



Handling Anomalous Data in the Lab: Students' Perspectives on Deleting and Discarding

Mikkel Willum Johansen¹ · Frederik Voetmann Christiansen¹

Received: 15 October 2017 / Accepted: 28 February 2020 / Published online: 12 March 2020
© Springer Nature B.V. 2020

Abstract

This paper presents and discusses empirical results from a survey about the research practice of Danish chemistry students, with a main focus on the question of anomalous data. It seeks to investigate how such data is handled by students, with special attention to so-called 'questionable research practices' (QRPs) where anomalous data are simply deleted or discarded. This question of QRPs is of particular importance as the educational practices students experience may influence how they act in their future professional careers, for instance in research. The ethical evaluation of QRPs however is not univocal. In parts of the literature QRPs are seen as unquestionably bad, while in other parts of the literature certain QRPs are seen as a necessary aspect of scientific practice. Results from the survey of Danish chemistry students shows that many students engage in certain types of questionable practices, and that a large minority of the students have been actively encouraged by their teachers to engage in such practices. The paper discusses to what extent and under what circumstances such instructional practices can be defended and suggests how the instructional practice connected to the handling of anomalous data can be improved.

Keywords Questionable research practices · Academic dishonesty · Professional ethics · Laboratory practice

✉ Mikkel Willum Johansen
mwj@ind.ku.dk

Frederik Voetmann Christiansen
fchristiansen@ind.ku.dk

¹ Department of Science Education, University of Copenhagen, Øster Voldgade 3, 1350 Copenhagen K, Denmark

Introduction: Entering the Grey Zone

The problem of so-called ‘questionable research practices’ (QRPs) plays a prominent role in current debates on scientific misconduct. Questionable practices are usually understood as practices that lie between carelessness and outright fraud. Although there is no clear and universally accepted definition describing precisely what the term ‘questionable practices’ covers, in the current literature, the term is taken to refer to practices such as bias in the design or publication of a study, misrepresentation of one’s own contribution to the research, inappropriate assignment of authorship, overlooking other’s use of flawed research data, filtration of research data, and other deviations from accepted norms of good scientific conduct (Fanelli 2009; John et al. 2012; Martinson et al. 2005; Steneck 2006; Swazey et al. 1993). Several large-scale surveys have made it clear that some of these questionable practices are widespread among active researchers. Thus, reporting from a survey answered by 3247 scientists who had received funding from the US National Institutes of Health, Brian Martinson and colleagues (2005) found that although few of the respondents admitted to having conducted outright misconduct such as falsifying or cooking research data (0.3%), 15.3% of the scientists who answered the survey admitted to “dropping observations or data points from analyses based on a gut feeling that they were inaccurate”, 15.5% admitted to “changing the design, methodology or results of a study in response to pressure from a funding source”, and 6% admitted to “failing to present data that contradict one’s own previous research” (Martinson et al., 2005, p. 737). Reporting from a survey of 2155 psychologists Leslie John and colleagues (2012) similarly found that although few of the respondents admitted to outright misconduct such as falsifying data (0.6%), questionable practices such as “selectively reporting studies that ‘worked’” (45.8%) and “deciding whether to exclude data after looking at the impact of doing so on the results” (38.2%), were prevalent (John et al. 2012, p. 525). Only numbers from the self-admission group are presented here. In a meta-study compiling data from 21 studies (including the one by Martinson and colleagues) it was concluded that 33% of the scientists covered by the surveys admitted to having engaged in questionable research practices other than falsification or fabrication of data (Fanelli 2009; see also De Vries et al. 2006).

These results corroborate the results of several historical case studies performed in the last half of the 20th century. Here, it was shown that renowned scientists such as Louis Pasteur (Farley and Geison 1974), George Ohm (Kipnis 2011) and Robert Millikan (Holton 1978) obtained their results through questionable practices such as filtration of data, one-sided analysis and suppression of contrary evidence. The use of questionable research practices, in other words, is not a new phenomenon, nor can it be said—judging from recent surveys—to be a thing of the past. Rather, questionable practices are an integral part of science which will somehow have to be dealt with.

If we turn to the ethical evaluation of QRPs, the literature is divided, perhaps because the very term “questionable practices” can be understood in different ways. In one interpretation, it suggests a suspense of ethical judgement where it is necessary to inspect the practices further before it can be determined whether

they are unethical or not. In another interpretation “questionable” is understood as dubious or, as Merriam-Webster puts it “attended by well-grounded suspicions of being immoral, crude, false, or unsound” (Merriam-Webster 2017). In the literature, both of these attitudes to QRPs can be identified. Thus, in parts of the literature QRPs are seen as unequivocally detrimental to science and consequently as unethical. It is recognised that QRPs are only minor violations of the norms, and yet, in this part of the literature, concerns have been raised that QRPs may in fact pose a more significant threat to the integrity of scientific knowledge than misconduct because QRPs are much more common than outright misconduct. To use the expression from Michael Zigmond and Beth Fischer (2002), the “little murders” of QRPs may do great damage to the scientific community, similar in scope to the “high crimes” of fabrication or falsification. Along the same lines Leslie John and colleagues (2012) argue that QRPs may lead to an arms race, where it is necessary to venture deeper and deeper into the grey zone of QRP in order to survive as a scientist:

QRPs are the steroids of scientific competition, artificially enhancing performance and producing a kind of arms race in which researchers who strictly play by the rules are at a competitive disadvantage. (John et al. 2012, p. 524)

However, prominent authors within history and philosophy of science defend at least some types of questionable practices as a normal and even productive part of scientific work. For instance, in a review of the literature on Mendel the American geneticist Theodosius Dobzhansky explains Mendel’s apparently filtered data by stating: “Few experimenters are lucky enough to have no mistakes or accidents happen in any of their experiments, and it is only common sense to have such failures discarded (Dobzhansky 1967, p. 156)”. Likewise, Millikan’s practice of publishing only a small part of his data has been defended (Holton 1978). In a more theoretical contribution Jerome Ravetz discusses the difficulty of determining when an experiment is ‘working properly’ (Ravetz 1971). Ravetz mentions the ‘fourth law of thermodynamics’ that “no experiment goes properly the first time” and explains that “anomalous readings always do crop up; and if one waited for them to vanish entirely, or tried to ‘explain’ each and every one of them, one would never get beyond this first stage of the work. In short, the scientist must be a craftsman with respect to his apparatus; and his judgement of when it is working ‘well enough’ must be based on his experience of that particular piece of equipment, in all its particularity” (Ravetz 1971, pp. 76–77). And here, the equipment working “well enough” is to be understood as well enough for the data to be kept and not immediately discarded as unsound. Although this supportive approach to QRPs was mainly aired in the late 20th century it is still voiced in parts of the modern literature (Kipnis 2011).

The theoretical and historical studies cited above have established that QRPs are an integrated part of science but, as the tension regarding the ethical judgement of the practices in the face of anomalous data illustrates, it is a part of science and scientific practice that is not very well understood.

In this paper, the focus is on the QRPs related to handling of unexpected, deviant or puzzling results, that is: data that somehow seems wrong. Furthermore,

the paper approaches the problem of QRPs from an educational point of view by investigating and discussing how certain QRPs are handled in an educational setting. This constitutes a real and important gap in the literature. Although the academic honesty and integrity of students has been the subject of several large-scale surveys (Hughes and McCabe 2006; McCabe 1997, 2005; McCabe et al. 2001; see also Davis et al. 2009, p. 41) these studies focus mainly on clear-cut misconduct such as using crib notes or copying from other students (and in short they show that these forms of misconduct are alarmingly frequent). Some of the surveys also touch upon laboratory behaviour. Donald McCabe (2005) and Hughes and McCabe (2006) thus both include the item “fabricating or falsifying lab data” and respectively 19% and 25% of the undergraduates who answered the question admit to having been engaged in those behaviours (Hughes and McCabe 2006; McCabe 2005). In the surveys, however, the practices are seen as unequivocally unethical and the fine line between clearly unethical cases of falsification and possibly defensible cases of questionable practice is not addressed. In contrast, this paper investigates this ‘grey-zone’ and how students navigate it.

A similar picture emerges in the literature on how to address the problem of academic dishonesty (Belter and du Pré 2009; Henslee et al. 2017; Miñano et al. 2017; VanDeGrift et al. 2017); the literature is mainly concerned with plagiarism and other examples of outright misconduct, whereas the problem of questionable practices and the ethical dilemmas created by anomalous laboratory data is not directly addressed.

In the substantial literature on students’ learning in laboratory practice (see Hofstein and Lunetta 2004) the learning opportunity presented by anomalous results has been treated indirectly, for instance by highlighting the need for students to develop scientific skills including not only skills of observation, but also skills of deduction and interpretation (Reid and Shah 2007). However, the difficult question of when to keep data, and when to discard data has not been investigated specifically in the educational setting.

This question is all the more important, because it is reasonable to assume that what students learn in the educational setting, may influence their actions in subsequent professional practice. Thus, Trevor Harding and colleagues (2004) have found that there are similarities in the types of justifications given by engineering students when asked to reason about various types of cheating and policy violations in educational and workplace contexts. These findings give credence to the hypothesis that students in their future professional practice may apply the same types of reasoning and justifications that they employed in their studies, including the more dubious and questionable practices. This is also supported in studies of the role of training effectiveness on ethical decision making (Mumford et al. 2008; Wu 2003). These studies point to the important role played by instructional cases in establishing mental models for ethical decision making. While the hypothesis of the role of instruction is ultimately empirical in nature, the hypothesis also has a normative aspect. A central aim of education in science and engineering is to prepare students to employ their disciplinary expertise in knowledge production, including performing value based judgments such as acting responsibly with respect to experimental data. This

is reflected, for instance, in the accreditation criteria for science and engineering programs.¹ Thus, it is expected that science and engineering education provides students with the ability to act responsibly and to make informed judgments, and it is recognized that such judgments rely heavily on the students' acquired disciplinary training and expertise. From this perspective, QRPs constitute a much more challenging problem than acts of misconduct such as test cheating. Where test cheating and plagiarism can be countered with clear rules ("do not do it"), QRPs call for reflected judgement where students learn how to assess individual cases. Thus, the educational techniques used to address outright misconduct might not be as effective in the case of QRPs.

Method

Participants

All 227 third-year chemistry students from the six Danish Universities that offer chemistry programs were invited to participate in the survey. The prerequisites for admission at the six programs are similar, i.e., a Danish High School (Gymnasium) exam or similar with highest obtainable educational level offered by the Danish high-school system in mathematics, and at least intermediate levels in two science subjects (physics, chemistry, biotechnology, geoscience). Some programs require a certain grade level average from the students' Gymnasium exam, but none of the programs are "elite" programs allowing only students with very high grades. All the universities offer graduate programs in addition to the bachelor programs considered. Curriculum in the bachelor programs is in many ways similar with initial courses within general chemistry, organic and inorganic chemistry (including lab work), in addition to courses in mathematics and physics. Computational chemistry, modelling, statistics, and philosophy of science are present in the programs to varying degrees, and some level of group based project work is an integral part of all programs. For all programs the teaching is conducted mainly by professors and Ph.D. students, assisted by teaching assistants and lab technicians.

Of the total population of 227 third year students who were invited to participate in the survey, 97 responded (42%). Of the 97 students who responded, one was screened out because that individual had abandoned the chemistry program. From the remaining 96, one did not answer any questions, and four were screened out because they had not completed at least 90 ECTS of their education (corresponding to one and one-half years of full-time study), which was an inclusion criterium used to ensure that all participants had enough relevant laboratory experience. Of the remaining 91 respondents 88 completed all the questions concerning lab conduct. The remaining three stopped the survey after question Q3, Q4 or Q5. The responses

¹ For instance, in the ABET criteria concerning Student Outcomes two of the eleven learning outcomes are: "(b) an ability to design and conduct experiments, as well as to analyze and interpret data" and "(f) an understanding of professional and ethical responsibility" (ABET 2017, p. 3).

came from students at all six institutions with the highest proportion being 25% of the total from one institution and the lowest being 5% of another ($n = 91$).

Instrument

The survey was conducted anonymously through an internet based platform. The survey was, in part, adaptive so subsets of the population were given specific questions, depending on their answers to previous questions. The survey consisted of three screening questions, a set of questions concerning lab conduct and a set of questions concerning test behaviour. Only the questions concerning lab conduct (Q2–Q12) are reported in this paper. In the question concerning lab conduct, students were asked whether they had engaged in certain questionable practices related to experimental work, about their conception of and motivation for engaging in such practices, and they had been encouraged to engage in ‘grey zone’ activities by their teachers. The exact questions can be seen in Tables 1 and 2 in our translation.

Prior to the distribution of the final survey, two test surveys had been developed and validated with sample chemistry student populations. Volunteering participants were contacted telephonically and asked to elaborate on their responses and their perceptions of the problems and their understanding of the items in the survey. Based on the feedback from the first round of interviews a substantially different version of the survey was developed, and based on the second round of interviews a final version with minor corrections was made. The survey was in Danish. All quotes from the survey below have been translated directly from Danish to English by the authors.

Procedures

Students were invited through their official university e-mail addresses provided by the institutions. In order to avoid double entries and to secure anonymity, students had to register with a unique, personal 8-digit key. A special concern was to ensure and make clear to the students that data they provided was treated with complete confidentiality. To meet this end students were informed that the study was reported to and approved by the Danish Data Protection Agency (J.nr. 2015-41-3903), and that all data would be recorded and stored in accordance with the guidelines of this agency.

Analysis

The final version of the questionnaire contained both closed questions, which were analysed quantitatively (reported in both absolute numbers and as percentages) and open-field questions where students were being asked to elaborate on specific questions. These open-field questions were analysed and coded in an iterative process following a grounded approach, whereby an initial interpretation and open coding of individual responses were followed by axial and selective categorical coding in relatable categories (Strauss 1987).

Table 1 Results on discarding experiments

Question	Answer	N
Q2 Have you experienced that you had to discard and repeat a full or partly completed experiment because the experiment had somehow gone wrong?	No	14
	Yes	77
Q3 ^a Have you experienced that you had to discard and repeat a full or partly completed experiment only because the results seemed wrong, that is, without you being able to identify the specific source of error? Consider only experiments you had to present to your teacher for instance in a report, presentation or similar.	No	50
	Yes, one time	22
	Yes, several times	5
Q4 ^{a,b} Did you mention the discarded experiment when you described your work to your teacher in a report, presentation or similar?	No	5
	Yes	17
Q4 ^{b,c} Did you mention the discarded experiments when you described your work to your teacher in reports, presentations or similar?	Rarely or never	1
	Sometimes	2
	Often	0
	Always or mostly	2
Q5 Have you experienced that a teacher encouraged you to discard a full or partly completed experiment only because the results seemed wrong?	No	72
	Yes	17
Q6 Describe the last occurrence, where a teacher encouraged you to discard a full or partly completed experiment only because the results seemed wrong	[Open-field question]	
Q7 What is your opinion of discarding an experiment only because the results seem wrong?	It is a completely illegitimate scientific practice	13
	It is an illegitimate scientific practice	32
	Neutral	25
	It is a legitimate scientific practice	11
	It is a completely legitimate scientific practice	7

^aOnly asked of students who answered 'Yes' to Q2

^bOnly asked of students who answered 'Yes, one time' to Q3

^cOnly asked of students who answered 'Yes, several times' to Q3

Table 2 Results on deleting outliers

Question	Answer	N
<i>Q8</i> Have you removed a deviating measurement (an outlier) from a data set only based on a feeling that there was a fault in the measurement? Consider only experiments you had to present to your teacher for instance in a report, presentation or similar.	No	41
	Yes, one time	26
	Yes, several times	21
<i>Q9a^a</i> Did you mention the measurement you had omitted when you described your work to your teacher in a report, presentation or similar?	No	3
	Yes	23
<i>Q9b^b</i> Did you mention the measurements you had omitted when you described your work to your teacher in reports, presentations or similar?	Rarely or never	1
	Sometimes	0
	Often	3
	Always or mostly	17
<i>Q10</i> Have you experienced that a teacher encouraged you to remove an outlier from a data set only based on a feeling that there was a fault in the measurement?	No	60
	Yes	28
<i>Q11</i> Describe the last occurrence, where a teacher encouraged you to remove an outlier from a data set only based on a feeling that there was a fault in the measurement?	[Open-field question]	
<i>Q12</i> What is your opinion of discarding an experiment only because the results seem wrong?	It is a completely illegitimate scientific practice	16
	It is an illegitimate scientific practice	23
	Neutral	24
	It is a legitimate scientific practice	23
	It is a completely legitimate scientific practice	2

^aOnly asked of students who answered 'Yes, one time' to *Q8*

^bOnly asked of students who answered 'Yes, several times' to *Q8*

Results

Quantitative Results

The part of the survey that had to do with lab conduct consisted of two parallel tracks of questions. The first track probed the students' practice in relation to discarding and redoing whole experiments, and the second their practice in connection to deleting outliers. Before the first question the respondents were explicitly instructed to base their answers only on experiences they had had on their current study program and not on experiences they had had in high school or at other university programs.

From the results (see Tables 1, 2) it is clear that a large proportion of the students had been engaged in questionable practices; 27 (30%) of the students had discarded an experiment "only because the results seemed wrong" and 47 (53%) of the students had removed an outlier "only based on a feeling that there was a fault in the measurement" (Q3 and Q8). Of the 91 students in the study, 58 (64%) stated that they had engaged in at least one of the two examined QRPs. A large proportion of the students were (as the students saw it) encouraged by a teacher to engage in these practices; 17 (19%) of the students were encouraged by a teacher to "discard a full or partly completed experiment only because the results seemed wrong" and 28 (32%) were encouraged by a teacher to "remove an outlier from a data set only based on a feeling that there was a fault in the measurement" (Q5 and Q10). These numbers clearly show that the two types of QRPs covered in the survey are a relatively common part of the student lab practice experience; a large minority of students admit to having conducted such practices and it is not uncommon that teachers actively encourage students to engage in the QEPs in question.

Although the two practices were relatively common among the students, there was no clear consensus about how the practices should be perceived from a moral perspective. Using a five-point scale ranging from "completely illegitimate" to "completely legitimate" we asked the students about their opinion of discarding an experiment only because the results seemed wrong (Q7). Forty-five students (51.1%) considered discarding an experiment to be completely illegitimate or illegitimate, while 18 (20.5%) considered it to be legitimate or completely legitimate and 25 (28.4%) were neutral. Using the same five-point scale 39 students (44.3%) considered removing an outlier to be completely illegitimate or illegitimate, while 25 (28.4%) considered it to be legitimate or completely legitimate and 24 (27.3%) were neutral (Q12). These numbers could be read as signalling a deficit in the student's lab related training: although students might occasionally engage in grey zone behaviour or even transgress the accepted norms of scientific practice, they should at least know these norms and have a clear idea about the legitimacy or illegitimacy of actions such as removing an outlier or discarding an experiment. So why was there such a large range of opinions in answers to these two questions?

To explore this question, the responses of the students who engaged in QRPs with those who denied doing so, were compared. Thus, the 27 students answering

positively on having discarded experiments only because the results seemed wrong (see Q3 in Table 1) were compared ($M_1 = 2.65, SD_1 = 1.09$) with the 50 students who answered that they had not done this ($M_2 = 3.69, SD_2 = 0.961$). Unpaired t-tests of these two groups' responses to question Q7 (see Table 1) show clear and significant differences in mean scores between the two groups ($t(45.73) = -4.0838, p = 0.0002$). Likewise, comparing the group of students who state that removed outliers based only on a feeling that it was faulty ($M_1 = 2.94, SD_1 = 1.11$) with the group denying having done so ($M_2 = 3.76, SD_2 = 0.97$) shows significant differences between the groups ($t(86) = -3.70, p = 0.0004$). Thus, the students stating that they have engaged in the grey zone behaviours are also more likely to consider such behaviour legitimate, while students who state that they have not engaged in such actions see them as (more) illegitimate. It appears, that experience with discarding experiments and removing outliers is accompanied by different judgments of whether these actions are acceptable, although from the data at hand only correlation, not causation or direction of causation can be inferred.

In order to probe into the context in which these behaviours took place the students who had answered "yes" to having discarded experiments or removed outliers were asked if they had mentioned the discarded experiments or deleted outliers in the reports or presentations they handed into their teachers (see Q4a and Q4b in Table 1 and Q9a and Q9b in Table 2). Most students reported that they did so. Mentioning the discarded experiment or deleted outlier may be seen as a way to justify the behaviour, and thus turn a perceived illegitimate action into a legitimate one. Whether this justification is sufficient or not is a difficult question that will be addressed further below.

Results from Open Questions

As reported above, a substantial number of students had been encouraged by a teacher to discard experiments (17 out of 89) or to delete outliers (28 out of 88) only because the result seemed wrong. Such encouragement by an authority figure can also justify students' engagement in questionable practices, although this type of justification by authority is itself questionable in certain cases as will be discussed below.

Of the 17 students who had been encouraged by their teacher to discard an experiment, 13 answered the open question and gave a description of the situation in which this occurred, and of the 28 who had been encouraged by their teacher to delete an outlier, 18 answered the open question. Most of these open text responses were quite elaborate. Thus, not only did students describe the situation, they also provided important insights into their thinking about their lab behaviour.

The answers to these open questions were analysed with a grounded approach. All comments were initially coded for content, and subsequently all codes were collected in groups and categorized through an iterative process (Strauss 1987). Separate analyses were done on the questions concerning discarding experiments and

deleting outliers, but the two analyses resulted in the same set of resulting categories of justifications, namely:

- Authoritarian arguments
 - Reference to authority of teacher
 - Reference to authority of theory
- Instrumental arguments
 - Reference to subsequent steps (use of product)
- Empirical arguments
 - Reference to lack of reliability in measurements
 - Reference to lack of continuity with other measurements
- Learning arguments
 - Reference to possibility for learning

Excerpts taken from the comments will be used to describe the content of these categories of justifications. All comments quoted in the following section have been translated from Danish by the authors.

Authoritarian Arguments

Many of the comments relied in part on authority for their justification. This is unsurprising, since the students were being asked to describe the last occurrence where a teacher encouraged them to discard experiments or remove outliers. But the data showed several different variants of authoritarian arguments. One was simply reference to the teacher as an authority in terms of having knowledge the student did not, as in the example below:

A detective exercise, where the teacher knew what substance I had been given, and my experiment showed it to be something different. I was encouraged to discard it and begin from the top, after which I showed the substance to be the right substance. (Respondent 80)

The data set was based on a well-researched phenomenon with a lot of literature on the subject. A single measurement was very deviant and my teacher could see that something had to have gone wrong during this measurement. It happened in the beginning of my education. (Respondent 28)

In other cases, the teacher was rather seen as an authority in terms of having the power to decide what the student should do:

In the physics laboratory. We were asked to 'just' make some new measurements that fit the data better. (Respondent 73) (quotation marks around 'just' were included in the original text)

Or in terms of deciding whether a course could be passed or not:

A report was written about an experiment and the report was delivered to the lab instructor. Some of the results fitted badly, and no good reason could be given for it. The instructor suggested to delete the results, so data fitted better and the report could be accepted. (Respondent 79)

Other types of appeals to authority involved a reference to how well the experiment or data matched the theoretical expectations, in addition to the encouragement by the teacher, as in the following example:

According to my calculations I got a theoretical mass percentage that was substantially lower than the measured value. It was at the end of the experiment, but my instructor thought it should be redone. (Respondent 62)

Instrumental Arguments

Different kinds of justifications revolve around the need to move on with the laboratory work:

I couldn't separate my product from my by-product, because their boiling points were so close. I tried for a full week to separate them. The product I should have gotten was to be used in the next step in the synthesis. (Respondent 8)

Here, the student had to discard the experiment (and presumably get the product from somewhere else) in order to be able to move to the next step in the laboratory work. Notice, that the comment also reflects reliance on theoretical authority (knowledge of the proximity of boiling points).

Empirical Arguments

A third type of justification revolves around mismatches with other empirical data obtained then and there in the laboratory:

A long series of measurements were made, where the first were discarded because they showed presence of oxygen in an oxygen-free mixture. Even though it can't be proved, I am not in doubt that the mistake was due to oxygen in the tubes (even though they had been "flushed" with nitrogen), which is why I think it is correct to exclude data from an experiment that has been tested lots of times. (Respondent 63)

The central point here is that some of the measurements deviated from the "long series" of measurements made. A causal argument (presence of oxygen in the tubes) is provided, backed by a temporal one (only the first measurements showed the deviating behaviour). Finally, the argument that this experiment has been done many times before is provided. This can be seen as an appeal to the authority of previous experimenters.

Several responses provide similar causal explanations, where likely causes for the deviations are provided (for instance measuring at the limits of the apparatus, that

the final concentration of a sample was identical to the initial concentration of the sample, etc.), as in the following example:

We used an assay to determine an agonist's affinity to a given receptor. One sample gave the same result as the reference. We deleted this sample as it most likely did not contain the agonist. (Respondent 11)

The limits and possible defects of measuring equipment can also play a role when students decide whether to trust a deviating result or not, as in this case:

It has only happened once and that was in connection to an assay that consisted of many measurements. One of these rows of measurements deviated significantly from the other measurements of the same. The teacher chose to remove these measurements because this mistake had happened often in this exercise and we deduced that it had to be caused by a defect in the measuring apparatus. (Respondent 15)

In some cases, the empirical arguments refer to lack of reproducibility of the specific measurements as in the following case:

It was said that we should remove the outlier, because we had possibly made a mistake in the measurement, as it seemed unlikely. We made the measurement again, and it [the result] changed. We discarded the first measurement. (Respondent 81)

In other cases, the outlier was identified by comparing with proximal measurements:

In connection to some measurements of lead in paint, where we had to do statistics, we were encouraged by the teacher to remove a measurement because it deviated so much from the other measurements that both we and the teacher believed it to be a measuring mistake. (Respondent 47)

Learning Arguments

Finally, a few students describe the surprising outcomes of experiments and outliers as opportunities for learning. For instance, one student writes elaborately about repeating experiments when something unexpected has happened:

At the moment, we are working on separating different components from snail slime, and on several occasions, we have had to repeat our chromatograms and solubility tests. As we are working with a number of things that are completely new to everyone in the project group this is often helpful either to identify faults or to get 'nicer' results without having identified the source of the error.

In addition to this we are often encouraged to repeat an experiment if we can't find the source of error ourselves as a part of the error identification process. Sometimes the experiment is supervised by an instructor who can help to pinpoint a mistake or a procedure that wasn't described sufficiently clear in the protocol and therefore was conducted in a wrong or insufficient way. In rela-

tion to the project we have also repeated a number of experiments in order to be certain that unexpected results are also constant [similar to the expected results]. (Respondent 4)

This comment expresses empirical arguments (ensuring that results are replicable) as well as an appeal to authority (being urged by the teacher to repeat the experiment), but it also contains the argument that the aim of the endeavour is learning on the part of the students (training in identification of faults) and even learning on the part of the teacher (in order to improve the experimental protocol).

Discussion

The data from the survey show that certain questionable research practices are common in the laboratory practice of chemistry students at Danish Universities as well as in the instructional practice of teachers. However, from the analysis of the qualitative data above, it can be seen that in many cases students have elaborate explanations and justification for removing outliers or discarding experiments, beyond being encouraged by their teachers. When these qualitative responses are included in the analysis, new dimensions of complexity are added to the problem. What this reveals cannot be easily understood as a case of rule violation, where students and teachers fail to observe and behave in accordance with accepted ethical norms. What can be seen is rather a case of conflict, where a rigid conception of the rules of scientific conduct does not correspond to the everyday norms of reasonable laboratory practice.

Empirical Arguments

In the justifications revolving around empirical arguments the conflict between ethical ideals and reasonable laboratory practice in many respects mimics the conflict in the ethical evaluation of QRPs that was identified in the literature above. The students are aware that unforeseen events happen during experimental work, and they use their knowledge about the causal nexus being explored, the experimental equipment being used, and plausible mistakes to critically evaluate the quality of anomalous experiments and measurements. In certain cases, this evaluation leads them to conclude that parts of their data and/or experiments were unsound and had to be discarded. From a practical point of view this is a completely reasonable way of proceeding. This, furthermore, is also how the students perceived it; the students who gave arguments coded as 'empirical' have a strongly more positive attitude towards discarding data than the average. And yet from an ethical point of view discarding data is a questionable practice that is normally considered a breach of scientific integrity.

One might of course demand that the students investigate the reasons for anomalous data further. This is exactly what happened in the justifications labelled above as 'learning arguments'. Here, the anomalous data is considered an opportunity to practice experimental techniques, to investigate the experimental protocol and

further explore the causal nexus involved. In some cases, this may be a fruitful way of proceeding, however, as pointed out by Ravetz (1971, also quoted above), one will never get anywhere if one launches a full investigation each time one gets an unexpected result. From a practical point of view, it is part of the skill set of a good experimenter, to be able to determine when anomalous data is simply due to a mistake and when a deeper investigation is called for.

Authority as Power and Authority as Trust

For the justifications revolving around an appeal to authority, things become even more complicated. In contrast to the empirical arguments, the students here do not only seem to breach the ethical rule that an experimenter should not discard data, but also the methodological rule that scientific inquiry should be based in direct empirical observations and not in dictates from an authority.

This methodological ideal has its origin in empiricist as well as rationalistic epistemologies, where reliance on empirical data and self-inquiry are seen as superior modes of establishing knowledge as opposed to reliance on authority. This narrative is very strong in the sciences, dating back to the scientific revolution and philosopher scientists such as Galileo Galilei and René Descartes (Shapin 1995). In the hierarchy of evidence in the sciences, empirical evidence thus takes prevalence over authority. In Nature-of-Science (NOS) research, reliance on authority has along the same line been identified as a characteristic of fields of study that are 'less'² scientific (Smith and Scharmann 1999). However, it is clear that this story is far from the whole truth. Twentieth-century philosophers of science highlighted the social aspects of scientific activity, including the reliance on authority, and stressed science education as an enterprise that relies heavily on authority to the degree that it may even be considered as fundamentally authoritarian in nature. For instance, historian and philosopher of science Thomas Kuhn (1959) described science education as fundamentally a "dogmatic initiation in a pre-established tradition that the student is not equipped to evaluate" (Kuhn 1959, p. 345). Likewise, chemist and philosopher of science Michael Polanyi (1967) describes how laymen will have to rely on authority in order to examine scientific tenets:

The popular conception of science teaches that science is a collection of observable facts, which anybody can verify for himself. [...] [T]his is not true in the case of expert knowledge, as in diagnosing a disease. But it is not true either in the physical sciences. In the first place, you cannot possibly get hold of the equipment for testing, for example, a statement of astronomy or of chemistry. And supposing you could somehow get the use of an observatory or a chemical laboratory, you would probably damage their instruments beyond repair before you ever made an observation. And even if you should succeed

² In the paper, Mike Smith and Lawrence Scharmann deliberately seek to replace the sharp demarcation between science and non-science with a qualitative continuum, where a given field can be seen as 'more' or 'less' scientific depending on certain characteristics.

in carrying out an observation to check upon a statement of science and you found a result which contradicted it, you would rightly assume that you had made a mistake. The acceptance of scientific statements by laymen is based on authority [...] (Polanyi 1967, pp. 63–64)

These ideas were later developed and expanded to cover the centrality of authority in scientific practice more generally. Philosopher John Hardwig, for instance, points out that scientists are knowers “that stand on each other’s shoulders in the way expressed by the formula: B knows that A knows that p. These knowers could not do their work without presupposing the validity of many other inquiries which they cannot (for reasons of competence as well as time) validate for themselves” (Hardwig 1985 p. 345). Similar ideas can be found in John Coady (1992) and parts of the social epistemology movement (cf. Smith 2013).

From this perspective, reliance on authority (in the form of textbooks and instruction by teachers) is inevitable in science education, and, one may add, understanding the nature of this reliance could well be considered as an independent and important learning goal in science teaching.

It is clear from the answers to the open questions that authority comes in various guises, depending on how it is exercised and how it is perceived by the students. In one form, the students see a teacher or an established theory as a source of authority because they acknowledge (or come to acknowledge) that the teacher possesses knowledge or skills that are superior to their own or because they believe the empirical verification of a known result to be more reliable than their own experiment. This kind of authority is for instance clearly visible in the response by Respondent 28, where the knowledge accumulated in the field is contrasted to the student’s own relative inexperience. In other cases, trust in an authority is not blind, but must be acknowledged as relevant by the students. Respondent 81 describes how an extra measurement convinced the students of the correctness of the teacher’s judgement to remove an outlier (see also Respondent 4).

This use of authority is hardly problematic [depending on how well the students evaluate the epistemic and moral trustworthiness of the source of authority (Hardwig 1985)]. On the contrary, if the ideas proposed by Kuhn, Polanyi and others are correct, then authority in this form is an integral part of science education—and a necessary condition for scientific knowledge production.

However, a radically different form of authority is also present in the data. In this form teachers simply exercise their equipped authority to make the students discard or repeat the experiments the teachers find somehow faulty or lacking (or at least this seems to be the students’ perception). This amounts to authority through the exercise of (sanctioning) power. There are several instances of this in the data. Respondent 73, for instance, describes how the teacher asked them to ‘just’ make some new measurements that better fit the data. This statement suggests that the student did not understand why this had to be done (or considered it unjust, unfair or inappropriate) (and furthermore, in the subsequent question concerning the ethical evaluation of removing outliers (see Q12 in Table 2), the student felt that removing outliers based only on a feeling that they were wrong was completely illegitimate). Similarly, Respondent 79 based the decision to delete outliers in the desire

to get a positive evaluation or grade from the teacher (see also respondent 62). Here the teachers sanctioning power is indirectly visible as the students transgress ethical norms in order to produce the results they believe the teacher will value the most.

These justifications indicate real instructional problems or—lacking the teachers' perspectives—at least, problematic learning outcomes. It is important to remember that the students were asked about situations, where they had been encouraged by their teacher to engage in the questionable behaviours, and in these cases the students are not giving reasonable justifications for an action, except that they were instructed to do it. It is reasonable to be concerned that the students' unintended learning outcome of such situations is that science is a power game, and that cheating is a way of playing it.

This finding is especially worrying if seen in the light of the survey presented in the work of Brian Martinson and colleagues (2005). Here, 15.5% of the respondents admitted to having changed “the design, methodology or results of a study in response to pressure from a funding source” (Martinson et al. 2005, p. 737). This result reflects the changes in the power structure of science that occurred during the transition from the traditional ‘free’ academic science to modern post-academic science, where scientific inquiry to a large degree depends on external sponsoring sources (Ziman 2000). The production of scientific knowledge now relies heavily on institutions and agents that do not always share the epistemic goals of the science (such as truth [or empirical adequacy], objectivity, precision, etc.). On the contrary, funding sources have, on several occasions, used their power to hide or discredit what they perceived to be uncomfortable truths (Michals 2007; Washburn 2005). This poses a real and significant threat to the integrity of scientific knowledge, and for this reason authorities that come in the form of power should be looked upon with concern—and so should instructional practices that (intentionally or unintentionally) reward students for complying with this type of authority.

So, in sum, the conflict the literature expresses concerning researcher behaviour seems to apply to the laboratory practice of students as well. The students in our study engage in certain QRPs on various occasions, but if the justifications the students give are factored in, it seems that at least some of the questionable behaviours can be justified—and so can the teachers' encouragement of the students to engage in them. However, not all of the justifications given by the students are equally valid, and at least some of the cases discussed above are ethically problematic. It appears that some of the students in the survey have come away from their laboratory instruction with the idea that one can deviate from the ideals of science by deleting outliers or discard experiments in order to reach a strategic goal such as verifying a pre-given result or producing the results a perceived authority requires.

In other words, the problem of QRP cannot be handled by prescribing simplified ethical norms or rules at the level of actions. Rather, to follow a point made by philosopher Arthur Fine (1998), procedural and ethical rules such as “Do not remove outliers!” or “Do not discard experiments!” may be considered useful heuristically because they generally lead to increasing trust in findings, but it may sometimes be necessary to disregard the rules in order to be able to trust the result. Thus, in order to teach students how to navigate the ‘grey-zones’ of QRPs it is necessary to move beyond the rules and to turn the students' attention to the justifications they are able

to give for their actions. In some cases, the students may be justified in following the simplified rules, while in others they may be justified in deviating from them.

Educational Implications

The absence of clear-cut behavioural rules constitutes a difficult instructional challenge. How can instruction support student learning of responsible laboratory conduct in the absence of fixed norms or clear rules?

As noted in the data section, answers to Q4 (Table 1) and Q9 (Table 2) show that almost all of the students who had deleted outliers or discarded experiments declared that they had mentioned the deleted data or discarded experiments in their final report. This can be seen as dealing with the grey zone behaviours by replacing one simple rule (do not do it!) with another (if you do it, report it!).

This change of rules may supposedly patch up the gap between the ideals of science and reasonable laboratory practice by allowing students to deviate from the ideal, as long as they are transparent about it. Although transparency is certainly a value to be appraised in laboratory work, such an appeal alone does not solve the problem, it merely displaces it. If a student deletes data that should not have been deleted or fails to discard an experiment that should have been discarded, that mistake is not justified because it is described in the report. It is only corrected when students and teacher actually trust that this was the best way to “get things right” (see Fine 1998). Transparency in reporting should be subservient to that goal – not the other way around.

The divergence of student opinions about the acceptability of deleting outliers and discarding experiments and the great variety in the quality of the justifications the students in the survey gave for deleting outliers and discarding experiments should be taken as an indication that the students will not necessarily learn how to navigate anomalous data simply by being exposed to (proper) laboratory practice. Thus, exposure to inquiry-based lab practice is not enough. As pointed out by Fouad Abd-El-Khalick (2013), courses aiming to teach students about the nature of science through inquiry-based instruction often fail in the attempt; evidence indicates that inquiry experiences do not in and by themselves change the students' epistemological ideas about the Nature of Science (NOS). From this, Abd-El-Khalick (2013) concludes that “an explicit-reflective framework is needed to achieve the goal of improving understandings about NOS among science teachers and students” (Abd-El-Khalick 2013, p. 2090). In other words, explicit NOS learning goals and activities promoting reflective practice must be embedded in the curriculum. The explicit inclusion of NOS objectives in the learning goals is not only intended to support the students' learning. It is also needed in order for teachers to allow themselves to pursue the students' acquisition of NOS understanding—and not just covering the science content.

At least some of these points made by Abd-El-Khalick can be transferred to the university context at hand. The results presented above seem to confirm Abd-El-Khalick's basic diagnosis; even though the students in the survey of chemistry students have been engaged heavily in lab practice for at least one and one-half

years of study, far from all of these students have developed clear understandings of how to handle anomalous data.

Following Abd-El-Khalick, establishing explicit-reflective frameworks may be a way to proceed. In such frameworks, appropriate treatment of anomalous data should be included as explicit learning goals in (at least some) courses followed by the students, and students should be provided the time and theoretical resources needed for them to verbalize and discuss their experiences. Clark Chinn and William Brewer (1993) provide a list of possible responses to anomalous data, which they consider to be “close to an exhaustive list”. They describe the responses with examples from science education as well as from the history of science. The list of responses (rephrased from Chinn and Brewer 1993) is outlined below:

1. Ignore the anomalous data
2. Reject the anomalous data
3. Exclude the data from the domain of the theory
4. Hold the data in abeyance
5. Reinterpret the data while retaining the theory
6. Make peripheral changes to theory
7. Accept the data and reject the theory, possibly in favour of a different theory

Such categorical frameworks—in an elaborated form—may provide a good starting point for reflective exercises on how to proceed in a scientifically sound way when facing anomalous data (outliers and ‘weird’ experimental outcomes). Moreover, such categorical frameworks may also assist teachers in making clear and explicit their own (possibly tacit) justifications for encouraging or discouraging rejection of anomalous data. A relevant future research path would be to explore the effects of developing and using such categorical frameworks as a way to handle anomalous data in university teaching—or explore the current and “best practices” among science and engineering teachers.

Conclusion

The survey of Danish chemistry students shows that more than half of the students engage in scientific practices that can be seen as questionable (64%), and that a large minority of the students have been encouraged by their teachers to engage in such practices (respectively 19% and 32% depending on the practice). The quality of the students’ justification for engaging in these types of practices differ widely and range from simplistic authority arguments to comprehensive ideas involving reference to the causal nexus under investigation, the student’s own capabilities as an experimenter and the need to promote learning. Although some of these arguments can be said to provide reasonable justification for the actions in question, at least from the perspective of reasonable laboratory practice, not all of the

arguments are equally valid. Especially, a subgroup of authority arguments give reason for concern. Here, students seemingly discarded data points or experiments only as a way to comply with the invested power of a teacher.

There is no consensus among students whether the questionable practices discussed here are scientifically acceptable or not, although students who admit to having engaged in these practices tend to evaluate them more positively than students who deny having engaged in them.

These results point to an instructional challenge; a large group of students - perhaps tacitly - adhere to an impossible ideal of science, while another group of students find that this very ideal is directly contradicted by their lab teacher. The ideal of science provides the students with little guidance when they are faced with anomalous data. In particular, it does not support the students in deciding whether discarding data is justified or not in any given situation and it does not help the students to develop the sound judgement needed in order to handle such difficult choices. The survey data contains clear examples where students lack sound judgement and discard data for clearly unacceptable reasons such as the wish to comply with power figures. This situation points to the need for more direct and realistic instruction about the nature of science. To teach students responsible lab practice requires movement from the level of actions (this action is right, and this is wrong) to the level of justification. Students need to learn what kinds of justifications are acceptable and what kinds of justifications are not. Or in other words, they need to be taught sound and informed judgements concerning their lab practice.

The result on the other hand also shows that real cases taken from the students own laboratory practice can be a valuable resource for such instruction. As seen above, some students are good at reasoning about their own practices, and experiences from their laboratory practice are in fact a way to change tacit (and outdated) assumptions about the nature of science. However, teachers should not expect such reasoning to materialize by itself; rather, they will have to employ a flexible reflective framework wherein the situations actually encountered by students can be grasped.

References

- Abd-El-Khalick, F. (2013). Teaching with and about nature of science, and science teacher knowledge domains. *Science & Education*, 22, 2087–2107.
- ABET (2017). Criteria for accrediting engineering programs—effective for reviews during the 2017–2018 accreditation cycle. *Engineering Accreditation Commission*. <http://www.abet.org/wp-content/uploads/2016/12/E001-17-18-EAC-Criteria-10-29-16-1.pdf>. Retrieved from 21 March 2018.
- Belter, R. W., & du Pré, A. (2009). A strategy to reduce plagiarism in an undergraduate course. *Teaching of Psychology*, 36(4), 257–261.
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1–49.
- Coady, C. A. J. (1992). *Testimony: A Philosophical Study*. Oxford: Clarendon Press.
- Davis, S., Drinan, P. F., & Gallant, T. B. (2009). *Cheating in School, What we Know and What we Can Do*. Chichester: Wiley-Blackwell.
- De Vries, R., Anderson, M., & Martinson, B. C. (2006). Normal misbehavior: Scientists talk about the ethics of research. *Journal of Empirical Research on Human Research Ethics: An International Journal*, 1(1), 43–50.

- Dobzhansky, T. (1967). Review: Looking back at Mendel's discovery. *Science, New Series*, 156(3782), 1588–1589.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4(5), e5738. <https://doi.org/10.1371/journal.pone.0005738>.
- Farley, J., & Geison, G. (1974). Science, politics and spontaneous generation in nineteenth-century France: The Pasteur–Pouchet debate. *Bulletin of the History of Medicine*, 48(2), 161–198.
- Fine, A. (1998). The viewpoint of no one in particular. In *Proceedings and Addresses of the American Philosophical Association*, 72(2), pp. 7–20.
- Harding, T. S., Carpenter, D. D., Finelli, C. J., & Passow, H. J. (2004). Does academic dishonesty relate to unethical behavior in professional practice? An exploratory study. *Science and Engineering Ethics*, 10(2), 311–324.
- Hardwig, J. (1985). Epistemic dependence. *Journal of Philosophy*, 82(7), 335–349.
- Henslee, A., Murray, M., Olbricht, S., Ludlow, L., Hays, G., & Nelson, R. (2017). Assessing freshman engineering students' understanding of ethical behavior. *Science and Engineering Ethics*, 23(1), 287–304.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88(1), 28–54.
- Holton, G. (1978). Subelectrons, presuppositions, and the Millikan–Ehrenhaft Dispute. *Historical Studies in the Physical Sciences*, 9, 161–224.
- Hughes, J. M. C., & McCabe, D. L. (2006). Academic misconduct within higher education in Canada. *Canadian Journal of Higher Education*, 36(2), 1–21.
- John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532.
- Kipnis, N. (2011). Errors in science and their treatment in teaching science. *Science & Education*, 20(7–8), 655–685.
- Kuhn, T. S. (1959). The essential tension: Tradition and innovation in scientific research. In C. W. Taylor & F. Barron (Eds.), *Scientific Creativity: Its Recognition and Development* (pp. 341–354). New York: John Wiley & Sons Inc.
- Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature*, 435(7043), 737–738.
- McCabe, D. L. (1997). Classroom cheating among natural science and engineering majors. *Science and Engineering Ethics*, 3(4), 433–445.
- McCabe, D. L. (2005). Cheating among college and university students: A North American perspective. *International Journal for Educational Integrity*, 1(1). <https://www.ojs.unisa.edu.au/index.php/IJIEI/article/view/14>. Retrieved from 21 March 2018. <http://dx.doi.org/10.21913/IJIEI.v1i1.14>.
- McCabe, D. L., Trevino, L., & Butterfield, K. (2001). Cheating in academic institutions: A decade of research. *Ethics and Behavior*, 11(3), 219–232.
- Merriam-Webster (2017). Online dictionary (<https://www.merriam-webster.com>). The online dictionary was accessed October 6 2017 with the search key “questionable”.
- Michaels, D. (2007). *Doubt is their product, how industry's assault on science threatens your health*. Oxford: Oxford University Press.
- Miñano, R., Uruburu, Á., Moreno-Romero, A., & Pérez-López, D. (2017). Strategies for teaching professional ethics to IT engineering degree students and evaluating the result. *Science and Engineering Ethics*, 23(1), 263–286.
- Mumford, M. D., Connelly, S., Brown, R. P., Murphy, S. T., Hill, J. H., Antes, A. L., et al. (2008). A sensemaking approach to ethics training for scientists: Preliminary evidence of training effectiveness. *Ethics and Behavior*, 18(4), 315–339.
- Polanyi, M. (1967). *The Tacit Dimension*. London: Routledge & Kegan Paul Ltd.
- Ravetz, J. R. (1971). *Scientific Knowledge and its Social Problems*. Oxford: Clarendon Press.
- Reid, N., & Shah, I. (2007). The role of laboratory work in university chemistry. *Chemistry Education Research and Practice*, 8(2), 172–185.
- Shapin, S. (1995). Trust, honesty and the authority of science. In R. E. Bulger, E. Meyer Bobby, & H. V. Fineberg (Eds.), *Society's Choices: Social and Ethical Decision Making in Biomedicine* (pp. 388–408). Washington: National Academies Press.
- Smith, M. U. (2013). The role of authority in science and religion with implications for science teaching and learning. *Science & Education*, 22(3), 605–634.
- Smith, M. U., & Scharmann, L. C. (1999). Defining versus describing the nature of science: A pragmatic analysis for classroom teachers and science educators. *Science Education*, 83(4), 493–509.

- Steneck, N. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics, 12*(1), 53–74.
- Strauss, A. L. (1987). *Qualitative Analysis for Social Scientists*. New York: Cambridge University Press.
- Swazey, J. P., Anderson, M. S., & Lewis, K. S. (1993). Ethical problems in academic research. *American Scientist, 81*(6), 542–553.
- VanDeGrift, T., Dillon, H., & Camp, L. (2017). Changing the engineering student culture with respect to academic integrity and ethics. *Science and Engineering Ethics, 23*(4), 1159–1182.
- Washburn, J. (2005). *University Inc., The Corporate Corruption of American Higher Education*. New York: Basic Books.
- Wu, C.-F. (2003). A study of the adjustment of ethical recognition and ethical decision-making of managers-to-be across the Taiwan Strait before and after receiving a business ethics education. *Journal of Business Ethics, 45*(4), 291–307.
- Zigmond, M. J., & Fischer, B. A. (2002). Beyond fabrication and plagiarism: The little murders of everyday science. *Science and Engineering Ethics, 8*(2), 229–234.
- Ziman, J. (2000). *Real Science: What it is and what it means*. Cambridge: Cambridge University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.