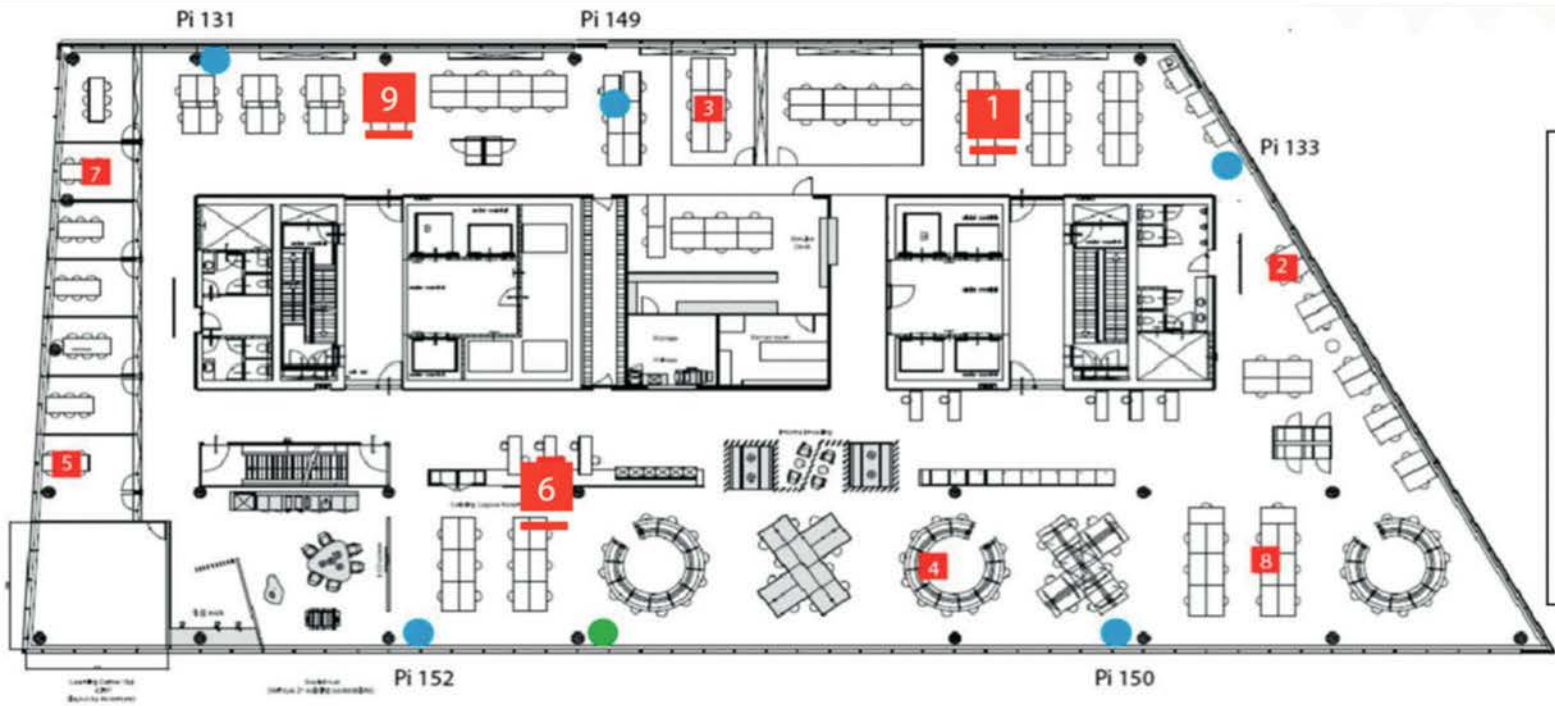
Pi 131 Pi 149 Pi 133 Pi 152 Pi 150

Legend:

1 Sensor box 9x

— Participants

● Rasberry Pi's 5x

● Internet router

Table 2: Data sources and variables used in the analysis

Qualitative	Variables	Values
Survey	experiences with self-reporting	preferences, suggestions
Email communication	experiences with platform and study	issues, complains

Quantitative	Variables	Values
Self-reported data [1]	thermal sensation vote	-3 (very cold) ... 0 ... 3 (very warm)
	pleasantness vote	1 (unpleasant) ... 5 (pleasant)
	noise level vote	1 (not noisy) ... 5 (noisy)
Monitored data [1]	Indoor temperature, CO_2, humidity, light and movement	numeric scale

[1] analysis focuses on working days (Monday to Friday) and working hours (7 to 19)
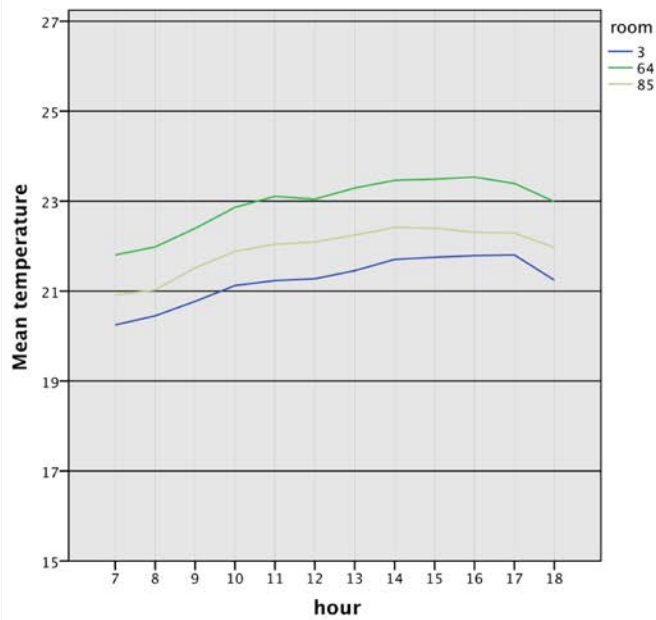
(c) A-mean hourly temperature
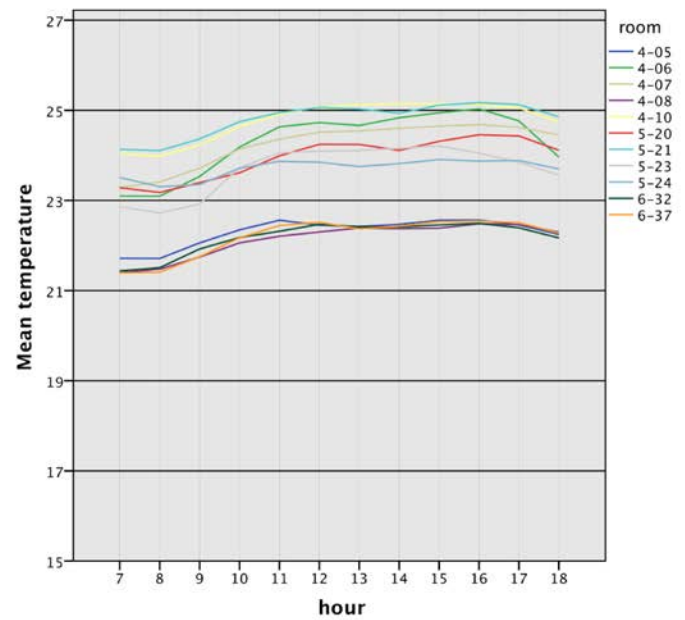
(d) B-mean hourly temperature

(e) thermal daily profiles

(f) thermal hourly profiles

Error Bars: +/− 1 SD
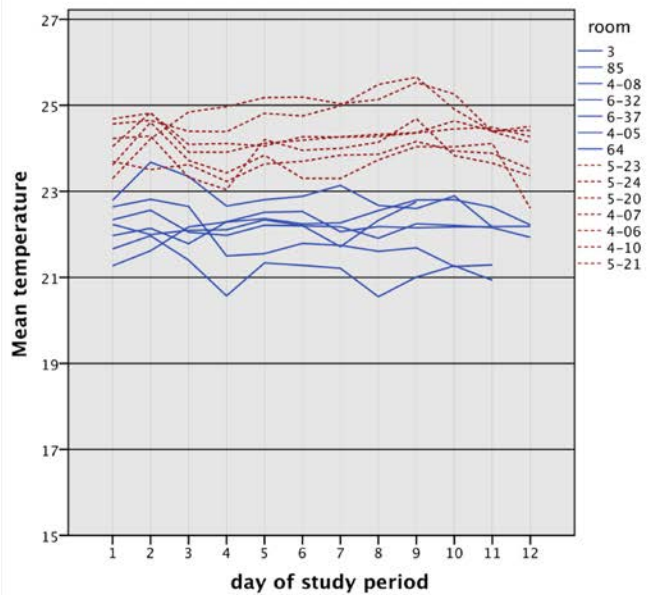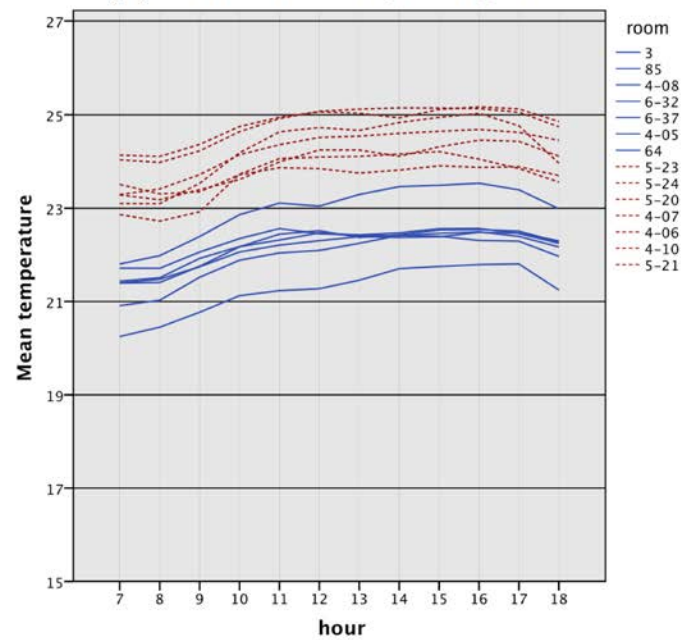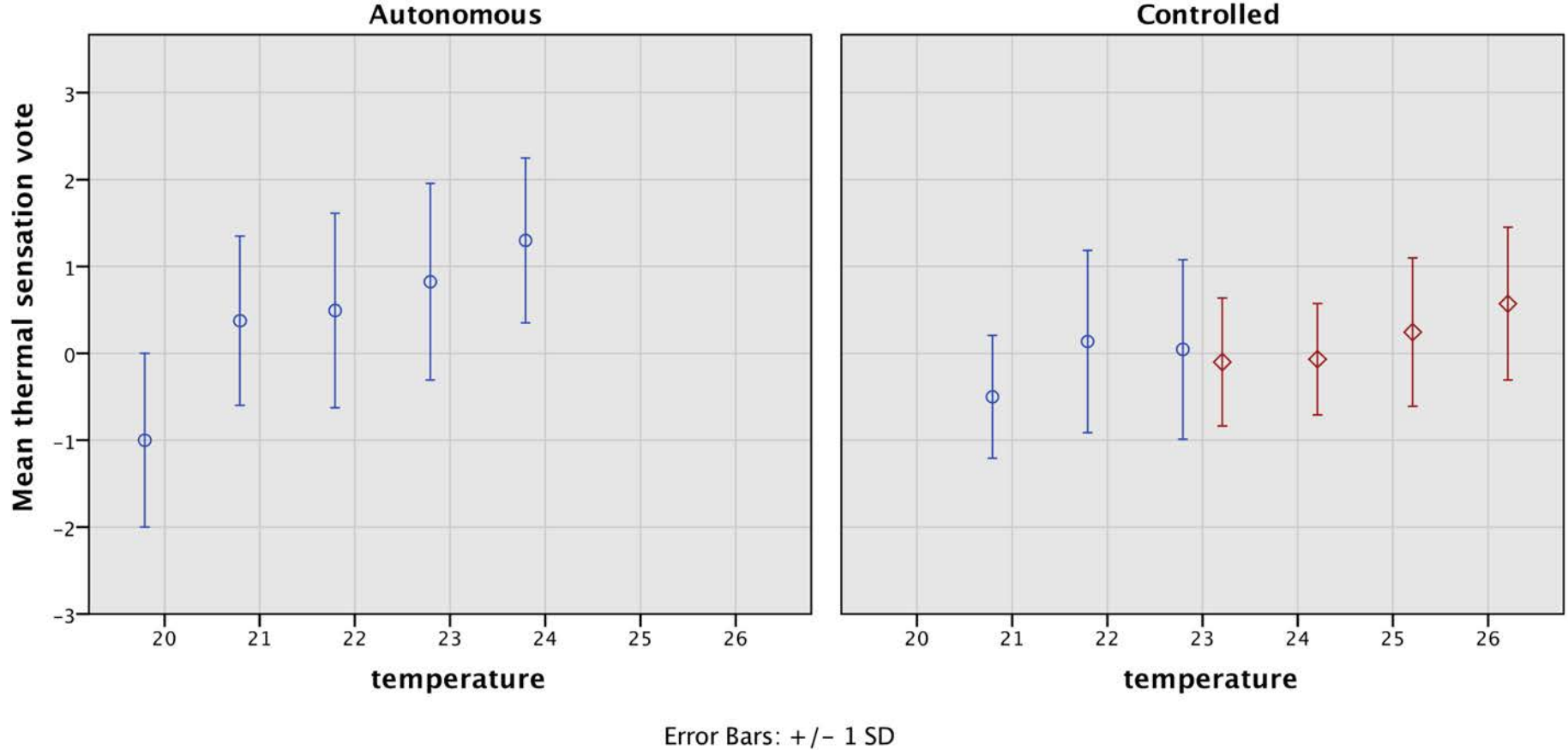
Fig. 6: Mean TSV with error bars representing the variation of the mean per "bins" of 1 degree of indoor temperature; diamonds represent warm rooms

Table 5: Summary of the working culture, infrastructure of each case study

	Building A	Building B
Working culture	*Flexible working style*; 1-2 days working from home or in other buildings *Type of work*; collaboration and consultation *Autonomous supportive*; employees feel responsible for their individual tasks; they decide when and where to work (as long as the goals are met) No formal break schedule; breaks are important and employees try to synchronise them so they become social activities; lunch is mostly a social event *Sustainable responsibility*; company separates garbage and make employees aware of their water usage in toilets	*Fixed working style*; full time working at the office *Type of work*; individual tasks mostly *Controlled oriented*; employees need to show their willingness to work; rushed and busy atmosphere; formal and hierarchical assignation of tasks *No formal break schedule*; breaks and lunch breaks are mostly an individual activity *Sustainable responsibility*; company asks employees to separate their waste
Working infrastructure	*Flexible desk*; open workspace; flex desk (employees select their working area; most of the time is in the same place); use of laptops *Other facilities*; food canteen; meeting and calling rooms; coffee corners; vending machine with office suppliers; terrace with bird sounds	*Fixed desk*; open workspace with fixed desks; use of desktop computers *Other facilities*; food canteen; meeting rooms; coffee corner; private working areas
Involvement in BOCS project	*Bottom up*; the company is a non-funded partner of the project; employees have been involved in earlier phases (8 of them being part of this study); they are able to express questions and concern about the project at all times; they could withdraw from the project at any time	*Top down*; employees have been involved in earlier evaluations of the platform (none of them participated in this study); employees have been requested to participate by the secretaries of their department; they could withdraw from the project at any time
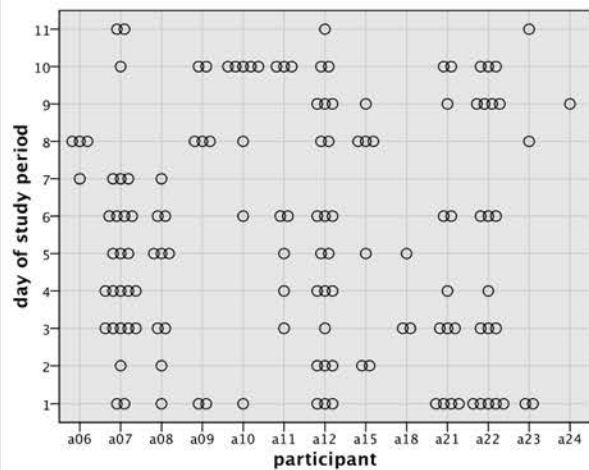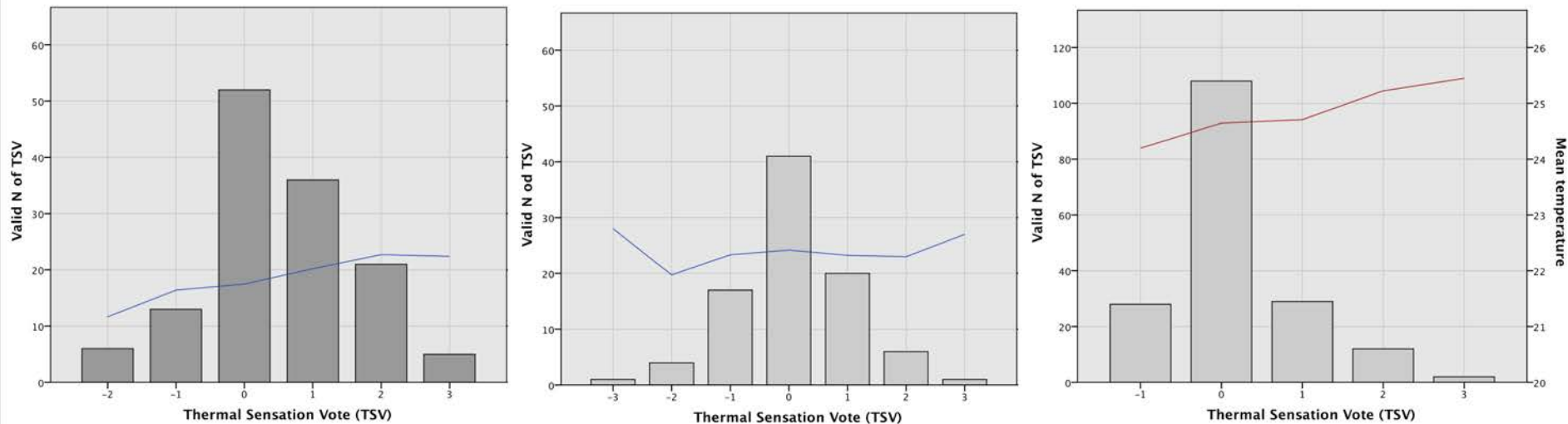
(a) auton-comf
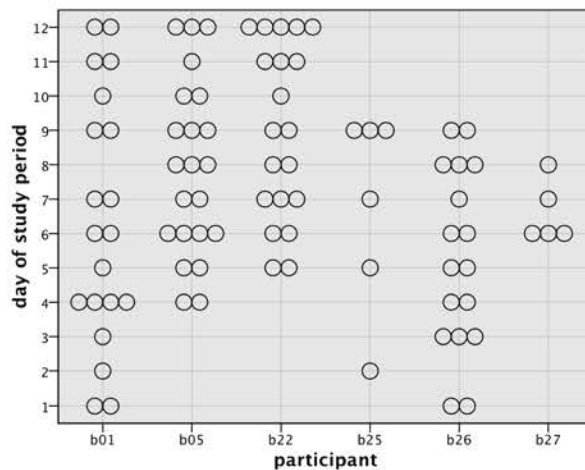
(b) controlled-comf
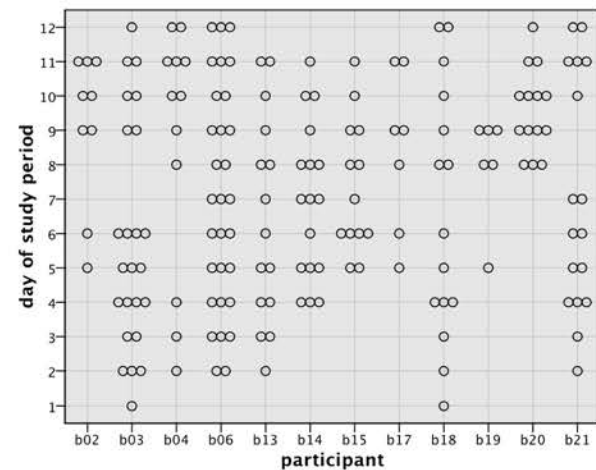
(c) controlled-warm

Fig. 9: 2-D dot plots of participants daily frequency overtime: number of circles in a cell represent the number of reports from participant x in day y ('null' or no reporting is represented by an empty cell)

Reporting behaviours

- **sample representation**, referring to the distribution of reports across participants/rooms
- **longitudinal representation**, defined as the distribution of reports across participants/time
- **voting representation**, defined as the distribution of reports across the voting scale (see Figure 7).

The sample and longitudinal representation of the reporting behaviour describe the **completeness** of the dataset (number of participants reporting at different times and days) and the voting representation impacts the **granularity** of the dataset (number of votes across the scale).

AUTONOMOUS	CONTROLLED
an inconsistent and low sampling representation as a small group of the active sample of participants contributed to the majority of reports (half of reports were reported by less than 25% of autonomous reporters)	a consistent and high sampling representation, as a large group of the total participants contributed to the majority of the reports (half of reports were reported by more than 40% of controlled reporters)
a sporadic and decreasing longitudinal representation as participants adopted a flexible reporting behaviour with a slight decreasing frequency at the end of the study	frequent and constant longitudinal representation as participants adapted their work to the reporting activity, adjusting a fixed schedule and a consisting even increasing tendency of reporting
a variability in the voting representation as votes covered a large range of temperatures and were distributed widely across the TSV scale (e.g. the IQR in Table 9 indicates that half of the autonomous voting covers a wider range of temperature (1.6 ∘ C) than the controlled profile)	a uniformity in the voting representation as votes covered a small range of registered temperatures and are concentrated in a small subset of the total TSV scale (e.g. the IQR shows that half of the controlled voting covers a narrower range of temperature (0.9 ∘ C) than the autonomous profile)

Conclusion

The typology of reporting behaviour could be used to assess the completeness and granularity of the dataset, and to consider contextual explanations of a quantified observed phenomenon. From the results of this study, it can be initially concluded that the autonomous reporting behaviour resulted in a dataset with a lower reliability but higher validity than in the controlled reporting behaviour.

Methodological recommendations

- Participation strategies should adjust for controlled and autonomous reporting styles by considering differences in fixed versus flexible reporting protocols, authoritative versus suggestive styles of reminders

- Self-reported analysis should assess the methodological accountability of embedding reporting behaviour analysis and contextual factors in the assessment of comfort

Mixed Methods: CHALLENGE

Skills

Mixed

- Experience in conducting both quantitative and qualitative
- Solid grounding in mixed methods

Time and Resources

Team work

- Is there sufficient time to collect and analyze two different types of data?
- Are there sufficient resources to collect and analyze both data sets?
- Are the skills and personnel available to complete this study (team work)?

Convincing Others

Newness

- Too new to understand its value / no time to learn it
- Object to the idea of mixing paradigms
- Too comfortable with one method
- Easy to make a mess of it

Recommended books Mixed Methods
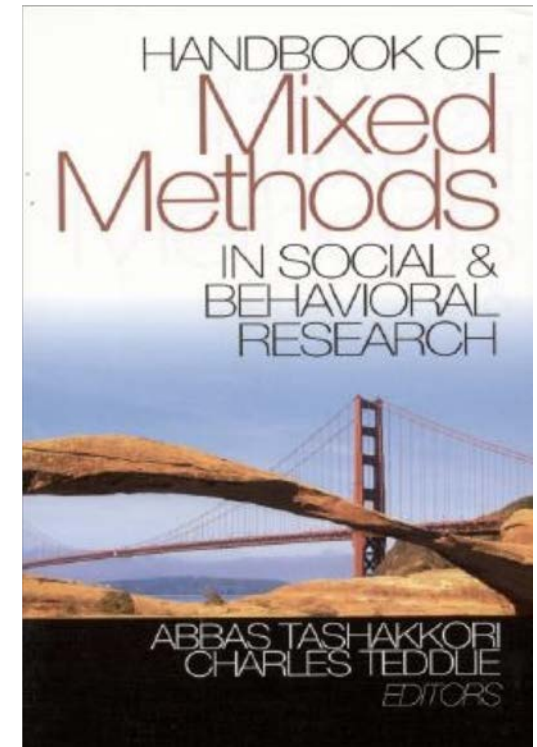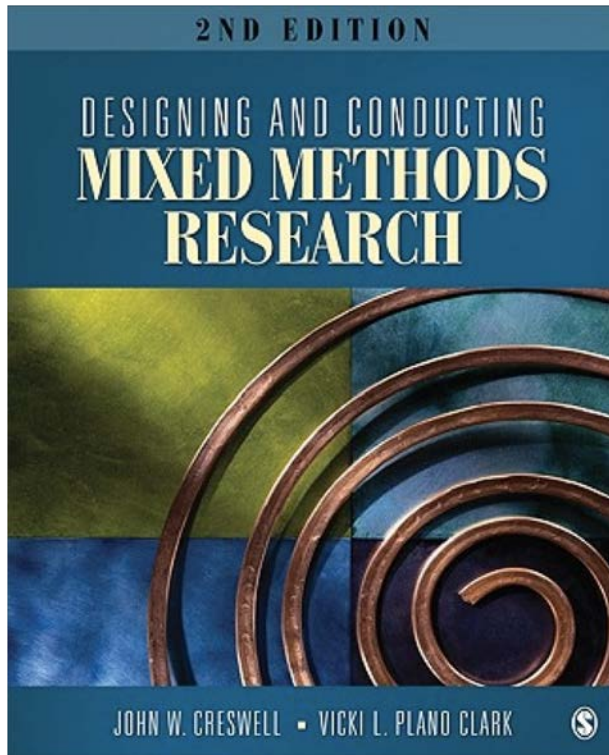
Assignment 1

What methods will you use and why?

↓

What data will you collect?

↓

What is the type of that data?

↓

How will you use it to make sense or prove a point?

Part 2

What are the requirements for rigorous (case study) research?

Quality in research?

- Quantitative?
  - Purpose of explaining
- Qualitative?
  - Generating understanding

quality, credibility, and trustworthiness
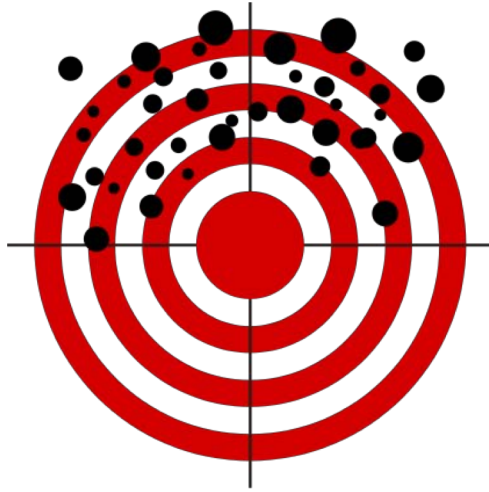
Unreliable & Unvalid
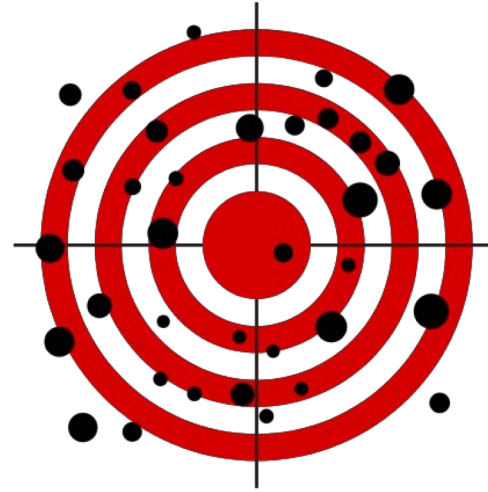
Unreliable, But Valid
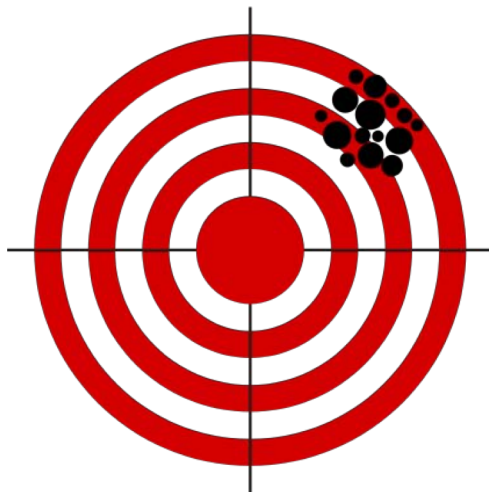
Reliable, Not Valid
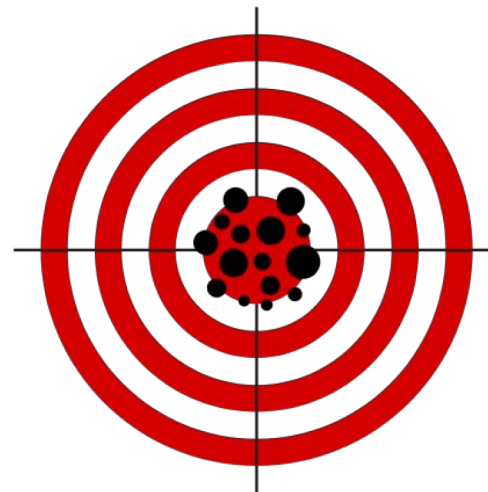
Both Reliable & Valid

Reliability

- Absence of random error
- Subsequent researchers come to the same insights

- **Transparency:** Low inference descriptors → detailed data presentations

- **Replication:** Inter rate reliability checks

Construct validity

The extend to which a study investigates what it claims to investigate

- Triangulation
- Clear chain of evidence

Internal validity

"logical validity" and refers to the presence of causal relationships between variables and results (avoiding confounding variables)

Develop a clear research framework

Pattern matching: avoid "anecdotalism" and include "deviant cases"

External validity

- Generalizability

- Difficult for case studies

- Cross cases will improve external validity

Good Practice Guidelines

Criteria	Possible methods
Transparency i) are your methods clear? ii) could others repeat your work?	Provide a clear account of procedure used, an 'audit trail' that others can follow (i.e. ensure that the evidence – fieldwork notes, interview transcripts etc - can be inspected independently, and that procedures for data analysis are clearly described and justified.)
Maximise validity	Ensure that your conclusions are based on supporting evidence and include analysis of cases that do fit within your conclusions and enough context for reader to judge interpretation.
Maximise reliability	Analyse the whole set of data Use more than one analyst / coder
Comparability	Compare data between and within cases in the data set Compare findings to other studies
Your role as a confounding factor	Account for the role of you, the researcher, in the research. Take into consideration the impact that you being part of a wealthy organisation such as MSF might have had on the responses given. Do you think people might have exaggerated certain problems for example? Don't be swayed by favourite findings! Be careful not to pull out only those findings that **you** find interesting. For example, don't just include in your analysis those who might say that MSF provides the right type of aid if others disagreed.

https://cloudfront.ualberta.ca/-/media/science/research-and-teaching/teaching/qualitative-research-methodology.pdf

Articles discussing quality in qualitative research

- Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of family medicine and primary care, 4*(3), 324.
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report, 8*(4), 597-606.

Verifying conclusions

- Checking for representativeness
- **Checking for researcher effects**
- **Triangulation across sources and methods**
- Checking the meaning of outliers
- Looking for negative evidence
- Ruling out spurious relations
- **Replicating a finding**
- Checking out rival explanations
- **Getting feedback from informants.**

Three strategies to report rigorous work
Gibbert, M. and Ruigrok

Strategy 1: Talk the Walk *(what is reported)*

- Report concrete research actions rather than abstract criteria: reader can appreciate the logic and purpose of the actions

- Case selection
- Transition between data collection and data analysis
- Selection of interviewees

- Reporting validity and reliability strategies

Three strategies to report rigorous work
Gibbert, M. and Ruigrok

Strategy 2: Priority Ordering of Validity Types
(how is reported)

- Internal and construct validity over external validity

Three strategies to report rigorous work
Gibbert, M. and Ruigrok

Strategy 3: Necessity is the Mother of Rigor
(why is reported)

- During case study research unforeseen things happen.
- Research strategies need to be changed along the way

- Creatively use setbacks and make best use of existing resources → 'messy' methodology sections

Article explaining 'requirements' in qualitative research

Gibbert, M. and Ruigrok, W. (2010) The "What" and "How" of case study rigor: Three strategies Based on Published work", *Organizational Research Methods*, 13(4) 710-737.

Assignment 2

How do you ensure that you meet the requirements?

PART 3

How (why) to publish your data?

Data management plan

Any research project will gain from the development and usage of a data management plan to understand how to store and publish all the data it creates. Indeed, nearly all funded projects within programs of the NWO and the EU Horizon 2020 now require a data management plan to be written within the first few months of the project.

Data management plan

A data management plan is a document that describes how the data will be generated or used within a given project, how they will be collected, managed, stored and made available during the study, and how they will be shared upon completion of the research project.

A data management plan is useful for purposes beyond merely satisfying the requirements of the research funder. It also helps the researcher to determine how the data can be managed efficiently and effectively. In addition, it can reduce the risk of data loss or other threats that could render the data illegible or unusable (e.g. the obsolescence of software).

Templates for writing a data management plan can be found on the RDM website of your university:

TU Delft

TU/Eindhoven

University of Twente

Advice

Please seek advice and support from your local front office on writing or updating your data management plan:

TU Delft RDM Support

TU/Eindhoven RDM Support

University of Twente

4TU.Centre for Research Data

lft.on.worldcat.org
s://tudelft.on.worldcat.org/external-searc...

DSA
2014
2017

- Home
- Planning research
- Conducting research
- Publishing research
  - Long-term storage
  - Uploading data
  - Licensing
  - Data description and formats
  - Data citation
  - Data journals
  - Data Refinement Fund
  - Data Paper Fund
- Training & Events
- About 4TU.ResearchData
- Researchers about 4TU.ResearchData
- Data Browser ▼

Publication of research

Store research data for the long term

When completing your study, you may store your data (or selections thereof) for the long term, in addition to publishing and disseminating it. 4TU.Centre for Research Data arranges standardised, well-documented storage, thereby guaranteeing that the data can be found and re-used.

Upload your own research data

Researchers may upload the data (up to approx. 10 GB) themselves. For the transfer of larger datasets or that consist of various parts, you can make use of our FTP server, SURFdrive (only for Dutch higher education institutions) or an external hard drive. Contact 4TU.ResearchData staff for assistance.

Cite data with a DOI

Every dataset in 4TU.ResearchData archive is provided with a unique, persistent Digital Object Identifier (DOI), which can be linked to or cited in publications. This allows you to claim your research data as your own. The DOI also simplifies the citation of the data.

The technical details of data storage and access

Storage, retrievability, re-use and citability are made possible through a technical infrastructure. 4TU.ResearchData closely follows the rapid developments in this field.

Publish in indexed journals

The number of 'data journals' is increasing. These journals offer researchers space in which to describe their data, to provide details on their collection and processing, to describe the software

Can we help?

Twitter

4TU CENTRE RESEARCH DATA

4TU.ResearchData
@4TUResearchData

"Celebrating Data! What's next?" 28 november organiseren @lcrdm, #RDNL, #NPOS, #NFU en de #UKB een interessante dag…
https://t.co/YFrFSH8QDw

Description

Descriptive metadata are indispensable for the preservation, retrieval and re-use of datasets. They provide answers to questions concerning the person creating the data, the subject of the data, the type of file, geographic information and other aspects. In other words, they are 'data about data'.

When uploading, you will be asked to enter the following descriptive metadata:

Creator*	Main researchers involved in producing the data.
Title*	Name or title by which the dataset is known.
Contributor	Institution where the data was created or collected. A person or organization responsible for making contributions to the dataset.
Publisher*	A holder of the data (including archives appropriate) or institution which submitted the work. Any others may be listed as contributors.
Publication year*	The year when the data was or will be made publicly available.
Date created*	Date the resource itself was put together; this could be a date range or a single date.
Description*	Concise description of the contents of the dataset. Describe the research objective, type of research, method of data collection and type of data.
Subject	Subject, keyword, or key phrase describing the resource.
Temporal coverage	Indicate the dates to which the data refer. Enter the year, or beginning and ending dates.
Spatial coverage	Describe the geographic area to which the data refer (e.g. municipality, town/city, region, country). The geographic coordinates of the area may be included, if desired.
Identifier	4TU.ResearchData automatically assigns a DOI to a dataset once the entire deposit procedure has been completed. In some cases, a dataset may be known by one or more other (persistent) identifiers.
Language*	The primary language of the resource. When no language is added, 4TU.ResearchData will automatically assign 'English'.
Link to publication	Include the web addresses or DOIs for any publication, important internal reports or other datasets that are related to your dataset.
Licence*	Terms and conditions on how the dataset may be used. Our recommended licence is CC0 as it makes your data maximally reusable. When no licence is selected, 4TU.ResearchData will automatically apply CC0 to your dataset.

Auxiliary information necessary to interpret the data, such as explanations of codes, abbreviations

Where to get help?

- Jeff Love (Data steward IDE)
  - J.S.Love@tudelft.nl