Preface

Writing and publishing are as important as research: your work is incomplete until you publish it. If no one cites your work, it is as if you haven't done anything (Kamat, 2015). Writing well increases how often people cite your work (Calgano, 2012).

Books, articles, and courses teach us how to write better but we are not reading them. Many articles drown in the passive voice, rely on crutch/zero verbs (use, perform, make, do, carry out, conduct) and introduce sentences with phrases like "It is seen that". Nature and Science insist that authors favor the active voice. Most of us don't publish in either journal. To write well, verbs are active and vigorous text is precise and concise and sentences are straightforward. Avoid the seven cardinal sins of writing: hedging, signposting, redundancy, self-consciousness, narcissism, boosting, and periphrasis.

How well do you write papers and present your research? We crafted 20 questions that test your knowledge of the minutiae of communicating science effectively

- forming sentences,
- formatting graphs and tables,
- expressing data appropriately, and
- recognizing the impact of your work through bibliometric measures.

If you answer 14 of the questions correctly, you communicate well. We substantiate our answers from the top scientific journals and other books dedicated to communicating science.

QUESTIONS

1. *Publishing*: Web of Science™ (WoS) Core Collection (2014) and Google Scholar (2014) are two citation databases that index publications of all forms. At the end of 2014, WoS indexed 39 million documents, including scientific articles (23 million), proceedings papers (6 million), meeting abstracts (4 million), book reviews (2 million) and editorials, letters, reviews, and news (4 million). What percentage of the scientific articles (in WoS) have been cited at least once?
   (a) more than 90%,
   (b) between 80% and 90%,

(c) between 50% and 80%,
(d) less than 50%.

2. *Publishing*: More people have cited Axel Becke's article on density functional thermochemistry (Becke, 1993) than any other between 1989 and 2014. How many citations has WoS indexed for this article?
(a) fewer than 5000,
(b) between 5000 and 15 000,
(c) between 15 000 and 45 000,
(d) more than 45 000.

3. *Writing style*: Passive voice, active voice, agents, and patients: writing clearly means that you identify the patient, the verb, and the agent. Select the best sentence?
(a) We measured the pressure periodically.
(b) The pressure was measured periodically.
(c) It was shown that the pressure varied periodically.
(d) Pressure varied periodically.
(e) The pressure measurement was performed at intervals.

4. *Writing style*: Fewer words can be more powerful than many words. The sentence below has 19 words and contains a hedge and an ionized verb. What is the minimum number of words to express the same information?
It is interesting to remark that these results indicate that deactivation phenomena are not important below 500 °*C for PdO/ZrO*$_2$*-Y.*
(a) less than 6,
(b) 6–10,
(c) 10–14,
(d) 14–19.

5. *Reporting data*: What do error bars in graphs or uncertainty in variables ($x \pm \Delta_x$) represent? (Multiple answers.)
(a) standard deviation, σ;
(b) standard error of the mean, $\frac{\sigma}{n}$;
(c) confidence interval, $t(\alpha, n-1)\frac{\sigma}{n}$;
(d) instrument resolution;
(e) maximum and minimum of a range of measurements.

6. *Reporting data*: Significant figures pollute literature data: don't carry more than your experiments warrant. Order the list below from most certain to most uncertain:
(a) 310 °C,
(b) 583.15 K,
(c) 310 °C with a relative uncertainty $\Delta_T = 1\%$,
(d) 583.15 K with a relative uncertainty $\Delta_T = 1\%$.

7. *Reporting data*: The International Bureau of Weights and Measures (Bureau International des Poids et Mesures, BIPM) maintains and updates SI writing conventions. All physical quantities can be expressed with the SI,

but the BIPM does accept nonstandard expressions (e.g., bar for pressure). Identify all acceptable SI expressions to describe silica content:

(a) The mass fraction of silica was 3%.
(b) The mass fraction of silica was 0.03.
(c) The silica content was $3\%_{wt}$.
(d) The silica content was 3% (w/w).
(e) The silica content was 3 wt%.

8. *Graphs*: Understanding typography—font types typefaces, character height, line weight—can help create an aesthetic graph. Identify all correct approximations for a 0.35 mm thick line.

(a) 0.014″,
(b) 1 pt,
(c) 1.3 px,
(d) 0.08 em,
(e) 0.35 pt.

9. *Graphs*: The ideal character height in a graphs ranges between 2.5% and 5% of the *y*-axis. It should be closer to 5% in presentations, posters, and graphical abstracts. What is the best Arial font size for a 50 mm graph (*y*-axis).

(a) 2 pt to 4 pt,
(b) 4 pt to 6 pt,
(c) 5 pt to 8 pt,
(d) 6 pt to 9 pt.

10. *Graphs*: Line weights in presentations are heavier than in journal articles. They also depend on frame dimensions—lines should be thicker for wider graphs. What line weight should frame a 70 mm graph for a paper (and poster)? What should the line weight be for a presentation or graphical abstract? (Multiple choices possible.)

(a) hairline, hairline;
(b) 3 pt, 3 pt;
(c) 0.75 pt, 1.4 pt;
(d) 0.25 mm, 0.5 mm;
(e) 0.25 mm, 0.25 mm.

11. *Tables*: Graphs are more effective than tables at communicating trends. Tables are better to report descriptive data, property data, and data with multiple response variables and factors. Select the best visual constructs to separate columns of data.

(a) vertical lines,
(b) shading,
(c) space,
(d) (a) and (b),
(e) (b) and (c),
(f) (a), (b), and (c).

12. *Papers*: Titles are to articles what poetry is to literature. They should be concise and precise. However, words in titles are also a source for Internet search engines. What should be the maximum number of words for a title?
(a) 12,
(b) 15,
(c) 18,
(d) 21,
(e) 24.

13. *Papers*: We reference other work to give credit, to give a framework to our research, to educate readers, and to direct readers to published procedures and methods (to keep your paper short). Between 2011 and 2014, *Nature* published over 3200 articles and letters that referenced 113 000 articles. What is the average age of the references (the difference between the year *Nature* published the paper and the year the reference article was published)?
(a) 4,
(b) 6,
(c) 8,
(d) 10,
(e) 12.

14. *Papers*: Talk to your readers and audience rather than lecture them. In the classic writing style, writers recognize that readers are competent and will recognize the truth as you lead them through your work. Soggy prose seeks "to argue for the truth" (Pinker, 2014). For each sentence, match one of the seven cardinal sins of writing on the left with the corresponding sentence on the right (multiple choices).

1	Hedging	a	These results are extremely significant statistically and compare favorably with validation studies (Hatzivassiloglou and McKeown, 1997).
2	Signposting	b	In general these results show that a system with zero-cost identities does not require centralized allocation of identities to encourage cooperation (Lai et al., 2003).
3	Redundant	c	Here we report our new results on the samarium-arsenide… (Ren et al., 2008)
4	Self-conscious	d	More recently researchers have attempted to quantify the effects of anxiety on foreign language learning (Horwitz et al., 1986).
5	Narcissism	e	Whether established pests are suitable for attempted eradication is extremely controversial (Myers et al., 1998).
6	Boosting	f	In this section we shall evaluate the rate of recombination for nonequilibrium conditions (Shockley and Read, 1952).

15. *Papers*: What are the minimum requirements for coauthorship?
- **(i)** Conceive and design the study (or parts of it).
- **(ii)** Collect and analyze data.
- **(iii)** Interpret data.
- **(iv)** Draft the article.
- **(v)** Revise parts of it.
- **(vi)** Approve the final version.
- **(vii)** Agree to be accountable for the results.
  - **(a)** (ii), (iv), (vi), (vii);
  - **(b)** (i), (ii), (iv);
  - **(c)** (iii), (v);
  - **(d)** (vi), (vii).

16. *Presentations*: There should be fewer slides in a presentation than the number of minutes allotted for it. For 15 min presentations, 10 to 12 slides is a good target. However, don't put 5 min worth of content into each slide. At most, how many images (or bullets) should a slide have?
- **(a)** 3,
- **(b)** 4,
- **(c)** 5,
- **(d)** 6,
- **(e)** 8.

17. *Presentations*: Some people speak too quickly, while others speak so slowly they put you to sleep (which may be a blessing). If you are speaking quickly, you have too much material and you will lose your audience: if you speak 100 words per minute, a 15 min presentation equals 1500 words, which is about half of the text in a paper. What is the optimal cadence of a presentation in words per minute?
- **(a)** less than 90,
- **(b)** 90–120,
- **(c)** 120–150,
- **(d)** 150–180.

18. *Posters*: How many figures and images should a poster have at minimum?
- **(a)** 2,
- **(b)** 3,
- **(c)** 4,
- **(d)** 5.

19. *Posters*: What is the minimum font size to read the poster title from 2 m (6.5 ft)?
- **(a)** 24 pt,
- **(b)** 36 pt,
- **(c)** 48 pt,
- **(d)** 60 pt,
- **(e)** 72 pt.

20. *Plagiarism*: Identify any cases of self-plagiarism:
- **(a)** You copy a graph from one of your articles to a book without permission from the copyright holder.
- **(b)** You copy several paragraphs from your thesis into an article.
- **(c)** You take data from a graph and reproduce it (not identically) in another figure.
- **(d)** You copy text from your patent to a scientific article.

21. *Bonus*: Ahiram's sarcophagus is inscribed with Phoenician text. All alphabets originate from Phoenician characters. Translate what we inscribed on the side of the lid on the front cover of the book (Figure 0.1).

FIGURE 0.1 We inscribed these Phoenicians letters on the side of the lid of King Ahiram's sarcophagus picture.

All books on scientific communication advocate writing clearly and concisely. However, most journal articles drown in the passive voice, and many rely on ridiculous phrases such as *It is seen that* to introduce an idea. Many researchers consider that anything other than the passive impersonal voice is unprofessional. Kirkman (1971) surveyed members of the Institution of Chemical Engineers to determine what kind of writing they preferred. Although he said the response "should destroy for ever the myth that most engineers prefer a heavy impersonal style," only 46% of the readers preferred the active voice. Is the active voice better than the passive voice for technical communication?

Writing conventions incite visceral reactions (arguments):

- *Data are* or *Data is*?
- Two spaces or one space after a period?
- Sentences can't start with *and*, *but*, or *because*.
- Infinitives should not be split.
- Don't use a contraction.
- Capitalize words following colons.
- (Unrolling toilet paper from the top or the bottom?)

Passive voice versus active voice in scientific literature is not a conflicting writing convention: write active sentences! They communicate more clearly and concisely, and are easier to understand.

We recommend ways of writing and presenting. Often we are assertive about these *rules*, but many journals require you to conform a writing conventions: The Chicago Manual of Style, AMA Manual of Style, The ACS Style Guide, and Scientific Style and Format: The CSE Manual for Authors, Editors, and

Publishers. Check these guides first before using a contraction or allowing data to be singular.

The first chapter is all about publishing and why it is important. We discuss impact factors for journals and *h*-indices for authors, and identify the countries that publish the most and the research fields that are growing and shrinking.

In the second chapter, we outline best writing practices, the active voice, and concise and plain language. We introduce some fundamental concepts of sentence structure. You will delete the phrase *It is seen that* (and its derivatives) entirely from your scientific repertoire of poor English after reading this chapter. Our premise is that expressing information with fewer words is better: sentences should not boost (e.g., *very, significantly, extremely*), hedge (*might be considered, is a distinct possibility, our results imply*), or mutilate/ionize/nominalize verbs to create noun phrases (*calcination was done, for the production of performed, made a measurement*). Writing in the active voice is difficult and requires practice. We reproduce texts from the scientific literature, demonstrate the problems with them, and suggest alternatives. At the end of each chapter we list exercises for practice and provide our solutions in Appendix A.

Language should be precise, and it is debatable how precise it needs to be, but data is rarely precise to more than three significant figures! We summarize statistics, uncertainty, and uncertainty propagation, and insist that you report your data correctly (precisely but no more precisely than your data merits).

We give explicit instructions to produce acceptable graphs and tables in Chapters 4 and 5. You will be an expert in creating graphs with the correct font size, line thickness height, and width after you read this chapter.

The following chapters on papers, presentations, and posters apply our guidelines to these three forums to communicate your results. We adopt the IMRAD format—introduction, methods, results, and discussion—and go through each of these sections for a standard scientific communication. We draw on the top journals for inspiration (and rules). Although we structure our discussion with the IMRAD format, we recommend you be more creative with your presentations and posters. Every second counts, and uninformative section titles waste time and space.

The consequences of plagiarism can be devastating (Chapter 9). If your name is on the paper, you are responsible for its content.

ACKNOWLEDGMENTS

The inspiration to write this book comes from graduate students who struggle writing their first paper, their second paper, and every paper after that. We started by compiling a glossary of overused phrases and expressions from students, and when the list became too large, we summarized it in a document that Elsevier posted on its website (Patience et al., 2013).

This document energized us to go further. We scoured the literature, and found many books that address writing English, or reporting data, or preparing scientific documents. Some of them also describe how to prepare graphs and tables. *Nature Methods* published a series on reporting data and the *Journal of Physical Chemical Letters* published 20 editorials on publishing papers. Who needs more? What researchers need to do more is to read these texts, but we feel that isn't enough: they need to practice, they need feedback, and they need explicit directives. This book complements the corpus of literature on scientific communication.

Many people contributed their artistic skills, and many more scrutinized our language and composition. In particular, Caterina Rocca wrote the Phoenician letters and contributed images, Jason Robert Tavares, Patrice Perreault, Federico Galli, Danilo Klvana, Cristian Trevisanut, Jitka Kirchnerova, Roland Malhamé, Clara Santanto, Joceyln Doucet, Pierre Sauriol, and Frank Ajersch all read a chapter or two and made helpful suggestions, Nicolas Patience compiled much of the data in the Appendices, and Christian Patience and Brendan Patience collaborated.

ANSWERS

Here are our answers to the quiz.

1. *Publishing*: (b). On average, 86% of all articles have been cited at least once. In 2014, 82% of the papers written in the early 1980s had been cited. Citation rates are increasing: in 2014, 90% of the articles written in 1996 had been cited at least once. The WoS *All Databases* has 92 million articles as of December 2015. (You can retrieve them by typing *":"* in the Basic Search window with *Topic* category.)
2. *Publishing*: (d). WoS registered over 47 000 citations as of December 2014. Google Scholar indexed 22% more citations of this paper. Lowry et al. (1951) wrote the most cited paper in 1951 (cited 308 000 times up until 2014).
3. *Writing style*: (d). This is the best sentence, but it can be better.
   - **(a)** We measured the pressure periodically.
     It doesn't matter who measured the pressure: *We* is an unnecessary agent. A better agent would be the instrument (e.g., a pressure transducer with a resolution of x mbar).
   - **(b)** The pressure was measured periodically.
     The is a classic passive sentence and it is vague. Identify an agent and the frequency.
   - **(c)** It was shown that the pressure varied periodically.
     It was shown that is superfluous. Removing this phrase doesn't change anything (except making it shorter).

(d) Pressure varied periodically.
This is the best sentence, but it is still vague. State the frequency.

(e) The pressure measurement was performed at intervals.
The word *measurement* makes your sentence longer unnecessarily. If you know what the pressure was, something measured it.

4. *Writing style* (a).
PdO/ZrO_2-Y is stable below 500 °C.

- *It is interesting to remark* is an unnecessary phrase.
- *not important* is better stated *unimportant*.
- *these results indicate*: state the results.

5. *Reporting data*: (a-e). All of them can represent uncertainty. You must state explicitly which one them you are using. Error bars in *Nature Methods* graphs are predominantly standard deviation (a) or standard error (b) (Krzywinski and Altman, 2013a). The JCGM (2008) states that (b) represents the *experimental standard deviation of the mean*, and that it is incorrect to refer it to as the *standard error of the mean*.

6. *Reporting data*: When the uncertainty is missing, we assume that it equals half the smallest significant figure. To convert degrees Celsius to kelvin we add 273.15. If the instruments don't record temperature to within ±0.01 °C, then don't add 0.15 K when you convert degrees Celsius to kelvin. Adding 273.15 K to 310 °C and keeping five significant figures increases the certainty 1000 times.

			Rank
(b)	583.15 K	±0.005 K	1
(c)	310 °C ± 1 %	±3 K	2
(a)	310 °C	±5 K	3
(d)	583.15 K ± 1 %	±6 K	4

7. *Reporting data*: (b). A space between 3 and % is missing in (a).

8. *Graphs*: (a-d). The line thickness of graphs ranges from 0.3 mm to 0.6 mm.

9. *Graphs*: (c) and (d). One inch equals 72 pt. A two-inch plot (51 mm) equals 144 pt. So, 2.5% and 4% of 144 pt are 3.6 pt and 5.8 pt, respectively. However, Roman numerals (and uppercase letters) are shorter than the total font size: they represent approximately 75%, so the font has to be bigger to meet our recommendation. The minimum font size in a graph is 6 pt, and the maximum is the caption font size.

10. *Graphs*: (c) and (d). Graph line weights in papers and posters range from about 0.75 pt to 1 pt (0.25 mm to 0.35 mm). Axis line weights in papers and posters are 1% of the length of the *y*-axis (in millimeters). We recommend that the line weight in presentations and graphical abstracts be about double what they would be in papers and posters.

11. *Tables*: (c). We prefer white space to vertical lines or shading.

12. *Papers*: (b). The top 1000 most cited articles in *Science* and *Nature* (in 2013) average 10 words. Only 13 of the top 1000 papers have more than 15 words. Among the top 1000 most cited articles in WoS (until the end of 2014) 130 titles have more than 15 words. The longest title has 32 words. The shortest title has one word and six characters. The titles of TED talks average 5.5 words per title with a standard deviation of 2 (1978 titles up until May 2015 - http://www.openculture.com/2014/06/1756-ted-talks-listed-in-a-neat-spreadsheet.html).

13. *Papers*: (c). References in *Nature* follow a Weibull probability distribution rather than a normal distribution. The cumulative distribution function, $P(t)$, has a modulus $m = 1$ and a scaling parameter $\theta = 8.1$:

$$P(t) = 1 - \exp\left(-\frac{t}{\theta}\right)^m.$$

This means 63% of the references in *Nature* are less than 8.1 years old, and 50% are less than 5.1 years old.

14. *Papers*: Some of the sentences have more than one cardinal sin of writing.

1	Hedging	b	In general these results
		d	have attempted to quantify
		e	attempted eradication
2	Signposting	f	In this section we shall evaluate
		c	Here we report
3	Redundant	c	new results
4	Self-conscious	e	is extremely controversial
5	Narcissism	d	More recently researchers
6	Boosting	a	extremely significant
		e	extremely controversial

15. *Papers*: (a). Taking responsibility for parts of the work is the minimum requirement for authorship (Royal Society of Chemistry, 2014), but that is insufficient. Authors must participate in writing and in manipulating data or designing the experiments (Shewan and Coats, 2010).

- conceived and designed the study **or** parts of it, **or** collected, analyzed, **or** interpreted data; and
- drafted the article **or** revised important parts of it; and
- approved the article's final version; and
- agreed to be accountable for the results.

16. *Presentations*: (d). Six 8 cm × 8 cm squares fit into a standard slide. This is the absolute maximum. More than 2 slides with 6 regions with graphs and text is too much information. Consider TED talks where speakers limit most slide to one image, one idea per slide and the maximum is 4. Gallo (2014) recommends that the first 10 slides should have no more than 40 words.

17. *Presentations*: (c). We have recorded people speaking at several conferences. Anything less than 100 words per minute seemed slow. Presentations approaching 160 words per minute were fast. An manager from General Motors gave a 1 h presentation and spoke over 160 words per minute for 1 h. The presentation was excellent, but there was so much information that it was difficult to follow. If you speak quickly, take more pauses. Dananjaya J. Hettiarachchi won the Toastmasters International World Championship of Public Speaking and spoke at 120 words per minute. Some TED lecturers—Al Gore, Hans Rosling, Becky Blanton—speak at rates between 130 words per minute and 160 words per minute, whereas others speak at rates exceeding 180 words per minute (Dlugan, 2012). Gallo (2014, pp. 82) holds speaking faster than 160 wpm is best (the rate of a casual conversation). However, the audience in TED talks are different than at international conferences many of whom English is their second or third language.
18. *Poster*: (e). Regardless how far people are away from your poster, the title should never be smaller than 72 pt. If you would like senior professionals, researchers and professors to read your text, bigger is better. Remember that 72 pt equals 1 in. (25.4 mm). In an oral presentation, the minimum font size is about 20 pt. Most poster text should be at least 32 pt.
19. *Poster*: (d). We surveyed the opinions on posters of 65 doctoral students, postdoctoral fellows, and engineering and chemistry professors: 80% agreed that posters should have at least three figures. Graphs and images must dominate a poster not text. At least five images are necessary for a poster.
20. *Plagiarism*: Self-plagiarism doesn't exist. If you are the copyright holder, you can copy it. When you publish a paper, you assign the copyright to the publisher. It is no longer yours. If you copy it without citing the source or getting permission, you are plagiarizing and infringing the copyright.
21. *Hint*: Phoenician reads from right to left.