Journal of University Teaching & Learning Practice

Volume 18
Issue 2 *Innovations and challenges in the teaching of statistics to non-specialists*

Article 02

2021

What statistics instructors need to know about concept acquisition to make statistics stick

Jochen Kruppa
Universitätsmedizin Berlin, Germany, jochen.kruppa@charite.de

Jessica Rohmann
Universitätsmedizin Berlin, Germany, Jessica.rohmann@charite.de

Carolin Herrmann
Universitätsmedizin Berlin, Germany, carolin.herrmann@charite.de

Miriam Sieg
Universitätsmedizin Berlin, Germany, miriam.sieg@charite.de

Kerstin Rubarth
Universitätsmedizin Berlin, Germany, kerstin.rubarth@charite.de

See next page for additional authors

Recommended Citation

Kruppa, Jochen; Rohmann, Jessica; Herrmann, Carolin; Sieg, Miriam; Rubarth, Kerstin; and Piper, Sophie, What statistics instructors need to know about concept acquisition to make statistics stick, *Journal of University Teaching & Learning Practice*, 18(2), 2021.
Available at:https://ro.uow.edu.au/jutlp/vol18/iss2/02

What statistics instructors need to know about concept acquisition to make statistics stick

Abstract

"The limits of my language are the limits of my mind. All I know is what I have words for" (Wittgenstein). When learning something completely new, we connect the unknown term to an already existing part of our knowledge. We can only build new ideas and insights upon an existing conceptual foundation. In the field of statistics, we educators frequently find ourselves met with great confusion when teaching novices. These students, entirely unfamiliar with even basic statistics, must connect the introduced statistical terms within their personal existing networks of largely non-statistical knowledge. Lecturers, on the other hand, who are well versed in statistics, have deeply internalized the content to be taught and its relevant context. The juxtaposition of the two roles may produce amusement in a lecturer upon gaining insight into the word associations made by the statistical novices. For example, a 'logistic regression' does not involve the 'shipping of goods in economically difficult times,' though this might seem entirely reasonable and intuitive to the statistics learner. Other times, these different perspectives can lead to headaches and frustration for both learners and their lecturers. In this article, we illustrate how simple statistical terms are initially connected to a student's pre-exiting knowledge and how these associations change after completing an introductory course in applied statistics. Furthermore, we emphasize the important difference between "term", "approach", and "context". Understanding this fundamental distinction may help improve the communication between the lecturer and the learner. We offer a collection of practical tools for instructors to help promote students' conceptual understanding in a supportive, mutually-beneficial learning environment.

Keywords

Basic learning and teaching; Curse of knowledge; Expert knowledge; Misconceptions; Statistical language

Authors

Jochen Kruppa, Jessica Rohmann, Carolin Herrmann, Miriam Sieg, Kerstin Rubarth, and Sophie Piper

Introduction

Those of us who teach statistics to non-statisticians are victims of the so-called curse of knowledge (Camerer *et al.*, 1989; Froyd & Layne, 2008; Birch & Bloom, 2007). The primary goal of teaching statistics is to transmit statistical knowledge and facilitate statistical understanding. The Guidelines for Assessment and Instruction in Statistics Education (GAISE) additionally recommend instructors “teach statistical thinking”, “focus on conceptual understanding”, “integrate real data with a context and a purpose”, “foster active learning”, and “use technology to explore concepts” (Carver et al, 2016). All these recommendations include the dissemination of concepts. To be able to understand a new concept, however, the student first needs to learn some vocabulary. In this paper, we focus on the very beginning of statistical teaching, where the first building blocks are put into place. We consider the earliest stages of learning, during which time the learner has little to no prior statistical knowledge. Simply stated, our setting can be boiled down to statistical experts teaching statistical novices. We, the teachers, take on the role of the statistical “experts”. We are not (necessarily) experts in teaching, but rather professionals working in the field of statistics who also teach. Very few of our students have previously heard of statistical concepts. The novices in question may be undergraduate students or experts in other scientific fields. It is well known that many non-statisticians initially struggle to learn introductory statistics due to common misunderstandings of even basic statistical concepts (Gardenier & Resnik, 2002). This can cause confusion, stress, or even feelings of anxiety that are counterproductive to the learning process (Perry, 2006). Misinterpretation in statistical learning has been “decried for decades, yet stays rampant” (Greenland *et al.,* 2016). The teaching approaches suggested to overcome these initial hurdles are manifold. Motulsky (2014) promotes the implementation of “intuitive biostatistics”. Larry Gonick & Wollcott Smith created a cartoon guide (Gonick & Smith, 1993). Other authors appeal to an audience considering themselves to be statistical “dummies” like the famous For Dummies series by John Wiley & Sons (Rauch, 2020). Van Emden used the title “Statistics for terrified biologists” (2008) to appeal to the prevailing feeling of many biology students. In each of these approaches, a learner-centered perspective is pivotal. The novices’ lack of comprehension is not only frustrating for themselves but also for their instructors. This is especially true for those instructors who invest great effort into developing teaching material but continue to face the same initial misunderstandings repeatedly (Gardenier & Resnik, 2002). In addition to the “curse of knowledge”, instructors find themselves also grappling with the “curse of efficiency” - essentially the ability to perform a given task more swiftly while expending fewer resources compared to the learners (Persky & Robinson, 2017; Logan, 2018; Hinds et al, 2001). This second “curse” leads instructors to massively underestimate the time a novice will need to understand the material; addressing this requires adequate awareness on the part of the instructor (Keysar et al., 1995). While we mainly focus on the link between knowledge and words or phrases, we also present several corrective suggestions to help address two phenomena teachers may remain unacquainted with: the curse of knowledge and the curse of efficiency (Heath & Heath, 2007).

What is the essential difference between a novice and an expert?

We operationalize the term *expert* as a person who has substantial knowledge and can think about a topic at a complex, theoretical level (Persky & Robinson, 2017). An expert can communicate abstractly. In contrast, a *novice* regards a newly introduced topic as a concrete matter and seeks specific numeric examples as well as simple explanations (Ghazi & Ullah, 2016). For instance, when learning about the statistical mean, a novice first wants to know how to calculate the mean for a set of measured numeric values. For an expert, on the other hand, the calculation of the mean is trivial, and it might be more important to explain that the mean - as location parameter of a normal distribution - is sensitive to outliers. During this very early stage of learning statistics, however, a discussion based on the expert's perspective is likely to raise more questions than provide useful information for the novice.

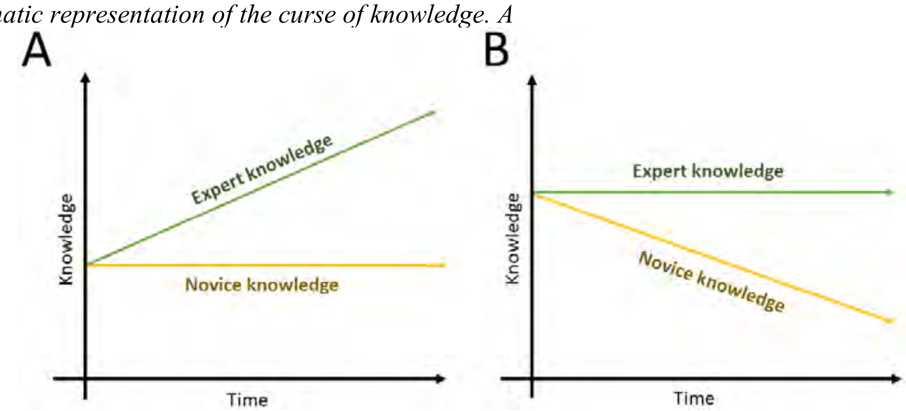
The curse of knowledge

We were all once novices in statistics and initially had to build up our expertise from an low level of statistical knowledge. Even now, we remain on a course to statistical understanding. We see learning more as a journey than as a process of arriving. Meanwhile the more we know, the greater the discrepancy between our level of understanding and that of a novice becomes. Figure 1 illustrates this simply. When progressively learning additional statistical concepts over time, the expert's knowledge continues to increase, leading to a larger gap between the expert's and the novices' understanding (Figure 1.A). However, from the experienced instructor's point of view, each new cohort of students appears less capable than the previous one (Figure 1.B), a common symptom of the so-called *curse of knowledge* (Camerer *et al.*, 1989). Crucially, the instructors are generally not perceptive of the broadening of their own understanding over time. Experts, who no longer write exams or other forms of formal assessment of their statistical knowledge, lack an indicator to measure gained knowledge. Compared to their own rising competence, novices seem less capable each year.

Newton (1990) demonstrated the curse of knowledge in a very simple experiment of *listeners* and *tappers*. Students at Stanford University were grouped into two groups: 40 tappers, who would tap the beat of a well-known song, and 40 listeners, who tried to guess the tapped song. Both tappers as well as listeners knew the pool of possible songs. The tappers estimated the rate of recognition of a song to be 50% [10%; 95%]. Much to the surprise of the tappers, the listeners could only determine 2.5% of the songs correctly, with only three correctly identified tunes out of 120. When 'hearing' the song in one's head while tapping, it is astonishingly easy to connect the tapping to the song. Contrarily, without this important element all the tapped rhythms sounded very similar to the listeners.

In a typical statistical teaching setting, a teacher may ask novices to "calculate the mean". However, not all students may be familiar with the Σ operator or the calculation behind it. Moreover, teachers may unconsciously assume that their intuitive understanding of mathematics is readily accessible to the students. Due to the *curse of knowledge*, experts are quick to misjudge the level of understanding of novices and may consequently hold them to unreasonable expectations (Persky & Robinson, 2017; Birch & Bloom, 2007).

Figure 1
Schematic representation of the curse of knowledge. A



Statistical knowledge of the expert increases with time, while new students participating in their entry level statistics courses enter with the same (low) level of knowledge. B: From the expert's perspective, the situation looks different. Since an expert does not actively perceive their own increase in statistical understanding over time, novices in their courses seem to be less capable in understanding basic concepts each year.

The curse of efficiency

However, the *curse of knowledge* is not the only obstacle experts face; they also drastically underestimate the time a novice will need to perform a given task. There is a difference between knowing something and doing it. Speed increases with more experience, tasks can be performed faster, and thoughts can be processed more efficiently. A faster uptake of skills or information is a sign of successful learning. Here too we instructors lack a real indicator allowing us to measure and perceive our own progress. As we become more efficient learners, we remain unaware of our own development. Thus, from an expert's point of view, novices will appear slower and slower with the uptake and implementation of new knowledge each year. This is called the *curse of efficiency* (Heath & Heath, 2007). Experienced teachers may find themselves puzzled by the time it takes a novice to calculate a mean or even to locate a specific value in a table. This misjudgment of the novice's ability to think and work efficiently is still present even if the expert remembers that their own processing speed was slower when they were first introduced to the concepts. Hinds (1999) demonstrated the effect of expertise on predicting novice performances in two studies, in which experts repeatedly underestimate novice performance times. Persky & Robinson (2017) concluded that experts were 1.3 to as much as 17 (seventeen) times faster in performing the same routine task as learners. This discrepancy can be explained by the experts' automatization of the task; because of its routine nature, experts do not need to actively think and concentrate but unconsciously make use of automatic 'paths' (Logan, 2018).

The combination of the learning trajectories of the expert and the novice constitute the so-called curses of knowledge and efficiency (Heath & Heath, 2007). Given these conditions, it is unsurprising that novice students feel overstrained when processing information from the first lectures. In the role of instructor, experts are hesitant to simplify the complexity of their field (and the demonstration of their knowledge) and, unwilling to leave anything out, instead choose to include an unmanageable amount of teaching material that reflects their understanding. The novice, not (yet) capable of adopting this abstract material, struggles to build a foundation for the basic concepts, let alone the more complex ones. Feeling overwhelmed, they may give up and thus block the uptake of even basic information.

Basic statistical language and cognitive development

To overcome these barriers to education, we must first find a common language in our basic statistics classes. The aim of Wittgenstein's work in 1922 was to identify the relationship between language and reality. For this manuscript, we explicitly distinguish between the elements *term*, *approach*, and *context*. *Term* describes a single word, which is coupled to an *approach,* a method or way of doing something. Therefore, a term is the nominalization of an approach. It is important to note that one term can have multiple synonyms, leading to the same approach. In regression analysis, the dependent and independent variable have a variety of terms like outcome, endpoint, risk factor, covariate and even more. Each term and the corresponding approach are associated with at least one *context.* Often, terms and approaches are used in a variety of contexts. A novice's education begins with the learning of single terms. Then, the novice connects single *terms* with instructions for action, an *approach*. Very often, students stop here because the next step, embedding the approach in a meaningful *context*, is the most time consuming and cognitively challenging step. We could assume that learning statistics is a form of cognitive development in a distinct knowledge field, as according to Jean Piaget's theory of cognitive development (Case, 1993). Piaget splits the cognitive development into four different stages: "sensorimotor stage", "preoperational stage", "concrete operational", and "formal operational". One graduates from one stage to another sequentially. While the time spent in each stage might vary from one individual to another, no stage may be left out. The "formal operational" stage of Piaget's theory is the most important one for learning mathematics. It is in this stage that abstract thoughts and logical thinking emerge. Dasen (1994) revealed that only one-third of adults might ever reach the formal operational stage in their lifetimes. For this reason, Ghazi & Ullah (2016) recommended, among other things, that teaching strategies be revised so that students receive educational material of a more tangible nature.

Table 1
Statistical examples for terms, approaches, and context.

Term	Approach	Context
Mean	$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$	Location parameter of a normal distribution, not robust to outliers.
Odds ratio (OR)	$OR = \frac{a/b}{c/d}$	OR is approximately equal to the relative risk under the rare disease assumption in case-control studies. If the outcome is common, the OR overestimates the relative risk.
Confidence interval (CI) for normally distributed values for significance level α	$\left[\bar{x} - z_{\left(1-\frac{\alpha}{2}\right)} \frac{s}{\sqrt{n}} ; \bar{x} + z_{\left(1-\frac{\alpha}{2}\right)} \frac{s}{\sqrt{n}}\right]$	As one of many possible contexts of the CI: The CI can be used to make null hypothesis statistical test decision with additional information on the point estimator $\bar{x}$ given the standard deviation s, the sample size n, and the $\left(1-\frac{\alpha}{2}\right)$-quantile $z_{\left(1-\frac{\alpha}{2}\right)}$ of the standard normal distribution.

Table 1 shows the phrases *term*, *approach*, and *context* with common statistical examples. The term "mean" is the nominalization of the approach "sum individual numbers" and "divide by sample size". In this example, the *approach* is a mathematical formula. The *context* of the "mean" is not needed for the direct application of the *approach*. Most students can easily calculate the mean, even without having any context in mind. One narrow *context* of the mean is the location parameter of a distribution that is sensitive to outliers.

Cargo cult science

What happens if students understand the term and the approach but cannot grasp the context? In such a case, students recognize the statistical term and can use the statistical approach behind the term without having knowledge of the related context. This circumstance produces *cargo cult science* (Feynman, 1974). The cargo cult itself originated in the late nineteenth and early twentieth century (Worsley, 1968) when the indigenous peoples of Melanesia were exposed to colonizing groups and the new goods and behaviors they brought with them. Many of the imported changes these colonizers brought with them, in lifestyle and utilities, provided direct advantages to and were adopted by the indigenous people. Even after the colonizing groups left, many of these habits survived. Perhaps the most remarkable behavior adopted was the building of airplanes and airfields. After the second World War ended, these indigenous communities continued to build these structures with the goal of receiving airdrops of cargo on the Philippines islands. The indigenous people did everything according to plan; they mastered the approaches ‘airfield’ and ‘towers’. Therefore, they formally performed ‘air traffic’ correctly. Still, none of the desired goods arrived. The critical context was missing.

Feynman (1974) stated that the scientific community is under constant danger to approach cargo cult status. He even coined the term ‘cargo cult science’. In statistics, this translates to using a statistical approach in a formally correct way but neglecting the critical statistical context behind it. A typical example is the distinction between mean and median, and the correct usages thereof. While novices are usually quickly able to calculate both, they require some sense of data distribution to decide which term and corresponding approach is best used when. When neglecting the context, they might calculate the mean for a binary variable like “smoking – yes/no” coded with zeros and ones. While the approach may be carried out correctly and provide an answer, this answer will be meaningless because the context is inappropriate. Moreover, results produced through the application of statistical approaches without context will most likely not be reproducible. Consider the usage of the t-test on data violating the normal distribution assumption. While a t-test approach can be applied, the result it produces may not be valid (e.g. if variance is zero). In a further example, the Wilcoxon rank sum test will produce non-significant results if the sample size is less than 4 in any group. The critical values are not defined for such sample sizes because the Wilcoxon rank sum test is a permutation test requiring at least 4 samples to permute. If students are only able to understand the practical “how” without also the importance of “when”, namely the statistical context in which these tools are appropriate, students will conduct “cargo cult statistics” (Stark & Saltelli, 2018).

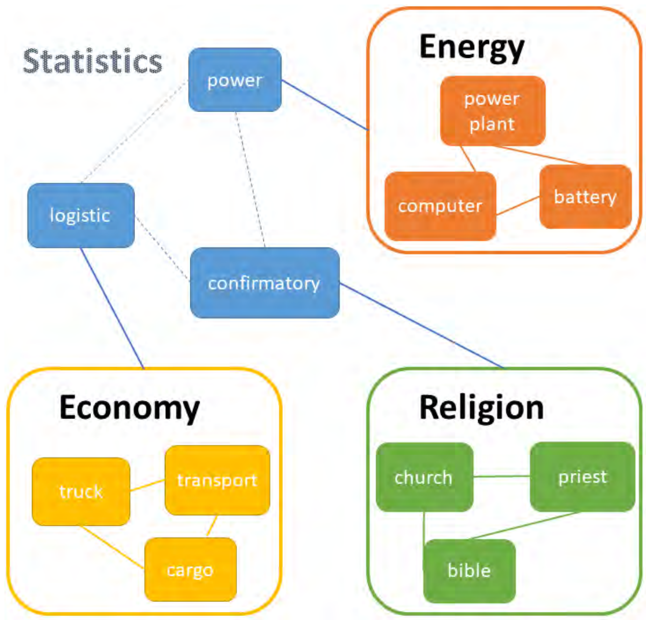
We’ll conclude this section with the story of Max Planck and his driver (Dobelli, 2013). After years of observing Max Planck giving numerous talks about his research, the driver suggested the pair switch roles. He claimed that he had heard the talk so often that he would be able to give it himself in the same quality. Planck agreed, and his driver gave the talk while he sat in the audience. Nobody noticed the switch. In the end, a professor from the auditorium asked a question. The driver said, “I am surprised by this simple question, I suppose even my chauffeur can answer it.” This amusing story contains a practical demonstration of cargo cult science by Max Planck’s driver. Giving the talk without context was feasible, but the unexpected question, delivering a new aspect rooted in theory and involving abstract connections, could not be answered. The driver was ultimately still dependent on his teacher for these elements.

A simplified learning model for statistical novices

Figure 2 shows a schematic diagram of our proposed learning model for statistical novices. We assume that novice students have no prior statistical background (Jaffe & Spirer, 1987). Conceptual statistical network do not yet exists in their minds. The students typically map newly encountered statistical terms to other existing, familiar contexts; perhaps connecting “power” with energy or “logistic” with the economy. This initial matching can cause confusion for both the student and the instructor. The students’ first connections may point the novice down a contextual path that is ultimately erroneous. At the same time, these paths of term associations confuse the instructor, who associates the same set of terms to their (correct) statistical context. To build up a statistics framework and adequate context comprehension, students must first accumulate enough statistical knowledge in the form of terms and then reallocate terms from other contexts to their developing statistical network. Any interested readers should consider Clarà's (2017) paper for a deeper discussion of the word-object relation in the context of learning theory.

Figure 2

Simplified learning model for statistical novices.



Lacking the requisite background knowledge, a novice has yet to develop a mental framework for statistics. Thus, new statistical terms will be associated with already stored network from existing knowledge in other domains. In the process of learning, the novice will reallocate and add new terms to a statistical context in their growing statistics network. It is this initial matching that leads to confusion for both teacher and student because students first associate terms to other contexts and their paths of associations cannot easily be followed by someone experienced who automatically associates the terms with the correct statistical context.

Pilot study: Poll among statistical novices

To illustrate our concept, we conducted a small pilot study among participants enrolled in the module "Statistics for biosciences I" at the Freie Universität Berlin, a series of fourteen in-class lectures with twelve supportive tutorials (in German language). For the vast majority of students in the course, regardless of their semester of study, this module of applied statistics for undergraduates was the first-ever encounter with statistics. The sessions covered descriptive statistics, simple statistical test procedures, and ended with a focus on linear regression and the corresponding assumptions. At the very beginning of the module, before any teaching had begun, students were given a short anonymous questionnaire in German language. Students could decline to participate without consequences. Participants were asked to respond to the following prompt: "Name the first terms or words (maximum three) that come to your mind when you read the following statistical terms." In total, the students were asked to report free word associations for twenty-two statistical terms. The questionnaire was administered in German with the aim of achieving free word association responses not hampered by language barriers. All 46 students took part in the first poll (Table 2).

Table 2
Baseline characteristics of the participants.

	October 2019 **N = 46**	**Retest in February 2020** **N = 31**
Gender	30 female	20 female
	12 male	8 male
	4 other	3 other
Semester at university	1^{st} = 4	1^{st} = 0
	2^{nd} = 0	2^{nd} = 0
	3^{rd} = 24	3^{rd} = 14
	4^{th} = 1	4^{th} = 0
	5^{th} = 8	5^{th} = 5
	6^{th} = 1	6^{th} = 0
	7^{th} = 4	7^{th} = 3
	8^{th} = 0	8^{th} = 0
	9^{th} = 1	9^{th} = 1

After completion of the lecture series in February 2020, we asked the same cohort of students to freely associate six preselected statistical terms presented in the first survey. In doing so, we tried to determine whether the students had changed or widened their statistical context. Due to stormy weather, only 31 students participated in the final lecture and thus were present to participate in the second round of our poll (Table 2, right column). For their use within this article, we have translated the provided terms and corresponding responses into English. Next, we assigned the students' associations to larger concepts like "religion", "economy", or "statistics". Both the translation and the grouping of words are certainly prone to our own contextual preconceptions, and are likely to be subjective to some degree. Nonetheless we feel our summary, presented as simple descriptive graphics, still illustrates our main points.

The frequency with which the most commonly given associations to the terms "regression" (German: "Regression"), "logistic regression" (German: "logistische Regression"), and "confirmatory" (German: "konfirmatorisch") occurred are presented in Figure 3. Column A shows the results of the first poll, the one before statistical teaching began (October 2019). Column B shows the results of the second poll in the same group of students at the end of the lecture series (February 2020). Figure 4 shows the corresponding results of both polls for the terms "distribution" (German: "Verteilung"), "estimate" (German: "schätzen"), and "power" (German: "Power"). Interestingly, the term "regression" was more often associated with "decrease" (Figure 3.A, n = 8) than with "increase" (Figure 3.A, n = 5) among our students. It seems that the natural association of "regression" (German: "Regression") might have something to do with the linguistic similarity to "recession" (German: "Rezession"). When teaching linear regression, teachers and textbooks often introduce the concept by drawing an increasing line to illustrate the terms slope and intercept. In light of the initial associations made with "regression" in our poll, novices may be more receptive to a decreasing line.

Figure 3

Frequencies of answers provided by statistical novices in a free word association exercise with the statistical terms "regression", "logistic regression", and "confirmatory" both before (A) and after (B) their first course in statistics. If students had no association for a term, 'NA' (missing) was recorded. Author JK translated the original answers into English and grouped them by context. A table of original German terms and English translations can be found in the Appendix.

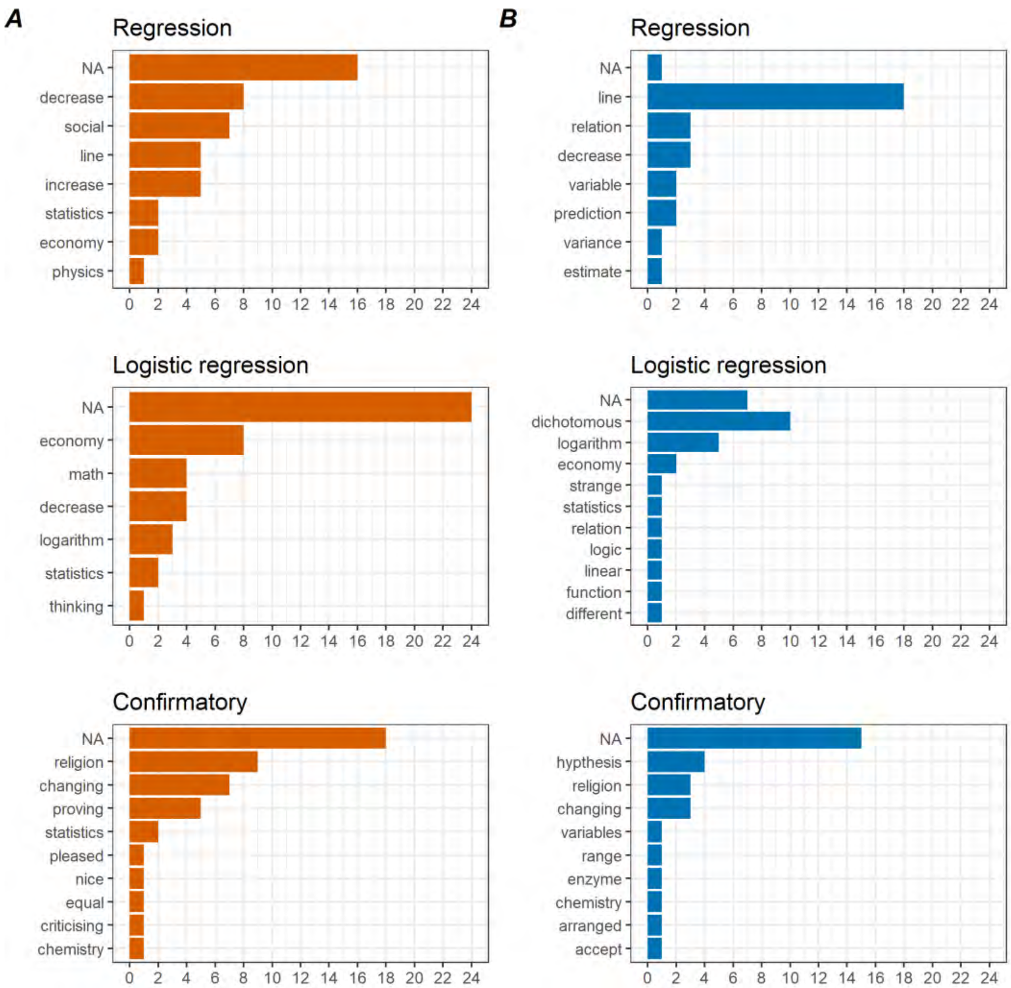
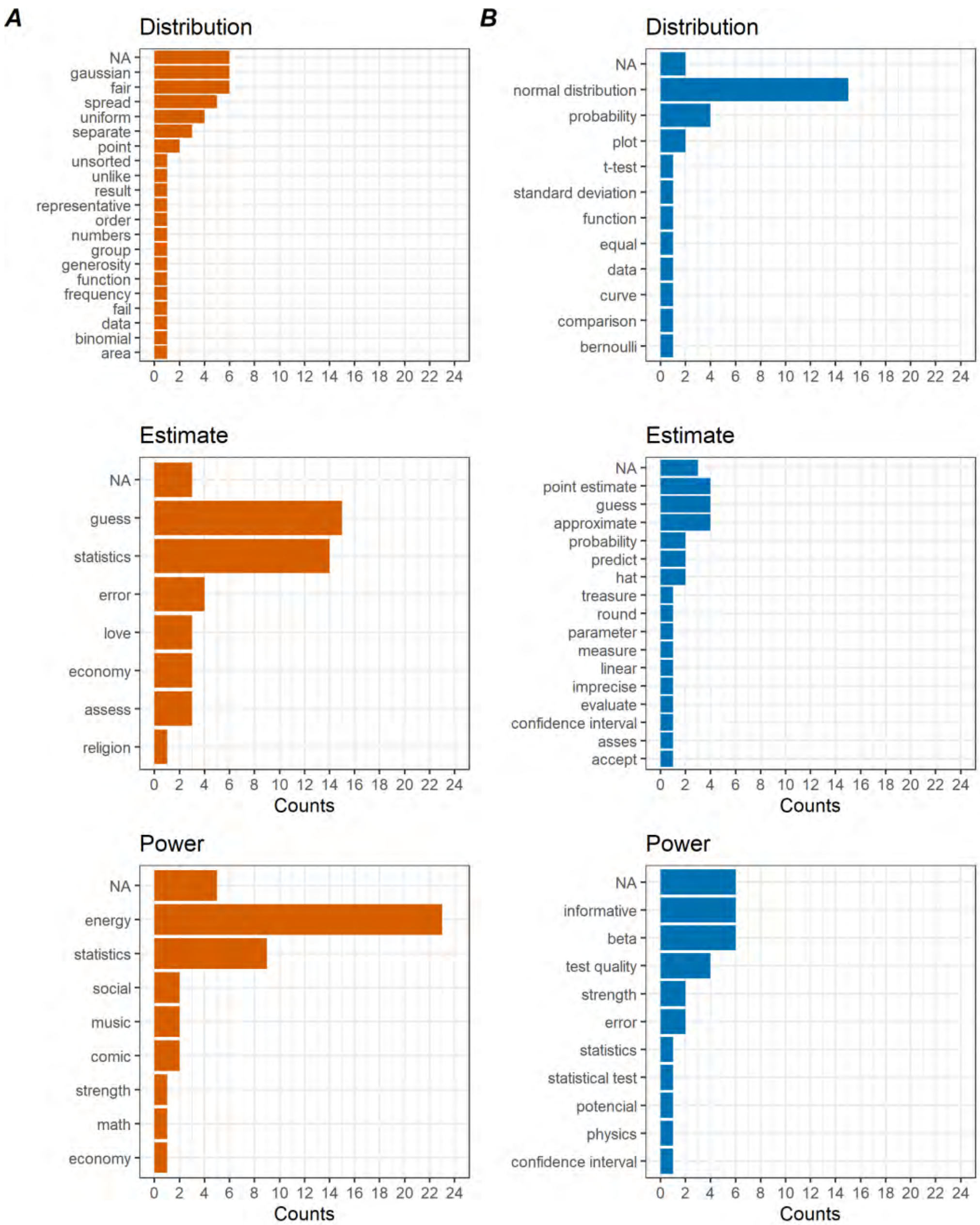


Figure 4

Frequencies of answers provided by statistical novices in a free word association exercise with the statistical terms "distribution", "estimate", and "power" both before (A) and after (B) their first course in statistics. If students had no association for a term, 'NA' (missing) was recorded. Author JK translated the original answers into English and grouped them by context. A table of original German terms and English translations can be found in the Appendix.



One can infer that teachers may benefit from being more mindful of their wording and attentive to the expectations of novices. A similar tendency can be observed in the free word associations with "logistic regression". No student freely associated it with an increase (Figure 3.A, n = 0) but some students associated it with a decrease (Figure 3.A, n = 4). Many students seemed unfamiliar with the term logistic regression and most respondents associated the term with economics. After the completion of the course, the term "regression" was most often associated with the term "line" (Figure 3.B). In the case of "logistic regression", the associations were more diverse. The initial support for the "economy" connection was diminished, with the term "dichotomous" appearing more often in the second round. The term "confirmatory" could be seen as a kind of control term. Though certainly related to statistics, as well as other fields, we did not formally introduce or emphasis this term throughout our course. A fluctuation in answers was observed but there was no clear shift compared to the other terms. The term "distribution" showed more connection to the concept "normal distribution" after our course (Figure 4.B). The translation of "estimate" has a dual interpretation in German, because the German word for estimate (German: "schätzen") can also mean "to cherish". That is probably why novices connected the term "estimate" to "love". After the course, the associations with "estimate" were very heterogeneous. Finally, the term "power" had a strong connection to "energy" before teaching started (Figure 4.A) and was diversely associated with statistical contexts at the end of the course (Figure 4.B). Though true causal inference cannot be drawn from this non-randomized, uncontrolled design, we observed an interesting shift in the association of terms with words from non-statistical contexts towards those from a statistical one.

Implications for practice

Statistics can be taught in various ways with different teaching tools. In the following section, we describe several teaching methods that we used throughout the course to help students build up their statistical knowledge networks. Of course, not all methods can be applied in every lecture, but many of them are generally suitable and can be applied broadly. We hope these suggestions help statistics teachers overcome the curse of knowledge and the curse of efficiency. We aim to help statistical novices generate a broader understanding while reducing the confusion and the frustration commonly accompanying the two curses.

Vocabulary unambiguity

Use only one term for each concept

It is common practice for teachers to vary their vocabulary for a single term in an effort to avoid repetition and to keep the material interesting, or to reflect the actual overabundance of different names for the same term. The usage of the term "outcome" instead of "response", for example, may increase the linguistic diversity but will be highly irritating for people trying to grasp basic statistics terms for the first time. The argument that students should become familiar with the many existing synonymous terms, requiring them to remember several names for the same approach, seems to do more harm than good. The efforts that requires is better invested in practicing the approach and understanding the relevant context. As previously noted, the novice has no pre-existing statistical knowledge network within which to "place" new concepts. Hence, Ghazi and Ullah (2016) recommend emphasizing concrete examples without deliberating abstract concepts for too long when teaching mathematics. Without such instruction, new terms are assigned to multiple contexts and it can take weeks or months of statistical teaching to "rewire" these confused connections to achieve a clear statistical understanding. For this reason, we encourage teachers to use one term exclusively. The repetition of this word will strengthen the initiated knowledge network, facilitating connection with the approach and, eventually, the relevant context. The term should be repeated far more frequently than the expert instructor thinks necessary. As Clarà (2017) details, assigning words a meaning is a process. The word-object relation is formed over time.

Use unambiguous mathematical symbols

Multiple mathematical symbols, such as p for the probability, α for significance level and the β for regression coefficient, are typically introduced in introductory statistics courses. Unfortunately, even in the context of basic statistics, many symbols have several multiple underlying terms, approaches and contexts. A p might be used to denote probability in general, a specific probability, the p-value of a statistical significance test, or alternatively, a proportion. A β may refer to a regression coefficient or the type II error in statistical testing decision theory. Confronted with the same symbol twice in a short course, students may think that we use regression analysis to determine the power, since the Greek letter β is used in both applications. In some introductory lectures, the p-value is introduced using the example of the binomial test. The binomial test, however, is used for proportions, which are often also represented by p. When the same symbols are used to denote multiple terms novices are easily confused, and may think a single approach applies to two distinct terms since they are denoted by the same symbol. It would be beneficial to the practice of teaching if, for example, all p-values would be consequently written in small letters with italic font, while proportions could be abbreviated with a capital P or 'Pr' for a distinct discrimination of terms and their underlying approaches and contexts. Another idea is to teach students to use $\bar{x}$ for the mean and s for the standard deviation when discussing data from a sample, while reserving the Greek symbols μ and σ for the theoretical context only, as the later represent the underlying true mean and standard deviation of the entire population.

It is definitely worthwhile to check lecture material carefully for ambiguous abbreviations and to avoid using the same symbol for different content wherever possible. One could ask a student in a higher semester for assistance in reviewing lecture materials before class, since they may more readily notice any inconsistencies in symbol use. Any additionally introduced statistical context should use new symbols. For larger courses, these symbols should be agreed on by the entire teaching team (including any student tutors) and be used consistently by all involved parties, as well as on any formal assessments. In addition, similar to the process of learning a new language, instructors should allocate sufficient time for students to practice and adopt these new symbols as part of their growing statistical vocabulary.

Chirume (2012) expresses general criticism for textbooks that change symbols too often, explaining that students may be prone to fail because teachers introduce new symbols and terms instead of using familiar ones. Groth et al. (2020) discusses the problem of shared "meanings" for mathematical words, especially when learning probability vocabulary. Teachers should avoid making assumptions about their students' familiarity with mathematical symbols. Begg and Pierce (2020) recommend asking questions about the mathematical symbols to reveal misinterpretations or misconceptions. In their words: "It is too easy to assume that our students are thinking what we are thinking as we use and read symbolic expressions."

Conceptual understanding

Show it, don't just tell it

Another tool is simple visualization. The novice demands concrete explanations, practical examples, and simple visualization to aid learning and understanding, as we have applied in this article. We used different simple and tangible illustrations for our learning model and to introduce the curse of knowledge and the curse of efficiency. Collins *et al.* (1991) described the practice of teaching complex processes by modelling, scaffolding, and fading as being akin to having a "cognitive apprentice". First, the big picture of a problem is presented. The teacher models the overall context and describes the problem with a concrete example. For instance, when teaching generalized linear regression models, the teacher may show a 3x3 cross table indicating the different possible levels of data in statistics. Both the dependent, the independent variable and the X can be continuous, discrete or binary. Next, schematic plots may be drawn and then numeric examples calculated. Here, visualization serves to help students understand the big picture (Supplementary Figure 1).

Tell a story whenever possible

A statistical concept without any context is difficult to remember. Explaining a concept using a story can help not only to guide the students through the lecture, but also to make the concept more memorable. Lawton and Taylor (2020) observed that students found listening to an instructor's personal stories particularly engaging. Zazkis and Liljedahl (2009) present a wide range of examples for teaching mathematics using storytelling. The educational "trick" in using storytelling is to display knowledge as a human-made product—a composite of ingenuity, energy, passion, hope and other emotions. Hansson *et al.* (2019) provide specific examples for teaching scientific principles in middle schools. In our own teaching, we use examples of the Ig Nobel Prize, which specifically honors research that "makes people first laugh and then think" (Abrahams, 2002). We have observed that students are intrigued by the publication of these interesting topics and begin to reflect on them independently. When teaching descriptive statistics and test theory, for example, we have used the comparison of jump performances of the dog flea and the cat flea (Cadiergues *et al.*, 2000). Another example is "the knuckling fingers" (Unger, 1998), which can be used to demonstrate the principles of a scientific abstract and the generation of a research question. Finally, we have used the work of Parkin *et al.* (2009) concerning the wearing of socks over shoes as an example of a clinical trial that explains the CONSORT statement (Boutron *et al.*, 2017). This is only a small collection of the possible publications that are suitable to construct an engaging story. Each scientific area has plenty of relevant publications listed on the website of the Ig Nobel Prize (https://www.improbable.com/ig-about/winners/; accessed on 12th February 2020). We agree that finding an entertaining story suitable for teaching is not always an easy task; however, once found, it can be instrumental in helping students anchor key concepts in their memories and recalling them for a lifetime.

Scaffolding expert thinking

Verbalize your train of thought

Tell your students what you think as you are thinking it. Many statistical solutions seem trivial to us because we already know the solution. However, the students do not know the solution prior to its presentation, and are unaware of the train of thought that accompanies it. Students cannot look into our heads and see how we think (thankfully!). Therefore, we must communicate our relevant thoughts, explaining as precisely as possible what we are doing (and why) as we do it (Collins, 1991). By slowly building up mathematical relationships, step by step, and explicitly stating our rationale for each step, we can help the students to follow our thought process. If a teacher reaches a difficult part of problem that has caused issues in the past, it might be useful to take extra time to articulate the possible source of the confusion. As an example, correctly conducting a statistical test relies on the knowledge and implementation of numerous statistical concepts. We recommend using a step-by-step narrative walk-through, in which the instructor stops and reviews all statistical concepts that arise throughout the process of conducting a statistical test.

Provide statistical mottos

Students frequently ask for a short summary of statistical rules, presented as concisely and tangibly as possible. Providing memorable mottos, proverbs, or rules of thumb might be a solution. We recommend using proverbs where possible because these are easy to remember. Van Belle (2011) introduces statistical rules of thumbs like "Use text for a few numbers, tables for many numbers, graphs for complex relationships", "Plan for missing data", or "Sample size calculations are determined by the analysis". A statistics lecturer could begin class with a relevant rule, to be explained during the lecture, such as "The mean should be reported with the standard deviation", "In a logistic regression, the harmful outcome is indicated by 1 and the positive outcome by 0" or "In prediction, you need training and testing data". Such statements are simple and concrete, and can guide students through practical problems.

Discussion and concluding notes

In this article, we sought to draw attention to two relevant educational phenomena unknown to many statistics instructors: the "curses" of knowledge and efficiency. We have described the path by which statistical understanding is acquired during the early stages of learning in introductory courses and have presented data illustrating statistical novices' free word associations with previously unfamiliar statistical terms. Moreover, we have compiled several suggestions for action with which the pitfalls of the two "curses" can be circumvented.

The generation of a conceptual network for statistical understanding takes time and must be constantly updated and optimized in a process common to all learning (Persky & Robinson, 2017). Whenever possible, using simple and concrete language while minimizing contextual overlap will help students to achieve a working proficiency as a first step towards mastering more complex concepts (Hinds et al., 2001; Ghazi and Ullah, 2016). At the introductory level, misunderstanding statistical terms and concepts is not uncommon among novices because the meaningful context and application have not yet been introduced to them (Jaffe & Spirer, 1987).

Black (2002) found instructors with *less* expertise experienced fewer boundaries in communicating because all participants (learners and lecturers) were more willing to learn from each other. This may explain why a junior teacher may form a closer connection to the students and demonstrate more understanding students' needs and expectations, from which they can tailor a suitable learning program. Less experienced lecturers are likely to approach the material very concretely, which might help explain why beginners perform better when they are taught by these individuals (Hinds et al., 2001). Persky and Robinson (2017) propose different solutions to overcome challenges arising from the so-called "expert abstraction". They conclude, however, that the *curse of efficiency* makes it almost impossible for experts to estimate how long beginners need for a routine task.

Our exploratory survey can be regarded as a proof-of-principle study. The techniques discussed above were incorporated into our teaching throughout the course and may have contributed positively to the shift in the students' identification of statistical terms with statistical concepts, though we caution against a causal interpretation of the results from this small, uncontrolled design. We were not anticipating many of the free word associations resulting from the first poll and were also surprised by the heterogeneity of the reported terms in the same exercise at the end of the course. Additionally, we grouped the reported words into higher order terms in German language and then translated each grouping term into English. This procedure explains why some otherwise inexplicable connections (e.g. of "love" and "estimate") occurred. Understanding, for example, the underlying German ambiguity of the word "schätzen", a verb meaning both "to estimate" and "to love/appreciate", makes this connection clearer. Many German statistical terms have dual-meanings in other contexts, and this is also true for English statistical terms. Actively drawing attention to this consideration early on in a statistics course may promote better communication and reduce frustration for both the instructor and the student. To present the results of our small, exploratory data collection we used simple descriptive graphics. In a larger study with substantially more students, automated language recognition might be a useful tool to group single word associations in larger word families.

The suggested approaches of presenting vocabulary consistently and unambiguously, promoting contextual understanding, and scaffolding are not only helpful when teaching statistical novices. They may also be beneficial when teaching experienced students who go on to major in statistics and related disciplines (Hinds et al., 2001). Minimizing the use of synonymous terms for the same concept is still helpful to intermediate and advanced learners. For these students, exposure to other possible terms one could encounter in the literature can be useful to familiarize students with the breadth of terminology encountered in the "real world". Nonetheless, using one term per concept throughout a class is indispensable to improve mutual understanding. In practice, using the same terms and the same symbols throughout the lecture was indeed the most challenging part for us as teachers. First, as the instructor, it requires much more preparation to be consistent and not to choose symbols "on the fly" that are readily accessible to you. Second, instructors may feel repeated use of single terms leads to a feeling of monotony and must actively work to suppress the urge to make it more diverse, ultimately sacrificing linguistic variety for clarity. For example, we totally eliminated use of the words "dependent" and "independent" variables when introducing regression and used "outcome" and "risk factor" instead. Though high level nuanced discussions exist pertaining to the meaning of "risk factor" in different contexts (e.g. causal versus prediction frameworks), these terms are more intuitive to students than "dependent" and "independent" variables. As another example, we solely used "β" to denote regression coefficients and the term "type II error" in statistical decision theory.

Our lecture series was held using "fully elaborated language" to allow the students to follow the instructors' train of thought as closely as possible. To this end, the teacher tried to explain each thought fully while teaching. There was *no* writing on the board without a corresponding verbal comment to transparently convey meaning in an effort to minimize potential for misinterpretation. In addition, the students were asked to interpret the formulas written on the board in words for the class. Using this feedback technique, it became immediately obvious if any mathematical symbols were misunderstood. This strategy aligns with proposals by Begg and Pierce (2020). In addition, introducing proverbs such as "In a logistic regression, the harmful outcome is indicated by 1 and the good outcome by 0" or "Use tibbles" instead of data frames in R" were repeated *ad nauseam* (at least, the instructors felt that way). We observed students recalling these verbatim in the oral exams, however, which spoke for the proverbs' memorability.

Lawton and Taylor (2020) investigated methods for fostering students' engagement in courses. Overall, in-class time spent practicing had the most engaging effect, as well as listening to the instructors' personal stories and viewing statistical cartoons. Based on our own experiences, we can highly recommend storytelling. We used different story arcs for different statistical topics in our lecture series. For example, we presented a Star Wars dataset (`starwars` implemented in the R Package `dplyr`) to introduce the nonsense of imputing missing values for variables like gender for R2-D2 and other robots. Later, we used the TV show "The X Files" to introduce the mysterious attributes of the "X" in regression analysis, including overdispersion, independence, collinearity, and missing values. Mixed models were introduced using a fictional dragon dataset; we crafted a scenario in which each dragon had completed an intelligence test in a world where groups of dragons live on different mountains. When tasking the students with performing an analysis of the intelligence scores of dragons, this story facilitated an introduction of the concepts of fixed and random effects.

When participating in scientific research as a part of a multi-disciplinary team, the statistician plays a collaborative role in the larger group of experts working together on a given project (Zapf *et al.*, 2019). When specialists from different fields come together, they often speak different languages and use different vocabulary (Black, 2002). Unfortunately, communication gaps in statistics can exist even in expert-level collaborations (Sharpe, 2013). The statistical expert will encounter collaborators who bring their own (other) expertise, those who may not have much statistical experience, and even some who may be statistical novices (Scholes & Vaughan, 2002). Given the growing importance and emphasis on interdisciplinary collaboration in research, we feel several of the suggestions we have presented could also be useful outside of the classroom at roundtable discussions and in collaborator meetings.

"It is too easy to assume that our students are thinking what we are thinking" (Begg & Pierce, 2020).

To conclude, experts should cultivate an awareness of the *curse of knowledge* as well as the *curse of efficiency,* and remind themselves regularly of the challenging task faced by novices to develop a mental statistical network building on heterogeneous prior knowledge. These "curses" add additional complexity to the more commonly considered elements instructors work to address, such as unique learning styles (Coffield, et al. 2004). If teachers of introductory statistics remain aware of the two "curses", they may foster a more productive learning environment, leading to a more fruitful and less frustrating experience for both their students and themselves.

Acknowledgement

We thank Hannah Grillmaier (Institute of Public Health, Charité – Universitätsmedizin Berlin) and Dr. Uwe Schöneberg (Institute of Biometry and Clinical Epidemiology, Charité – Universitätsmedizin Berlin) for the thorough proofreading of the text.

References

Abrahams, M. (2002). *Ig Nobel Prizes: The Annals of Improbable Research*. Orion.

Begg, M., & Pierce, R. (2020). Symbols: the challenge of subscripts. *International Journal of Mathematical Education in Science and Technology*, 1-8. https://doi.org/10.1080/0020739X.2020.1794071

Birch, S.A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*(5), 382-386. doi: https://doi.org/10.1111/j.1467-9280.2007.01909.x

Black, L.J. (2002). *Collaborating across boundaries: Theoretical, empirical, and simulated explorations* (Doctoral dissertation). Massachusetts Institute of Technology.

Bloom, L. (1998). Language acquisition in its developmental context. In D. Kuhn & R. S. Siegler (Ed.), W. Damon (Series Ed.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th ed., pp. 309–370). Wiley.

Boutron, I., Altman, D.G., Moher, D., Schulz, K.F., & Ravaud, P. (2017). CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Annals of Internal Medicine*, *167*(1), 40-47. https://doi.org/10.7326/M17-0046

Cadiergues, M.C., Joubert, C., & Franc, M. (2000). A comparison of jump performances of the dog flea, Ctenocephalides canis (Curtis, 1826) and the cat flea, Ctenocephalides felis felis (Bouché, 1835). *Veterinary parasitology*, 92(3), 239-241.
https://doi.org/10.1016/S0304-4017(00)00274-0

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, 97(5), 1232-1254. https://doi.org/10.1086/261651

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., ... & Wood, B. (2016). *Guidelines for assessment and instruction in statistics education (GAISE) college report 2016*. https://commons.erau.edu/publication/1083

Case, R. (1993). Theories of learning and theories of development. *Educational Psychologist*, 28(3), 219-233. https://doi.org/10.1207/s15326985ep2803_3

Chirume, S. (2012). How does the use of mathematical symbols influence understanding of mathematical concepts by secondary school students?. *International Journal of Social Sciences & Education*, 3(1).

Clarà, M. (2017). How instruction influences conceptual development: Vygotsky's theory revisited. *Educational Psychologist*, *52*(1), 50-62. https://doi.org/10.1080/00461520.2016.1221765

Coffield, F., Moseley, D., Hall, E., Ecclestone, K., Coffield, F., Moseley, D., ... & Ecclestone, K. (2004). Learning styles and pedagogy in post-16 learning: A systematic and critical review. http://evidence.thinkportal.org/handle/123456789/62

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, 15(3), 6-11.

Dasen, P. (1994). Culture and cognitive development from a Piagetian perspective. *Psychology and Culture*, 145-149.

Dobelli, R. (2013). *The art of thinking clearly: better thinking, better decisions*. Hachette UK.

Feynman, R.P. (1974). Cargo cult science. *Engineering and Science*, *37*(7), 10-13.

Gardenier, J., & Resnik, D. (2002). The misuse of statistics: concepts, tools, and a research agenda. *Accountability in Research: Policies and Quality Assurance*, 9(2), 65-74. https://doi.org/10.1080/08989620212968

Ghazi, S.R., Ullah, K., & Jan, F.A. (2016). Concrete operational stage of Piaget's cognitive development theory: An implication in learning mathematics. *Gomal University Journal of Research*, 32(1), 9-20.

Gonick, L., & Smith, W. (1993). *The cartoon guide to statistics*. Collins.

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., & Altman, D.G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, 31(4), 337-350. https://doi.org/10.1007/s10654-016-0149-3

Groth, R.E., Bergner, J.A., & Austin, J.W. (2020). Dimensions of learning probability vocabulary. *Journal for Research in Mathematics Education*, 51(1), 75-104. https://doi.org/10.5951/jresematheduc.2019.0008

Hansson, L., Arvidsson, Å., Heering, P., & Pendrill, A.M. (2019). Rutherford visits middle school: a case study on how teachers direct attention to the nature of science through a storytelling approach. *Physics Education*, 54(4), 045002. https://doi.org/10.1088/1361-6552/ab07e7

Heath, C., & Heath, D. (2007). *Made to stick: Why some ideas survive and others die*. Random House.

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied*, 5(2), 205. https://doi.org/10.1037/1076-898X.5.2.205

Hinds, P.J., Patterson, M., & Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of Applied Psychology*, 86(6), 1232. https://doi.org/10.1037/0021-9010.86.6.1232

Jaffe, A.J., & Spirer, H.F. (1987). *Misused statistics: straight talk for twisted numbers*. Marcel Dekker.

Keysar, B., Ginzel, L.E., & Bazerman, M.H. (1995). States of affairs and states of mind: The effect of knowledge of beliefs. *Organizational Behavior and Human Decision Processes*, 64, 283-293.

Lawton, S., & Taylor, L. (2020). Student Perceptions of Engagement in an Introductory Statistics Course. *Journal of Statistics Education*, 28(1), 45-55.

Logan, G. D. (2018). Automatic control: How experts act without thinking. *Psychological Review*, *125*(4), 453. http://dx.doi.org/10.1037/rev0000100

Motulsky, H. (2014). *Intuitive biostatistics: a nonmathematical guide to statistical thinking*. Oxford University Press.

Newton, L. (1990). Overconfidence in the communication of intent: Heard and unheard melodies (unpublished doctoral dissertation). Stanford University.

Parkin, L., Williams, S.M., & Priest, P. (2009). Preventing winter falls: a randomised controlled trial of a novel intervention. *The New Zealand Medical Journal*, 122(1298), 31-8.

Persky, A. M., & Robinson, J. D. (2017). Moving from novice to expertise and its implications for instruction. *American Journal of Pharmaceutical Education*, 81(9). https://doi.org/10.5688/ajpe6065

Perry, B. D. (2006). Fear and learning: Trauma-related factors in the adult education process. *New Directions for Adult and Continuing Education*, 110, 21-27. https://doi.org/10.1002/ace.215

Rauch, G., Neumann, K., Grittner, U., Herrmann, C. & Kruppa, J., (2020). *Medizinische Statistik für Dummies.* Wiley-VCH.

Scholes, J., & Vaughan, B. (2002). Cross-boundary working: implications for the multiprofessional team. *Journal of Clinical Nursing*, 11(3), 399-408. https://doi.org/10.1046/j.1365-2702.2002.00634.x

Sharpe, D. (2013). Why the resistance to statistical innovations? Bridging the communication gap. *Psychological Methods*, 18(4), 572. https://doi.org/10.1037/a0034177

Stark, P.B., & Saltelli, A. (2018). Cargo-cult statistics and scientific crisis. *Significance*, 15(4), 40-43. https://doi.org/10.1111/j.1740-9713.2018.01174.x

Unger, D.L. (1998). Does knuckle cracking lead to arthritis of the fingers?. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 41(5), 949-950. https://doi.org/10.1002/1529-0131(199805)41:5<949::AID-ART36>3.0.CO;2-3

Van Belle, G. (2011). *Statistical rules of thumb* (vol. 699). John Wiley & Sons.

Van Emden, H.F. (2019). *Statistics for terrified biologists*. John Wiley & Sons.

Wittgenstein, L. (1922). Tractatus logico-philosophicus (TLP). Kegan Paul.

Worsley, P. (1968). *The trumpet shall sound: A study of "cargo" cults in Melanesia (vol. 156).* Schocken Books.

Zapf, A., Huebner, M., Rauch, G., & Kieser, M. (2019). What makes a biostatistician?. *Statistics in Medicine*, *38*(4), 695-701. https://doi.org/10.1002/sim.7998

Zazkis, R., & Liljedahl, P. (2019). *Teaching mathematics as storytelling*. Brill Sense.

Appendix 1

Original questionnaire

Semester:
Geschlecht: ◻ weiblich ◻ männlich

Nennen Sie die ersten Begriffe oder Worte (maximal drei), die Ihnen in den Sinn kommt, wenn Sie folgende statistischen Ausdrücke lesen. **Nicht lange nachdenken, sondern frei assozieren**.

Beispiel: Huhn – Bauernhof

Translation:

Name a maximum of three terms or words that come to your mind when you read the following statistical expressions. Don't think long, but freely associate.
Example: chicken – farm

Orignal asked terms:

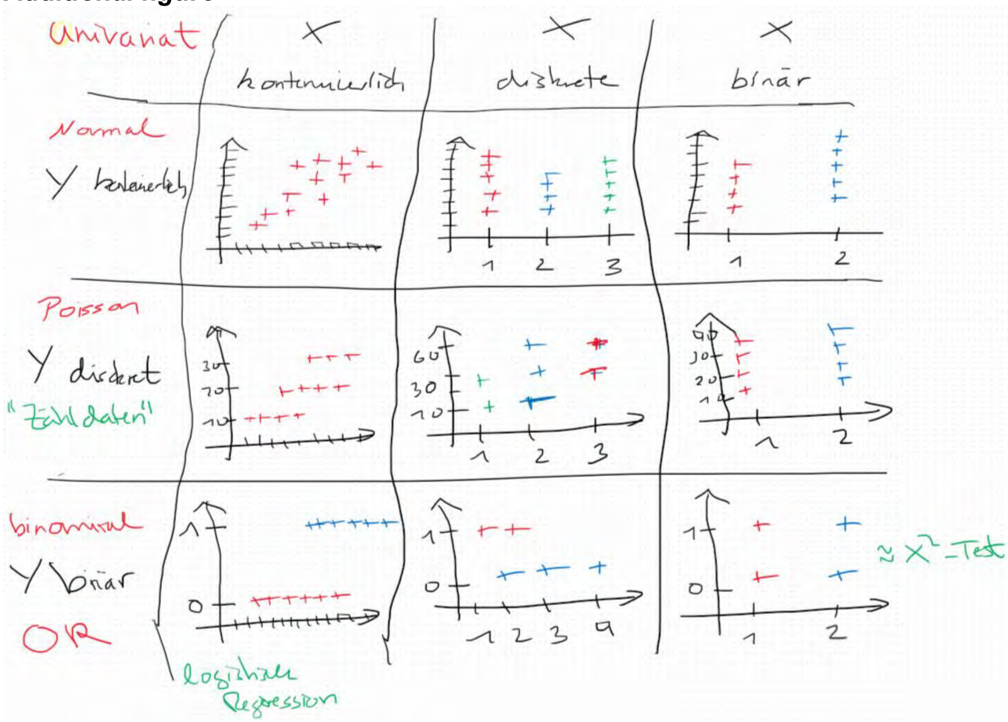
Schätzen, Regression, Logistische Regression, Konformatorisch, Interferenz, Evidenz, Endpunkt, Parameter, Deskriptiv, Explorativ, Korrelation, Hypothese, Power, Verteilung, Dichte, Funktion, Normalverteilung, Fehler, Stichprobe, Modell, Frequentist, Unabhängig

Translation:

Estimation, regression, logistic regression, confirmatory, interference, evidence, endpoint, parameter, descriptive, exploratory, correlation, hypothesis, power, distribution, density, function, normal distribution, error, sample, model, frequentist, independent

The German word "Schätzen" is also true in the sense of "I esteem you". The word "Power" is common on many electrical devices.

Additional figure



Supplementary Figure 1. A 3x3 cross table displaying the different levels of data in statistics versus the corresponding analyses by a generalized linear model. The example is the scaffold from the introductory lecture "Statistics for life sciences II" at the Freie Universität Berlin.