

Writing with Data Workshop

Tommaso Venturini (tomm.venturini@gmail.com / www.tommasoventurini.it)

The concept

This workshop is meant to provide a collective and hands-on reflection on the techniques of academic writing, with a specific focus on writing for the social sciences. It is explicitly inspired by the “writing workshops” organised by Howard Becker in Chicago and Bruno Latour in Paris, but adapts such model to the writing of essays that are based on the analysis of data.

The workshop considers all types of writing (visual as well as textual) and explores their role not only in the presentation of findings but also in the staging of the research protocol.

Writing with data is the art (or rather the artisanship) of crafting coherence between the conceptualisation, the operationalisation and the presentation of a scientific research.

The audience

5 to 15 Ph.D. students or researchers (or *very good* master students)

Preparation

One week before the first session of the workshop, each of the participants submit to the facilitator:

1. A tentative **title** for the essay she/he would like to write
2. A **1000/2000-word text** on the essay she/he imagines writing divided in two parts:
 - a short **introduction** presenting the object and objective of the essay;
 - a short **conclusion** presenting the (potential) take-away message of the essay.Both drafts of future essays and extract from past paper or projects are acceptable.
3. A step-by-step description of the **research protocol** that she/he used or intends to use
4. The **dataset(s)** that she/he intends to explore (or a sample of it) and a short description of it. Qualitative datasets (e.g. interviews transcription, observation notes) are perfectly fine but need to be handed in as well (as least a sample of them).
5. (If possible) **1 to 3 diagrams**, each visualising one dimension of the information contained dataset or one of the conceptual arguments conveyed in the text.

The items above are collected and distributed to all the participants, who are required to read them carefully before the workshop.

The workshop's golden rules

The objective of the workshop is to discuss collectively the conceptualisations, operationalisations and visualisations proposed by each participant *as conveyed in the her/his introductory text, research protocol or dataset*. Participants are encouraged to criticize each other, politely but nitpickingly, according to two golden rules:

1. *No substantial critiques allowed.* Participants should abide by the proposals contained in the text, protocol and dataset. The *form* of these proposals, however, can and should be thoroughly inspected to point out:
 - ambiguities or lack of explicitness
 - inconsistencies between different parts.
2. *No general critiques allowed.* All observation should make explicit and precise reference to a sentence in the text, a step in the protocol, or an item in the dataset.

The exercises

The workshop is organised around a series of exercises designed to spark the discussion among the participants. Each exercise is performed individually by all participants on the same text/protocol/dataset (indicated by the facilitator) and the results are discussed collectively. A (non-exhaustive and non-ordered) list of the exercise is presented below.

1. Unfolding the multiple paper

Good academic texts revolve around one single idea (possibly declined under different angles). Often, however, multiple stories, research questions and textual objectives are folded into one piece of writing thereby overcharging the text and producing ambiguity for the readers. Sealing a clear deal with the reader (about what the text will do or not do) facilitates understanding and avoiding the risk of being accused of not delivering what promised. While the definition of “textual deal” is entirely up to its author, academic texts, as any piece of literature, are always part of a “genre” which has rules and create expectations that must be respected (or at least acknowledged).

1.a. What’s the deal? Consider the text indicated by the facilitator and identify the “deal” that it proposes to its reader. If more than one answer can be identified for each of the items of the templates below, flag it as a problem.

- This project investigates (or tells the story of) [name and define the *explanandum*],
- in order to show how it is [chose one of the following relations]
 - different from (or similar to)
 - connected to
 - compose by (and hence can be decomposed into)
 - grouped or sorted according to
 - ...
- [name and define the *explanans*],
- through the example of [name and describe the *case study(ies)*]
- and a dataset composed of [name and describe the type data available for the research]

If more than one answer has been identified for the same items, discuss collectively which one should be dropped in order to isolate the deal that promises the best combination of interest and feasibility (in the light of the data available for the research).

1.b. Genres. Consider the paragraph(s) indicated by the facilitator and rewrite them using a different style in order to comply with a different textual genre. Possible styles/genres include:

- objectifying, positivist, natural-sciences-like essays;
- thick description, anthropology-like essays;
- personal recollection essays;
- normative, recommendations essay;
- ...

2. One title to rule them all

The consistency of academic texts is often threatened by a lack of consistency between the aspirations expressed in the conceptualisation, the operationalisations allowed by available data, and the findings actually delivered in the presentation – in most cases, through an increasing cutback from the ones to the others. This shift does not necessarily invalidate an academic text, but it must be controlled and explicitly justified.

Consider the set of items (texts, protocol/datasets, diagrams) indicated by the facilitator and sum up each of them in a single sentence (be benevolent, but honest). Compare the three sentences and discuss if the shift from one to the other is justifiable and actually justified in the text.

A good way of testing the coherence of an academic text is to try to consider its title and decide if it can be equally applied to its three main dimensions (conceptualisation, operationalisation and presentation).

If this is not the case for submission indicated by the facilitator, discuss whether the problem

- is in the title (and can be fixed by proposing an alternative title);
- or in the textual consistency (and can be fixed by tweaking one of the dimensions).

3. Characters, techniques of staging and casting

Academic texts are characterized by a complex form of storytelling. On the one hand, like all narrative texts, they should tell the story of a few identifiable characters acting in a well-delimited situation (conceptual definition). On the other hand, they should establish a robust connection between such characters and the phenomena that they are supposed to impersonate (operational definition).

2.a staging. Highlight all the characters present in the text indicated by the facilitator. Mark them in yellow in their first appearance and in green the following. Consider whether all characters are:

- clearly identified (it is easy to distinguish between new and recurring actors?)
- necessary (are they all doing something? Do they all make a difference? Is the stage too crowded?)

In passing, consider if the researchers themselves are presented as characters of the text.

2.b casting. Work together to fill the table below:

Character name	Conceptual definition(s)	Operational definition(s)	False positives	False negatives

- Discuss whether the casting is complete and univocal: does the research protocol provide clear and unique conceptual and operational definitions for all the characters?
- Compare the span of the conceptual and operational definitions, are they coextensive? Can you think of possible false positives (things that are captured by the operational definition, but not covered by the conceptual one) or false negative (things covered by the conceptual definition but not captured by the operational definition).

4. Writing like Hemingway, designing like Tufte

The writer Ernest Hemingway is known for is legendarily concise style (which, according to some, derived from his experience as a news correspondent at the time of the telegraph).

Similarly, the designer Edward Tufte is known for having coined the notion of data-ink ratio, the idea that diagrams should convey the data they represent (or more precisely the dimensions of the data they focus on) with the maximum possible visual economy.

3.a Writing like Hemingway. Rewrite the section indicated by the facilitator trying to use as few words as possible, but...

- ... do not make the argument any less explicit or detailed
- ... do not make the nargument flatter or less interesting.

Do the exercise above in three steps:

- first, by removing all unnecessary words and replacing longer constructs with simpler ones;
- then, by dropping unnecessary actors and notions;
- finally streamline the argument by moving sentences and rewriting connections

3.b Designing like Tufte. Consider the diagram (or the set of diagrams) indicated by the facilitator and enunciate as clearly as possible the argumentative objective that it pursues (the pattern in the dataset or the argument in the text that the diagram is meant to represent).

Contemplate each of visual variables employed in the visualisation and identify which dimension in the dataset or elements in the argument it represents.

- Are all relevant dimensions/items present in the diagram and translated in a visual variable?
- Does the diagram visualise some dimensions/items that are not relevant for its objective?
- Is the same dimension/item represented by more than one visual variable?
- Are the chosen visual variables the best fit for the dimensions/items that they represent?

NB. Keep in mind that

- *proportional variables* (whose values can be quantitatively compared) are best to represent numerical values;
- *ordered variables* (whose values are linearly sorted) are best to represent rankings;
- *associative variables* (whose values have equal visibility) are best to represent categories.

Use the table below as a guide to discuss:

- How the diagram can be improved
- What diagrams could be more effective to fulfil the argumentative purpose of the author

Visual variable	Data (or argument) dimension	Alternative visual variable
Position (proportional)		
Size (proportional)		
Brightness (ordered)		
Hue (associative)		
Form (associative)		
Others (orientation, saturation, texture...)		

5. Unpacking theoretical shortcuts

From the point of view of writing, theoretical concepts are only useful only inasmuch as they fulfil a role in the mechanism of the argumentation. To do so their meaning as well as their textual function must be sufficiently clear. Also restraint is advisable in the use of theoretical jargon, which should be employed only when strictly necessary.

Highlight all the theoretical notions present in the text indicated by the facilitator and, for each them,

- Check if the text provides
 - (a) the reference of the text in which the notion is developed,
 - (b) a short explanation of the meaning of the notion.
- Identify whether the notion
 - fulfils a legitimate rhetorical function (connecting the text to a larger academic discussion, being supported or defeated by empirical evidence, articulating two arguments, explaining a finding, guiding the interpretation...);
 - or is used as an illegitimate shortcut (to avoid getting into the necessary details of a complex argument) or as a boast of academic sophistication.
- Find a common language expression that could replace the notion.

Pay particular attention to “substantivized” words (e.g. constructive-*ism*, intersection-*ality*, platform-*isation*, disenchant-*ment*)

6. Harnessing the paragraph

The paragraph is the crucial textual unit of academic writing (and of most writing in general). The paragraph is a textual “unit” in two different senses: in the sense that it should contain one and only one argument as well as everything needed to make the argument, but also in the sense that it should be explicitly articulated to the other units of the texts (the other paragraph). A good paragraph is thus self-contained and well-connected.

Considering the paragraph indicated by the facilitator:

Tommaso Venturini – Writing with Data Workshop

- Highlight the sentences (or part of sentences) that makes the argument and the sentences that makes the connection with previous and following paragraphs. Is the distinction clear?
- Represent the paragraph as a simple “input/output machine”: what does the paragraph import from the text that precedes it, and what does it export to what text that follows.