Teacher's Handbook Measurement Uncertainties

Karel Kok*, Burkhard Priemer
Humboldt-Universität zu Berlin

May 12, 2026

Contents

1 Introduction **2**

2 Concepts **3**
2.1 Point and Set Paradigm [ ] 3
2.2 True Value [ ] 4
2.3 Repeated Measurements [ ] 6
2.4 Mean Value [ ] 8
2.5 Uncertainty [ ] 9
2.6 Measurement error (Deviation) [ ] 12
2.7 Measurement Result [ ] 13

3 Determining the Uncertainty **16**
3.1 Type A [ ] 16
3.2 Type B [ ] 18
3.3 Graphs [ ] 19
3.4 Uncertainty Propagation [ ] 21
3.5 Relative uncertainty [ ] 23

4 Data Comparison **25**
4.1 Simple Comparison [ ] 25
4.2 Indicators [ ] 26
4.3 Comparing Multiple Results [ ] 28

*karel.kok@physik.hu-berlin.de

1 INTRODUCTION

Welcome to this digital handbook on the topic of measurement uncertainties. This is the teacher's manual giving conceptual as well as didactical background to the digital learning environment (DLE) for students that can be found here: https://lernen.physik.hu-berlin.de/measurementuncertainties. After working through this manual, you will:

- know about the underlying concepts of measurement uncertainties;
- know about students' preconceptions of measurement uncertainties;
- know some strategies to address these preconceptions;
- know some didactical strategies to introduce the topic of measurement uncertainties;
- have some practical examples to address measurement uncertainties;
- have seen some concepts that go beyond the DLE.

This support manual is conceptually oriented, focusing not on numbers and definitions, but on the underlying principles, meaning, and interpretation of the concepts. The concepts are introduced based on real-world examples and data.

The manual consists of three parts, that are divided into several steps. Each step consists of several thematical blocks. These blocks are labeled with symbols. These symbols indicate what type of content is discussed in the block:

- Textbook basics. This is the subject matter where definitions are supplemented with their meaning and interpretation.
- Didactics. This is where you will find students' views and preconceptions, ideas for teaching strategies, and ideas for didactic reduction that will help you address the topic of measurement uncertainties in your teaching.
- Examples. This is where you will find (links to) practical examples or further information that connect to the contents of the step.
- Advanced concepts. These are some optional concepts that go beyond the concepts of the students' DLE but might be essential for certain experiments. Advanced concepts are also labeled with a or to indicate whether it addresses the basics or the didactical part. Since these concepts are not addressed in the students' DLE, you will have to introduce these concepts yourself if you want your students to work with these concepts!

In the first part, the basic concepts regarding measurement uncertainties are introduced. In a conceptual approach, you will learn about the fundamental concepts as well as students' views and preconceptions about these concepts.

In the second part, the determination of the uncertainty is discussed. Here, you will learn about calculating the measurement uncertainty, graphing, propagation, and analyzing the uncertainty budget. This subject matter will be supplemented by a didactical analysis in which you will get to know some didactical strategies and ideas for didactical reduction.

The last part focuses on the application of measurement uncertainties in terms of data comparison. This entails the formal rules for data comparison as well as a didactical tool to identify students' views on measurement uncertainties in data comparison problems.

Table 1: Some contrasting characteristics of the point and set paradigms.

Point paradigm	**Set paradigm**
Every experiment has a "true value" that can be determined when taking the correct procedures with the correct equipment.	In an experiment, a certain measurand can be determined. This measurand is a precise as possible description of what is to be measured.
Uncertainties can be eliminated, avoided, or reduced to zero. They are a sign of something having gone wrong.	Uncertainties are omnipresent in every experiment. The goal is to control them, reduce them to a desired level, and to quantify them. They are an indication of the quality of an experiment.
Repeated measurements are taken to: confirm an outcome, practice measuring, and be able to calculate a mean value.	Repeated measurements are taken to estimate the spread in measurements and quantify the measurement uncertainty.
The mean value represents the "true value" of an experiment.	Every measurement result—regardless if it comes from a single measurement or multiple measurements—contains a best value (often the mean value) and an uncertainty interval (the range of values) around it.

2 CONCEPTS

2.1 Point and Set Paradigm [ ]

Research on students' understanding of measurement uncertainties has identified two paradigms, see Tab. 1: the point and set paradigm [1]. These paradigms describe two extremes in students' viewpoints regarding measurement data. This part starts with a description of the two paradigms which will then be used as a foundation to talk about students' difficulties regarding measurement uncertainties.

Point paradigm [ ]

In the point paradigm, students think about measurements as being single isolated events that, given the right instrumentation and practice, lead to the "true value" that can be measured in an experiment. Measurement uncertainties are the result of imperfections in the measurement instruments and/or procedures that can, in principle, be eliminated.

Set paradigm [ ]

In contrast, in the set paradigm, students look at a dataset as a whole. The fluctuations in the measurements are used to determine the measurement uncertainty and are, hence, an indication of the data quality. Students let go of the idea of a true value but rather try and determine the uncertainty interval which can be seen as a set of values. Hence, a result of a measurement can never be a single value but rather a range of values.

Goal [ ]

Research has repeatedly found that the vast majority of students, even at the university level, are firmly located in the point paradigm. The goal of many first-year lab courses is to shift students' thinking from the point toward the set paradigm. Research has also shown that this is best done by first addressing the fundamental underpinnings of measurement uncertainties [2, 3], before moving toward statistical procedures. This approach is also chosen in the Digital Learning Environment (DLE) that was developed for students in secondary education.

Table 1 shows some contrasting ideas associated with the point and set paradigms that students have. The concepts in this table will be further elaborated on in the next subsections.

Relevancy of measurement uncertainties [ ]

Depending on the goal of the experiment, the evaluation of the measurement uncertainty is a necessity for the success of the experiment. In all cases where conclusions are data-based, an analysis of the measurement uncertainty is required. In other instances, e.g., when the relation between two variables is to be illutrated or when general trends are shown, this evaluation is not needed at all. To read more about this, see Kok et al. [4].

For teaching strategies on measurement uncertainties, see Holz and Heinicke [5].
https://www.youtube.com/watch?v=uMvT02mHkss

References

[1] Buffler, A., Allie, S., & Lubben, F. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education*, *23*(11), 1137–1156. https://doi.org/10.1080/09500690110039567

[2] Volkwyn, T. S., Allie, S., Buffler, A., & Lubben, F. (2008). Impact of a conventional introductory laboratory course on the understanding of measurement. *Physical Review Physics Education Research*, *4*(1), 010108. https://doi.org/10.1103/PhysRevSTPER.4.010108

[3] Séré, M.-G., Journeaux, R., & Larcher, C. (1993). Learning the statistical analysis of measurement errors. *International Journal of Science Education*, *15*(4), 427–438. https://doi.org/10.1080/0950069930150406

[4] Kok, K., Boczianowski, F., & Priemer, B. (2020). Messdaten im Physikunterricht auswerten – wann sind Messunsicherheiten wichtig? *MNU Journal*, *73*(4), 292–295. https://doi.org/10.18452/27175

[5] Holz, C., & Heinicke, S. (2020). Tipps für Lehrkräfte. *Unterricht Physik*, *31*(177/178), 39–43.

2.2 True Value [ ]

Suppose one is interested in the falling time of an object that is dropped from a height of 1 m. This time can easily be calculated using the equation: $h = \frac{1}{2}gt^2 \rightarrow t = \sqrt{\frac{2h}{g}} = \sqrt{\frac{2\cdot 1\,\mathrm{m}}{9.81\,\mathrm{m\,s^{-2}}}} = 0.451\,523\,6\ldots$ s.

In principle, this falling time can be calculated precisely. However, when experimentally verifying this exact number, problems emerge. Human reaction time will influence the measurements and no perfect measurement instruments exist. One can start to wonder if the theoretical result can be confirmed experimentally.

Measuring to absolute precision [ ]

To confirm this falling time, an experiment needs to be done. The setup of the experiment has to be such that the object can fall 1 m. This is easier said than done. Using a tape measure, the height can only be measured up to a precision of 1 cm. This means that one can "only" be sure that the height is between 1.005 m and 0.995 m. One can decrease this interval by taking a more precise measurement instrument, but the precision will always have a limit due to the resolution of the instrument. Furthermore, this assumes that the measurement instrument has been calibrated exactly. This, of course, can never be the case since it would require another perfect instrument to be compared to.

Not only the instrument, but also the surroundings will affect the measurement. Suppose that in the setup a ball is dropped from a table. A temperature change will cause the steel legs of the table to expand or contract. This effect is negligible for everyday measurements, but to acquire absolute precision, this effect has to be taken into account.

Ultimately, one can never be one hundred percent sure that the experimental setup has a height of *exactly* $1.\overline{000}$ m.

And remember, these are only the problems for the setup. The falling time also needs to be measured to absolute precision, one needs to have an absolute value for g, account for air resistance, etc.

The measurand [ ]

Of course, this example is a bit exaggerated. However, it illustrates the problems of measuring something to *absolute precision*. At some point, measurement uncertainties will *always* affect measurements in the real world. Therefore, these conditions have to be included in the definition of the quantity that is being measured: the falling time of an object from a height of 1 m, at a temperature of 20 °C, at longitude x and latitude y, ... This description of the quantity that is being measured and its surrounding conditions that affect its value is called the **measurand**. Since one is unable to define the measurand in absolute detail, i.e., all the experimental conditions (that can have uncertainties themselves), this means that a **"true value"** of the measurand—an idealized result with zero uncertainty—cannot exist! Note that this inexistence goes beyond the impossibility of not being able to determine the "true value". Hence, the term "true value" is not used any more in metrology (the science that is concerned with measuring).

For good scientific practice, results have to be reproducible. Therefore, when reporting an experiment, one has to clearly define the measurand. This description is often done in measurement equations but should also include what was measured and report the conditions that could have affected the measurement result. Furthermore, the measurement uncertainty, a quantification of the variability in measurement results, should be reported. This shows the precision of the measurement result (a small uncertainty indicates a precise result) and is an indication quality of the experiment.

There are cases in which certain information about the measurand can be omitted. In the example before, the air temperature could have affected the falling time of the object. However, using everyday measurement instruments, the uncertainty in height and time measurements are much larger than the uncertainty due to temperature fluctuations. Therefore the influence of temperature changes will be impossible to measure and can be omitted from the measurand.

Conditions that can affect the measurement result should be reported. For instance, when stating the height of the table, one should include the uncertainty. When the time measurements are done by hand, an estimation of the reaction time should be reported and integrated into the evaluation of the uncertainty.

Students preconceptions about the "true value" [ ]

The idea of the existence of a "true value" is very persistent among learners [1, 2, 3, 4]. Students often pursue to obtain the "exact answer" and want to know whether their result is correct. This belief in a single and exact number coincides with point paradigm thinking.

Although some students might understand that their measurement result has an uncertainty, they still might think of reference values as a "true value" to compare their own result with. Unfortunately, the uncertainty of these reference values is very often not reported in school books. Although the uncertainty is not reported, this does not mean that the reference values do not have an uncertainty. Sometimes, the reported values are reported up to a decimal place that remains unaffected by its uncertainty. For instance the value $G = 6.674 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$ has fewer decimal places than the known value of $G = (6.674\,30 \pm 0.000\,15) \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$. Despite practical reasons to omit the uncertainty, students should be aware that all reference values (apart from seven fundamental constants that are defined as absolute quantities [5]) have been measured or rely on measurements and thus have an uncertainty. Some helpful prompts for students could be:

- How is the reference value determined?
- Was this measurement subject to measurement uncertainties?
- How certain are you about your measurement?
- You say U has a value of 5 V, how certain are you that it's not 5.1 V?
- Describe to me exactly what you have measured in your experiment.
- Are there any sources of uncertainties in your setup?
- Discuss with the person sitting next to you how much sense it makes to talk about the true value of your height.
- Can you estimate the precision of your measurement instruments?
- Find a way to describe—in utmost detail—how to measure the time it takes you to get to school as exactly as possible.

Using these prompts, students can be guided toward a set paradigm understanding. They start to look critically at their experiment. Ultimately, they start to describe in finer detail what they are measuring in terms of the measurand. Also, they start to express the limitations of their measurement result in the form of an uncertainty interval.

How long is a banana? [💡]

For a practical example of how to introduce the measurand to students, see: [6].

References

[1] Allie, S., Buffler, A., Campbell, B., & Lubben, F. (1998). First-year physics students' perceptions of the quality of experimental measurements. *International Journal of Science Education*, *20*(4), 447–459. https://doi.org/10.1080/0950069980200405

[2] Fairbrother, R., & Hackling, M. (1997). Is this the right answer? *International Journal of Science Education*, *19*(8), 887–894. https://doi.org/10.1080/0950069970190802

[3] Lubben, F., Campbell, B., Buffler, A., & Allie, S. (2001). Point and set reasoning in practical science measurement by entering university freshmen. *Science Education*, *85*(4), 311–327. https://doi.org/10.1002/sce.1012

[4] Smith, E. M., Stein, M. M., & Holmes, N. G. (2020). How expectations of confirmation influence students' experimentation decisions in introductory labs. *Physical Review Physics Education Research*, *16*(1), 010113. https://doi.org/10.1103/PhysRevPhysEducRes.16.010113

[5] International Bureau of Weights and Measures. (2019). *The International System of Units* (9th ed.). Le Bureau international des poids et mesures.

[6] Musold, W., & Kok, K. (2024). Wie lang ist die Banane? — Über die Relevanz, eine Messgröße zu definieren. In B. Priemer & K. Kok (Eds.), *Messunsicherheiten im Physikunterricht* (1st ed.). Berlin Universities Publishing. https://doi.org/10.14279/depositonce-21608

2.3 Repeated Measurements [📖 🛠]

Measurements are repeated so that one gets an idea about the variability of the measurements. Given enough repeated measurements, these measurements will follow the normal distribution with a mean value and a standard deviation. The question arises of how many measurements are required.

Reasons for repeated measurements [📖]

Some scientists have developed an intuition as for to when to stop, others use rules of thumb, again others continue measuring to increase precision. If one is interested in the variability of the measurements, one will want to know the mean and the standard deviation. Since the standard deviation is not affected by the number of measurements (see 3.1), one can stop taking measurements as soon as the distribution adequately fits the normal distribution. The latter is usually done with a statistical test, which is beyond the scope of this unit.

Alternatively, one can look for changes in the value of the standard deviation and the mean value with additional measurements. When the first two significant digits of the standard deviation (and the corresponding digits of the mean) do not change with additional measurements, this is an indication that the data reasonably fits a normal distribution.

Students' ideas about repeated measurements [🛠]

The scientific reasons to repeat measurements can only be understood once a correct understanding of measurement uncertainties has been established. It is therefore no surprise, that the reasons students give for why they need to repeat their measurements differ greatly from those described above.

Taking students' belief in a "true value" into account, it becomes clear that students do not see a need to repeat measurements. If you measure correctly, you will get the correct value. Some students will measure again to confirm their result [1]. When the second measurement is not numerically identical, some students become perplexed or get the feeling of having done something wrong. Others will continue to measure until recurring values are found.

Table 2: Example dataset: with the rule of thumb, one can stop measuring after seven measurements.

Measurement	Time [s]
1	2.18
2	3.02
3	2.82
4	3.61
5	2.47
6	3.31
7	2.24

For some students, the order of the measurements has an important meaning. Some students place more trust in the first measurement since the setup might wear after measuring [2]. Although this might actually happen, this means that the setup needs to be fixed after every measurement.

Another practice found is that students see repeated measurements as a way of practicing their measurement technique [3]. Afterward, these students report the last measured value since they had the most practice in measuring when recording this measurement. In itself, taking some measurements for practice is a good strategy—one knows what to expect and how to take the measurements properly. However, if these measurements are truly practice rounds, they should not be part of the dataset, since mistakes are expected to occur.

Still, other students take repeated measurements by routine: they take six measurements because they were told to do so, because they did that last time, to be able to calculate a mean value, or because "more is just better" [2, 4, 5].

All these reasonings can be associated with the point paradigm. Students rely on single values and hence see no reason for a set of repeated measurements.

The reasons to take repeated measurements and when to stop are very complex. Probably too complex for secondary education students, especially knowing when to stop taking measurements. However, students can have a conceptual understanding of why repeated measurements are taken: to get an idea about the variability of the measurements. As for when to stop, simple rules of thumb can be used.

Some rules of thumb [ ]

The most simple rule of thumb is to simply state how many measurements to record. Usually, six to ten measurements can give a pretty good idea about the variability of the data and start to approach a normal distribution. Simulations of several alternative quantifications of the uncertainty (see Kok and Priemer, Kok and Priemer [6, 7]) showed eight repeated measurements to be a favorable number. The exclude extremes quantification, see Sec. 3.1, approximates the value of the standard deviation with this many repeated measurements.

A more sophisticated rule of thumb is to stop taking measurements until three consecutive measurements are not smaller/larger than the smallest and largest measurement in the dataset. Table 2 shows an example dataset of seven repeated measurements. After measurement number four, the next three consecutive measurements (5–7) are between the minimum (2.18 s) and maximum (3.61 s) of the first four measurements in the series. One could decide to stop measuring after the seventh measurement.

One major issue with this rule is that the order of measurements becomes an (explicitly) important aspect. Despite this issue, the dialog of why this rule is used will pave the way toward set paradigm understanding: one wants to see the variance and one wants to know when this is good enough. After learners have fully grasped the concepts of the mean value and the uncertainty, they can start to see the reasoning behind the underlying reasoning regarding repeated measurements.

References

[1] Lubben, F., & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education*, *18*(8), 955–968. https://doi.org/10.1080/0950069960180807

[2] Séré, M.-G., Journeaux, R., & Larcher, C. (1993). Learning the statistical analysis of measurement errors. *International Journal of Science Education*, *15*(4), 427–438. https://doi.org/10.1080/0950069930150406

[3] Warwick, P., Linfield, R. S., & Stephenson, P. (1999). A comparison of primary school pupils' ability to express procedural understanding in science through speech and writing. *International Journal of Science Education*, *21*(8), 823–838. https://doi.org/10.1080/095006999290318
[4] Buffler, A., Allie, S., & Lubben, F. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education*, *23*(11), 1137–1156. https://doi.org/10.1080/09500690110039567
[5] Ryder, J., & Leach, J. (2000). Interpreting experimental data: The views of upper secondary school and university science students. *International Journal of Science Education*, *22*(10), 1069–1084. https://doi.org/10.1080/095006900429448
[6] Kok, K., & Priemer, B. (2022). Comparing Different Uncertainty Measures to Quantify Measurement Uncertainties in High School Science Experiments. *International Journal of Physics and Chemistry Education*, *14*(1), 1–9. https://doi.org/10.48550/arXiv.2205.04102
IJPCE doi: 10.51724/ijpce.v14i1.214.
[7] Kok, K., & Priemer, B. (2023). Messunsicherheiten quantifizieren: Welche Maße gibt es dafür? *MNU Journal*, *76*(4), 330–333. https://doi.org/10.18452/27043

2.4 Mean Value [📘 ✎]

The arithmetic mean value of a series of measurements can be calculated as:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{1}$$

where N is the number of repeated measurements and x_i are the individual measurements.

The mean value is considered the **best estimation** of the value of the measurand from a series of measurements—the most central value of the series. One can also think of it like this: if one had to guess what the next measurement would be, the mean value will—on average—be closest to this new measurement.

This coincides with the idea that one can stop measuring when the mean value (and the uncertainty) have stabilized. At that point, the mean value coincides with the most central value of measurements that the experiment produces.

The arithmetic mean is most often used as the best estimation of the measurand. However, there are alternative quantifications. One possibility is to use the **median**: the middle value of a sorted dataset. Another measure is the **modus**: that measurement that most often occurs. Lastly, in the case when repeated measurements all give the exact same result, the best estimation is just that one value.

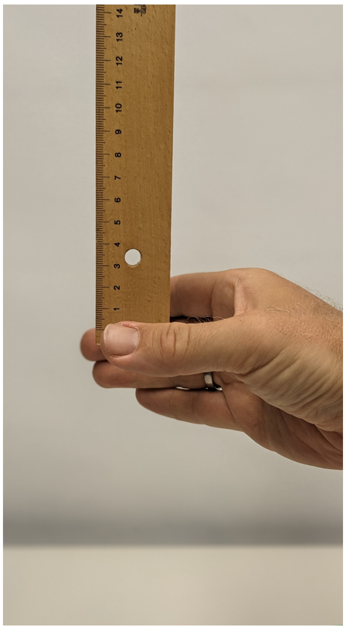
Suppose, one wants to measure the reaction time of a student. This can be done by measuring the falling distance of a ruler, see Fig. 1a. Using the equation $h = \frac{1}{2}gt^2$ one can determine the reaction time. The measured data is shown graphically in Fig. 1b. The most central point of these measurements is the mean value, which is indicated in Fig. 1c as a red cross.

Students' ideas about the mean value [✎]

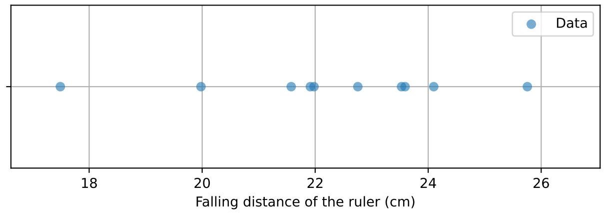
Calculating the mean value usually does not pose a problem for students. However, students' reasoning about what the mean value represents often remains superficial [1, 2].

Sometimes the mean value is mentioned as a reason for gathering repeated measurements. Although, of course, one has to collect multiple measurements to calculate a mean value [3], this is often a routine and automated procedure and the mean value is regarded as an isolated value. Some students go even further and will regard this value as a "true value" [4]. Although the process of calculating a mean value from a series of repeated measurements can be associated with the set paradigm, regarding the mean value as an isolated value is considered point paradigm thinking [5]. To move to an adequate set paradigm reasoning, students need to supplement it with the uncertainty.

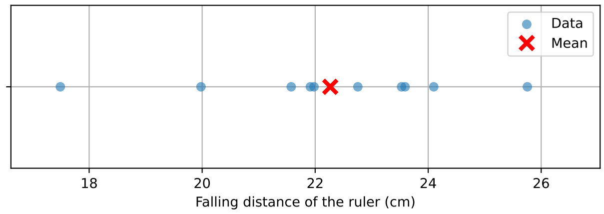
Another idea about the mean value that exists, is that it is "more precise" [6]. Although a (stabilized) mean value is the best estimation of the value of the measurand, this does not say something about precision—which is indicated by the uncertainty.



(**a**) Measuring the falling distance of a ruler to determine the reaction time.



(**b**) Ten repeated measurements of the falling distance of a ruler.



(**c**) The mean value (red cross) is the most central point of this dataset.

Figure 1: Measurement data of an experiment in which the reaction time is determined by measuring the falling distance of a ruler.

References

[1] Majiet, N., & Allie, S. (2019). Student understanding of measurement and uncertainty: Probing the mean. *2018 Physics Education Research Conference Proceedings*. https://doi.org/10.1119/perc.2018.pr.Majiet

[2] van Kampen, P., & Gkioka, O. (2021). Undergraduate students' reasoning about the quality of experimental measurements of covarying secondary data. *European Journal of Physics*, *42*(4), 045704. https://doi.org/10.1088/1361-6404/abfd27

[3] Allie, S., Buffler, A., Lubben, F., & Campbell, B. (2002). Point and Set Paradigms in Students' Handling of Experimental Measurements. In H. Behrendt, H. Dahncke, R. Duit, W. Gräber, M. Komorek, A. Kross, & P. Reiska (Eds.), *Research in Science Education - Past, Present, and Future* (pp. 331–336). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47639-8_47

[4] Ford, M. J. (2005). The Game, the Pieces, and the Players: Generative Resources From Two Instructional Portrayals of Experimentation. *Journal of the Learning Sciences*, *14*(4), 449–487. https://doi.org/10.1207/s15327809jls1404_1

[5] Buffler, A., Allie, S., & Lubben, F. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education*, *23*(11), 1137–1156. https://doi.org/10.1080/09500690110039567

[6] Coelho, S. M., & Séré, M.-G. (1998). Pupils' Reasoning and Practice during Hands-on Activities in the Measurement Phase. *Research in Science and Technological Education*, *16*(1), 79–96. https://doi.org/10.1080/0263514980160107

2.5 Uncertainty [ ]

The uncertainty of a measurement is a quantification of the variability of the measurements in an experiment. In most cases, the values of repeated measurements in an experiment follow the normal distribution. Therefore, it comes as no surprise that the standard deviation (or the standard deviation of the mean) is the most common scientific quantification of the uncertainty.

Of course, there are measurements in which no variability appears (for instance measuring the length of a table with a tape measure). Here, the uncertainty is determined by the measurement instrument, more on that in Sec. 3.2.

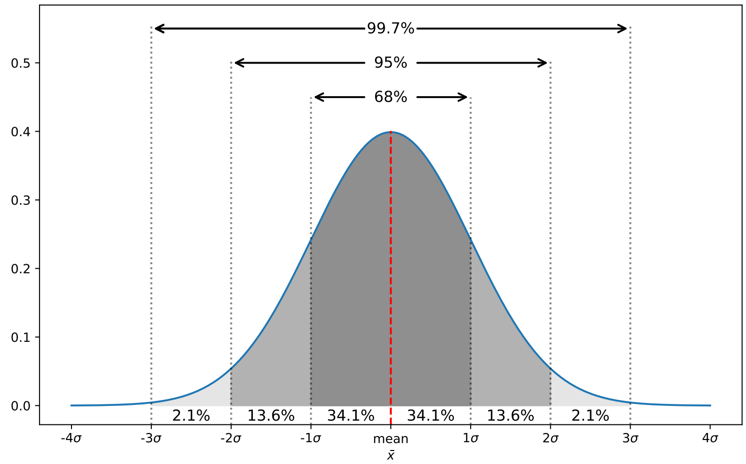


Figure 2: A normal distribution with a mean value at zero (red dashed line) and standard deviation σ. The dotted lines indicate the one, two, and three sigma boundaries. The percentages describe the how many measurements can be expected in the corresponding intervalls.

Standard deviation of a distribution [📘]

The standard deviation of a distribution is the scientific standard for describing the variability of a measured variable around its mean value. It can be calculated as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}, \tag{2}$$

where N is the number of measurements, x_i are the individual measurements, and $\bar{x}$ is the mean value.

When data are normally distributed, see Fig. 2, the area under the normal distribution between $-\sigma$ and $+\sigma$ covers 68 % of the total area. This means that 68 % of measurements will lie in the 1σ uncertainty interval (i.e., the range of values between $\bar{x} - \sigma$ and $\bar{x} + \sigma$). For the 2σ interval, this percentage increases to 95 %.

Looking back on the data of the falling distance of the ruler, see Fig. 3a, this data also has an uncertainty. Using equation (2) this uncertainty can be calculated, this is shown in Fig 3b, the uncertainty interval is shown in Fig. 3c. Since the uncertainty interval includes 68 % of the measurements, not all measurements lie within the uncertainty interval. Figure 3d also indicates the total range of values, which encompasses all measurements.

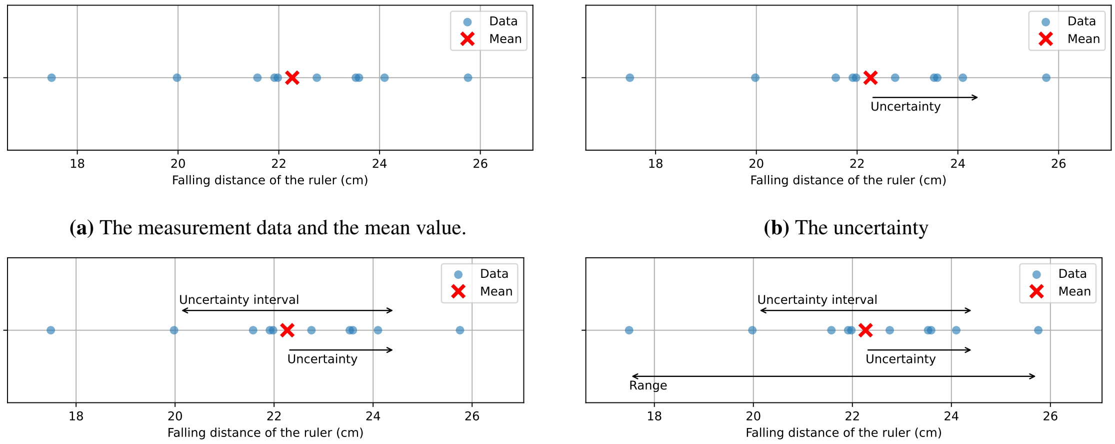
Standard deviation of the mean [★]

The standard deviation of the mean, as the name suggests, gives the uncertainty of the mean value (rather than the uncertainty of the individual measurements).

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}. \tag{3}$$

The standard deviation of the mean is sometimes called the standard error. Due to the negative co-notation of the word error, the use of the term standard deviation of the mean is preferred.

In contrast to the standard deviation, the standard deviation of the mean becomes smaller with an increasing number of measurements. Conceptually, this can be understood because the increasing number of measurements, although they might have large uncertainties, starts to show the underlying distribution more and more clearly. From this distribution, the center (the mean value) can be determined with more certainty.



(a) The measurement data and the mean value.

(b) The uncertainty

(c) The uncertainty and the uncertainty interval

(d) The range, the uncertainty and the uncertainty interval.

Figure 3: Measurement data of the falling distance of a ruler, now with the uncertainty.

Students' ideas about the uncertainty [ ]

Many students will see the uncertainty as something that can be eliminated in an experiment [1, 2, 3]. This heavily ties in with the belief in the existence of a "true value", see Sec. 2.2. Some students believe that, given enough practice, the uncertainty can be reduced to zero, others believe that this can only be done given "real" measurement instruments, again others think this can only be done by "professional" scientists. However, uncertainties will always show up in every experiment. They are *omnipresent* in every experiment and are unequal to zero.

Some students and even textbooks refer to measurement uncertainties as "errors". It is known that the use of this word misleads students into thinking that they have done something wrong [3, 4, 5, 6, 7, 8]. Also, the Guide to the Uncertainty of Measurements (short GUM [9]) explicitly refrains from using the term "error". The word "error", in the context of measurement uncertainties, represents an actual error: a (non-fixable) mistake or something that has gone wrong. Measurement uncertainties, on the other hand, are a part of every scientific experiment and cannot be avoided.

The omnipresence of measurement uncertainties and the notion that all scientists have to deal with this—and the resulting uncertainty—can be a comforting thought for students. It shows them that they are not doing something wrong, but rather engage in authentic scientific practice. This simultaneously develops their view on the nature of science [10, 11, 12].

Although the standard deviation might be too complex or time-consuming in secondary education (alternatives will be discussed later in Sec. 3.1). One common misconception is worth mentioning:

This is the belief that more measurements will lead to a smaller standard deviation [13]. However, this is not the case. More measurements (x_i) will not lead to a smaller value of the standard deviation (σ), see equation (2). Rather, with more measurements one gets a better description of the distribution of the measurements, i.e., the shape of the normal distribution becomes more smooth. But the width of this shape does not change. However, the standard deviation of the mean ($\sigma_{\bar{x}}$) does become smaller. But this quantification is not always known to students or they cannot conceptually distinguish between the two.

References

[1] Coelho, S. M., & Séré, M.-G. (1998). Pupils' Reasoning and Practice during Hands-on Activities in the Measurement Phase. *Research in Science and Technological Education*, *16*(1), 79–96. https://doi.org/10.1080/0263514980160107

[2] Munier, V., Merle, H., & Brehelin, D. (2013). Teaching Scientific Measurement and Uncertainty in Elementary School. *International Journal of Science Education*, *35*(16), 2752–2783. https://doi.org/10.1080/09500693.2011.640360

[3] Pillay, S., Buffler, A., Lubben, F., & Allie, S. (2008). Effectiveness of a GUM-compliant course for teaching measurement in the introductory physics laboratory. *European Journal of Physics*, *29*(3), 647–659. https://doi.org/10.1088/0143-0807/29/3/024

[4] Goedhart, M. J., & Verdonk, A. H. (1991). The development of statistical concepts in a design-oriented laboratory course in scientific measuring. *Journal of Chemical Education*, *68*(12), 1005–1009. https://doi.org/10.1021/ed068p1005
[5] Heinicke, S. (2012). *Aus Fehlern Wird Man Klug: Eine Genetisch-Didaktische Rekonstruktion des Messfehlers*. Logos Verlag Berlin GmbH.
[6] Kampourakis, K., & McCain, K. (2019, December). *Uncertainty: How It Makes Science Advance*. Oxford University Press.
[7] Kirkup, L. (2002). A guide to GUM. *European Journal of Physics*, *23*(5), 483–487. https://doi.org/10.1088/0143-0807/23/5/305
[8] Rollnick, M., Dlamini, B., Lotz, S., & Lubben, F. (2001). Views of South African Chemistry Students in University Bridging Programs on the Reliability of Experimental Data. *Research in Science Education*, *31*(4), 553–573. https://doi.org/10.1023/A:1013102108541
[9] Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement – guide to the expression of uncertainty in measurement* (tech. rep. No. JCGM 100:2008). JCGM. Paris, France.
[10] Heinicke, S., Glomski, J., Priemer, B., & Rieß, F. (2010). Aus Fehlern wird man klug - Über die Relevanz eines adäquaten Verständnisses von "Messfehlern" im Physikunterricht. *Praxis der Naturwissenschaften – Physik in der Schule*, *59*(5), 5–15.
[11] Lederman, N. G. (2007). Nature of Science: Past, Present, and Future. In S. K. Abell, K. Appleton, & D. Hanuscin (Eds.), *Handbook of Research on Science Education* (pp. 831–879). Routledge.
[12] Priemer, B., & Lederman, N. G. (2021). Nature of Scientific Knowledge and Nature of Scientific Inquiry in Physics Lessons. In H. E. Fischer & R. Girwidz (Eds.), *Physics Education* (pp. 113–150). Springer International Publishing. https://doi.org/10.1007/978-3-030-87391-2_5
[13] Séré, M.-G., Journeaux, R., & Larcher, C. (1993). Learning the statistical analysis of measurement errors. *International Journal of Science Education*, *15*(4), 427–438. https://doi.org/10.1080/0950069930150406

2.6 Measurement error (Deviation) [ ]

The word error is sometimes, mistakenly, used to refer to measurement uncertainties. However, errors, in the sense of procedural errors or mistakes, are something that should be avoided. Whereas measurement uncertainties cannot be avoided.

There is another concept that includes the word error and that is in the form of a deviation: the absolute difference between a best estimation and a reference value. To avoid confusion, the word error will only be used in the context of (fixable) mistakes.

Definition [ ]

Error is the difference between a best estimation and a reference value:

$$\text{error} = |\text{best estimation} - \text{reference value}|. \tag{4}$$

If all goes well in an experiment, the reference value should be within the uncertainty interval. The error is said to be small and within uncertainties.

However, there can be **systematic effects** that lead to a shift in the measurements, the best estimation, and the uncertainty interval (but not the uncertainty itself!) with respect to a reference value. For instance, a measurement device can be falsely calibrated, the starting point of a moving object has been displaced, etc. In these cases, *all* measurements have the same shift in their value. When these systematic effects are identified, one can correct for this offset.

The existence of error can also be the starting point to look for causes. Is there a systematic effect, are the uncertainties underestimated in general, is the measurand defined correctly, is the reference value appropriate, etc.?

Students' ideas about error [ ]

As discussed in Sec. 2.7, some students refer to measurement uncertainties as errors, which could lead them to think they have made a mistake [1, 2, 3, 4, 5, 6]. In the physics classroom, it is advised to avoid using the word error in the context of measurement data in general. Instead, one could refer to measurement

$$\left.\begin{array}{l} I = 0.227459\ \text{A} \\ u_I = 0.0050\ \text{A} \end{array}\right\} \quad I = (0.2275 \pm 0.0050)\ \text{A}$$

Figure 4: The best estimation of the current gets rounded according to the uncertainty of the current, i.e., four decimal places.

uncertainties to describe the variance in measurements and to a deviation to indicate a difference between the best estimation and a reference value (which could also be another measurement).

Another prevalent idea is that a small error is always better. Although a small error is an indication of good accuracy, this does not necessarily mean it is a good measurement result. Suppose two groups try to verify the resistance of a $R = 100\,\Omega$ resistor. Group A measures $R_A = (99 \pm 10)\,\Omega$ and group B measures $R_B = (103 \pm 5)\,\Omega$. The error of group A, $e_A = 1\,\Omega$, is smaller than that of group B, $e_B = 3\,\Omega$. However, the measurement uncertainty of group B is half that of group A, indicating a better precision. Since both measurement results are compatible with the reference value (more on the comparison of measurement results in Sec. 4.1), one could argue that the result of group B is better.

References

[1] Goedhart, M. J., & Verdonk, A. H. (1991). The development of statistical concepts in a design-oriented laboratory course in scientific measuring. *Journal of Chemical Education*, *68*(12), 1005–1009. https://doi.org/10.1021/ed068p1005

[2] Heinicke, S. (2012). *Aus Fehlern Wird Man Klug: Eine Genetisch-Didaktische Rekonstruktion des Messfehlers*. Logos Verlag Berlin GmbH.

[3] Kampourakis, K., & McCain, K. (2019, December). *Uncertainty: How It Makes Science Advance*. Oxford University Press.

[4] Kirkup, L. (2002). A guide to GUM. *European Journal of Physics*, *23*(5), 483–487. https://doi.org/10.1088/0143-0807/23/5/305

[5] Pillay, S., Buffler, A., Lubben, F., & Allie, S. (2008). Effectiveness of a GUM-compliant course for teaching measurement in the introductory physics laboratory. *European Journal of Physics*, *29*(3), 647–659. https://doi.org/10.1088/0143-0807/29/3/024

[6] Rollnick, M., Dlamini, B., Lotz, S., & Lubben, F. (2001). Views of South African Chemistry Students in University Bridging Programs on the Reliability of Experimental Data. *Research in Science Education*, *31*(4), 553–573. https://doi.org/10.1023/A:1013102108541

2.7 Measurement Result [ ]

Since measurement uncertainties indicate the quality of an experiment, no measurement result is complete without it. Next, the different ways of denoting the measurement result are shown followed by the rules on the number of significant figures that need to be shown.

Notation of measurement results [ ]

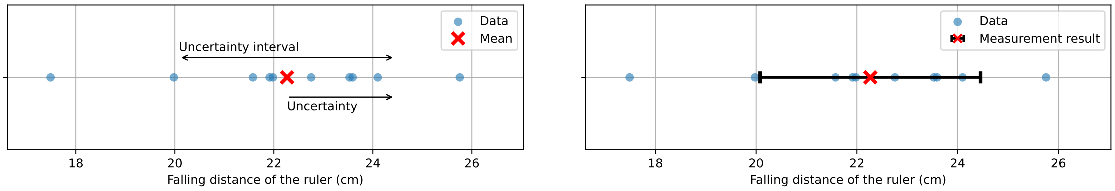
A complete measurement result always shows the best estimation as well as the uncertainty. The most common way to denote this is:

$$\text{measurement result} = \text{best estimation} \pm \text{uncertainty}. \tag{5}$$

The best estimation is most often the mean value, $\bar{x}$, and the uncertainty the standard deviation, σ.

The rounding of the best estimation goes in accordance with the digits of the uncertainty. The best estimation gets as many decimal places as the uncertainty has. For instance, suppose the following current is measured: $I = 0.227\,459\,\text{A}$ with an uncertainty $u_I = 0.0050\,\text{A}$, see Fig. 4. The uncertainty has four decimal places, hence, the best estimation is rounded to four decimal places:

$$I = (0.2275 \pm 0.0050)\,\text{A} \tag{6}$$



(a) The uncertainty and the uncertainty interval

(b) The measurement result with uncertainty bars.

Figure 5: Measurement data of the falling distance of a ruler, like in Fig. 1 now with the complete measurement result.

Table 3: The number of significant figures (# SF) in blue for given quantities.

Quantity	# SF
$l = 17.2\,\mathrm{cm}$	3
$t = 10.02\,\mathrm{s}$	4
$V = 0.0250\,\mathrm{L} = 25.0\,\mathrm{mL}$	3
$\lambda = 640.0\,\mathrm{nm} = 6.400 \times 10^{-7}\,\mathrm{m}$	4
$d = 0.000\,010\,\mathrm{m} = 10 \times 10^{-6}\,\mathrm{m} = 0.010\,\mathrm{mm}$	2

Another way to denote the uncertainty is by placing it in parentheses. This is best illustrated using the example from before. With the parentheses notation, this becomes:

$$I = 0.2275(50)\,\mathrm{A}. \tag{7}$$

This notation is slightly less intuitive for students. The DIN favors this notation for the industry since it sets itself apart from tolerance that has the same notation as (6). Tolerance indicates the maximum deviation between measurements and a reference value that is *allowed* in e.g., production processes.

The uncertainty is an indication of the precision of a measurement result and, hence, an indication of the quality of the experiment. The uncertainty spans an **uncertainty interval** around the best estimation. The uncertainty interval ranges from:

$$\text{uncertainty interval} = [\bar{x} - \sigma; \bar{x} + \sigma]. \tag{8}$$

The uncertainty interval can be seen as the range of values in which the measurand is to be expected (with a certain degree of confidence).

Again looking back on the measurement data of the falling distance of a ruler, see Fig. 5a, this measurement result can be represented graphically. The standard way to plot this is, is by using uncertainty bars (sometimes referred to as errorbars) as shown in Fig. 5b. Note that in this case, the uncertainty is shown for the variable on the$x$-axis, for uncertainties on the y-axis, the uncertainty bars are vertical.

Number of significant figures [📖]

In the example from before, the uncertainty of the current was given in two **significant figures**. The number of significant figures is the number of digits without counting the leading zeros, see Tab. 3. The number of significant figures of the uncertainty determines the number of significant figures of the best estimation, which is rounded in accordance with the uncertainty, see Fig. 4. It should be noted that, the number of significant figures of the best estimation is always larger or equal to the number of significant figures of the uncertainty. There exist three rules for the number of significant figures of the uncertainty:

- The uncertainty is written using one significant figure.
  Examples:
  $u = 3.581\,\mathrm{cm} = 4\,\mathrm{cm}$
  $u = 149\,\mathrm{m} = 1 \times 10^2\,\mathrm{m} = 0.1\,\mathrm{km}$
  $u = 0.005\,01\,\mathrm{A} = 0.005\,\mathrm{A} = 5\,\mathrm{mA}$
  $u = 0.029\,48\,\mathrm{L} = 0.03\,\mathrm{L}$

- The uncertainty is written using two significant figures.
  Examples:

$u = 3.581\,\text{cm} = 3.6\,\text{cm}$
$u = 149\,\text{m} = 1.5 \times 10^2\,\text{m} = 0.15\,\text{km}$
$u = 0.005\,01\,\text{A} = 0.0050\,\text{A} = 50\,\text{mA}$
$u = 0.029\,48\,\text{L} = 0.029\,\text{L}$

- The uncertainty is written using one significant figure unless the first significant figure is a 1 or a 2, in which case two significant figures are indicated.
  Examples:
  $u = 3.581\,\text{cm} = 4\,\text{cm}$
  $u = 149\,\text{m} = 1.5 \times 10^2\,\text{m} = 0.15\,\text{km}$
  $u = 0.005\,01\,\text{A} = 0.005\,\text{A} = 5\,\text{mA}$
  $u = 0.029\,48\,\text{L} = 0.029\,\text{L}$

The GUM [1] does not prefer a specific rule for the number of significant figures of the uncertainty (although they specify to use a maximum of two digits).

For the **rounding** of the uncertainty, there are two options: normal rounding (as was done above) or rounding up. Always rounding up (e.g., $0.005\,01\,\text{A} = 0.0051\,\text{A}$ or even $0.005\,01\,\text{A} = 0.006\,\text{A}$) is a very conservative procedure that can lead to overestimations of the uncertainty. The GUM is somewhat ambivalent here but seems to prefer normal rounding. They do state that it is sometimes appropriate to round uncertainties up using common sense. This would mean to round $u = 0.029\,48\,\text{L}$ up to $0.030\,\text{L}$ but to round $u = 0.005\,01\,\text{A}$ down to $0.0050\,\text{A}$.

Students' ideas about the measurement result [ ]

When asked how one should report a measurement result, many students (even at the university level) indicate that the mean value should be reported [2]. However, for a complete measurement result, this should be complemented by the uncertainty. To prompt students to think about this, one could ask them how their measurement result reflects the quality of their experiment.

References

[1] Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement – guide to the expression of uncertainty in measurement* (tech. rep. No. JCGM 100:2008). JCGM. Paris, France.

[2] Leach, J., Millar, R., Ryder, J., Séré, M.-G., Hammelev, D., Niedderer, H., & Tselfes, V. (1998). *Survey 2: Students' images of science as they relate to labwork learning. Working paper 4, labwork in science education project* (tech. rep. No. Project PL 95-2005). Centre for Studies in Science and Mathematics Education. Leeds.

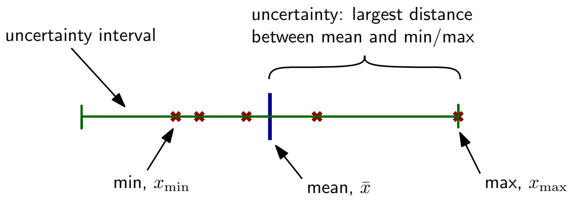


Figure 6: Graphical depiction of how the maximum deviation is calculated. The red crosses are the individual measurements, the blue bar indicates the mean value, and the uncertainty is the largest distance between the mean and either the min or max value of the data set. The green bar indicates the complete uncertainty interval around the mean value.

3 DETERMINING THE UNCERTAINTY

In the previous step, all the basic concepts and terminology were covered. This step is concerned with the quantification of the uncertainty. This is done for repeated measurements, single measurements, and graphical analysis. After that, some rules for uncertainty propagation and the relative uncertainty are covered. Each part starts with the basics followed by didactical considerations.

3.1 Type A [ ]

When measurements are subject to variability, one should take repeated measurements to determine the measurement result (see 2.3). The uncertainty in these instances can be determined by statistical means. This type of uncertainty evaluation is called a **type A** uncertainty evaluation.

Reduction of complexity [ ]

The standard deviation and the standard deviation of the mean (see Sec. 2.7) are the most common quantifications of the uncertainty in scientific practice. Often, students experience difficulties in the process of calculating and interpreting these [1, 2, 3]. To aid them, one could to automate the process of calculating the uncertainty, however, it has been found that students consequently lose a feeling for the credibility of the outcome [3].

Alternatively, one could graphically represent the measurements. This visualizes the variability of the measurements, which reduces the cognitive load on the students [4, 5]. However, this approach lacks a numerical quantification of the uncertainty, something favored by researchers, and something that would also be required for schools [1, 3].

Alternative quantifications [ ]

Another option is to make use of other quantifications that are easier to calculate and conceptually understand. Next, four alternative quantifications will be defined. Research has shown that, when comparing these quantifications with the standard deviation, increasing mathematical complexity also increases the statistical quality [6].

The most simple quantification is the **maximum deviation**. To calculate this, one sorts all measurements from smallest (x_1) to largest (x_N) and calculates the mean value ($\bar{x}$). The biggest difference between the mean value and the smallest or largest value of the series is the uncertainty:

$$u_{\text{min-max}} = \max\left(\bar{x} - x_1, x_N - \bar{x}\right). \tag{9}$$

This measure is graphically shown in Fig. 6. The advantage of this quantification is its mathematical simplicity. Its disadvantage is that it is a big overestimation of the uncertainty (in comparison to the standard deviation) that is highly sensitive to outliers.

To compensate for outliers, this procedure can be adapted to exclude the smallest (x_1) and largest (x_N) value of the series and repeat the procedure for the remaining measurements. This is called the **exclude extremes** quantification:

$$u_{\text{excl.extr.}} = \max\left(\bar{x} - x_2, x_{N-1} - \bar{x}\right). \tag{10}$$

This quantification is almost as simple as the **maximum deviation**, but is less sensitive for outliers. However, with an increasing number of repeated measurements, this quantification becomes equally sensitive for outliers.

For larger sample sizes, one could decide to repeat the procedure for the central half of the measurements, this is called the **middle 50%** measure. To do so, one starts excluding pairs of extreme values one by one until the remaining number of measurements would be less than half the total number of measurements. For instance, for 4–7 measurements, one pair of extremes is excluded, for 8–11 measurements two pairs are excluded, etc. In the form of an equation, this can be written as:

$$u_{\text{mid.50\%}} = \max\left(\bar{x} - x_{1+N/4}, x_{N-N/4} - \bar{x}\right). \tag{11}$$

The downside for this measure is that it might feel unsatisfying for students. They might wonder why repeated measurements were taken at all when half their dataset is disregarded.

The last alternative is the **mean absolute deviation (MAD)**. This is the average of all individual measurements' deviations from the mean:

$$u_{\text{MAD}} = \frac{1}{N}\sum_{i=1}^{N} |x_i - \bar{x}|. \tag{12}$$

Since the calculation of the MAD involves the calculation of a mean value, it is conceptually easy to understand. This uncertainty gives a value that is systematically lower than the standard deviation.

Further reading [💡]

For further information on the different uncertainty quantifications and some examples, see Kok and Priemer [7].

For some ideas for experiments where the analysis of the measurement uncertainty is a necessity for drawing a correct conclusion see: [8, 9, 10, 11, 12, 13]

References

[1] Deardorff, D. L. (2001). *Introductory Physics Students' Treatment of Measurement Uncertainty* [Doctoral Thesis]. North Carolina State University.

[2] Séré, M.-G., Journeaux, R., & Larcher, C. (1993). Learning the statistical analysis of measurement errors. *International Journal of Science Education*, *15*(4), 427–438. https://doi.org/10.1080/0950069930150406

[3] Zangl, H., & Hoermaier, K. (2017). Educational aspects of uncertainty calculation with software tools. *Measurement*, *101*, 257–264. https://doi.org/10.1016/j.measurement.2015.11.005

[4] Kramer, R. S. S., Telfer, C. G. R., & Towler, A. (2017). Visual Comparison of Two Data Sets: Do People Use the Means and the Variability? *Journal of Numerical Cognition*, *3*(1), 97–111. https://doi.org/10.5964/jnc.v3i1.100

[5] Susac, A., Bubic, A., Martinjak, P., Planinic, M., & Palmovic, M. (2017). Graphical representations of data improve student understanding of measurement and uncertainty: An eye-tracking study. *Physical Review Physics Education Research*, *13*(2), 020125. https://doi.org/10.1103/PhysRevPhysEducRes.13.020125

[6] Kok, K., & Priemer, B. (2022). Comparing Different Uncertainty Measures to Quantify Measurement Uncertainties in High School Science Experiments. *International Journal of Physics and Chemistry Education*, *14*(1), 1–9. https://doi.org/10.48550/arXiv.2205.04102
IJPCE doi: 10.51724/ijpce.v14i1.214.

[7] Kok, K., & Priemer, B. (2023a). Messunsicherheiten quantifizieren: Welche Maße gibt es dafür? *MNU Journal*, *76*(4), 330–333. https://doi.org/10.18452/27043

[8] Kok, K., Boczianowski, F., & Priemer, B. (2020). Messdaten im Physikunterricht auswerten – wann sind Messunsicherheiten wichtig? *MNU Journal*, *73*(4), 292–295. https://doi.org/10.18452/27175

[9] Boczianowski, F., & Kok, K. (2020). Modelle empirisch prüfen - Frequenzmessung an stehenden akustischen Wellen mit dem Smartphone. *MNU Journal*, *73*(4), 295–299.

[10] Kok, K., & Boczianowski, F. (2021). Acoustic Standing Waves: A Battle Between Models. *The Physics Teacher*, *59*(3), 181–184. https://doi.org/10.1119/10.0003659

[11] Kok, K., & Priemer, B. (2023b). Using measurement uncertainties to detect incomplete assumptions about theory in an experiment with rolling marbles. *Physics Education*, *58*(3), 035007. https://doi.org/10.1088/1361-6552/acb87b

[12] Musold, W., & Kok, K. (2024). Wie lang ist die Banane? — Über die Relevanz, eine Messgröße zu definieren. In B. Priemer & K. Kok (Eds.), *Messunsicherheiten im Physikunterricht* (1st ed.). Berlin Universities Publishing. https://doi.org/10.14279/depositonce-21608

[13] Wagner, S., Maut, C., & Priemer, B. (2021). Thermal expansion of water in the science lab—advantages and disadvantages of different experimental setups. *Physics Education*, *56*(3), 035022. https://doi.org/10.1088/1361-6552/abeac4

3.2 Type B [★]

When repeated measurements do not show any variance or when only one measurement can be taken, a type A uncertainty evaluation is not appropriate. In the case of repeated readings, it probably means that the measurement instruments are not sensitive enough to measure the variance. This does, however, not mean that there is no measurement uncertainty! In these cases, the uncertainty has to be estimated by other means with the help of a list of sources of uncertainty and their estimated quantified contribution to the uncertainty. This is called a **type B** uncertainty analysis.

Sources of uncertainty [★ 📖]

To make a list of uncertainty contributions, one starts to look at the experiment and searches for sources of uncertainties. These uncertainties can be classified into several categories: the experimental procedure, the environmental conditions, the measurement instruments, mathematical rounding, and the experimenter self [1, 2]. These uncertainty contributions will have to be estimated and added together to form the "final" uncertainty.

The uncertainty by the measurement instrument itself can be divided into three components: gauge uncertainty (usually indicated on the device), linearization uncertainty (how accurate the markings on a scale are or the digitalization of the instrument is), and the scale uncertainty (the finite number of markings on the scale) [3]. For the scale uncertainty rules of thumb exist for digital and analog instruments:

Digital instruments: The uncertainty for a digital instrument is one increment on the scale. For example, see Fig. 7a, suppose a temperature of $T = 19\,°C$ is measured using a digital thermometer. The instrument has increments of 1 °C, thus the scale uncertainty is $u_{\text{scale}} = 1\,°C$. The reason for choosing one entire increment as uncertainty is because one cannot know how the instrument rounds the values (is the value 7.6 °C cut off to 7 °C or correctly rounded as 8 °C). With that, the measurement result becomes $T = (19 \pm 1)\,°C$.

Analog instruments: The uncertainty for an analog instrument is half an increment on the scale. For example, see Fig. 7b, a spring scale measures a force F. The indicator is between 0.2 N and 0.3 N, so the best guess would be 0.25 N. The instrument has markings every 0.1 N, since it is an analog instrument, the uncertainty is $u_{\text{scale}} = 0.05\,\text{N}$. The reason for this, is that one can determine how to round the reading of the analog scale. So the measurement result is $F = (0.25 \pm 0.05)\,\text{N}$.

Simplification [★ 🛠]

Most of the time when a type B uncertainty evaluation is performed, one will make practical considerations. For instance, when measuring the time that students take to run a 100 m stretch, one could include an estimated reaction time of 0.5 s.

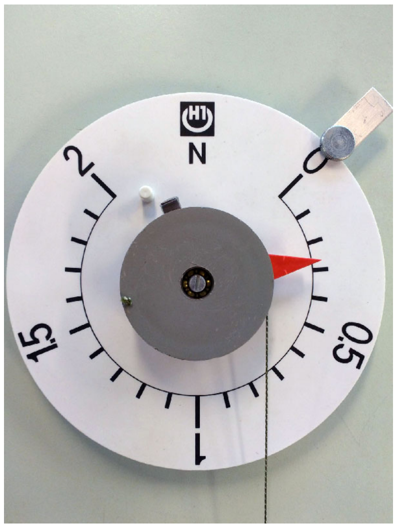
When taking the measurements as shown in Fig. 7, one could decide to only include the scale uncertainty, but not be bothered with gauge and linearization uncertainties (which are much harder to estimate).

Lastly, one could estimate a range of values in which one could plausibly expect the measurement to be. For instance, suppose that, when measuring the length of a classroom with a 1 m board ruler with 1 cm markings, the ruler has to be shifted eight times. Shifting the ruler will result in an additional uncertainty to the 0.5 cm scale uncertainty. One could decide to include a 1 cm uncertainty for each shift. This will lead to a total uncertainty of $8 \cdot 1\,\text{cm} + 0.5\,\text{cm} = 8.5\,\text{cm}$.

Further reading [★ 💡]

For further information and some practical examples on the uncertainty of measurement instruments, see Nagel [3].

(a) A thermometer measuring (19 ± 1) °C.



(b) A spring scale measuring (0.25 ± 0.05) N.

Figure 7: Two measurement instruments.

References

[1] Hellwig, J. (2012). *Messunsicherheiten verstehen : Entwicklung eines normativen Sachstrukturmodells am Beispiel des Unterrichtsfaches Physik* [Doctoral Thesis]. Ruhr-Universität. https://nbn-resolving.org/urn:nbn:de:hbz:294-36561.

[2] Hennes, M. (2007). Konkurrierende Genauigkeitsmaße–Potential und schwächen aus der sicht des anwenders. *Allgemeine Vermessungs-Nachrichten*, *7*, 136–146.

[3] Nagel, C. (2021). Sicher ist sicher! Fachliche Klährung für die didaktische Rekonstruktion von Messunsicherheiten im Unterricht. *Plus Lucis*, (4), 7–11.

3.3 Graphs [★]

Sometimes, determining a quantity is done, not by repeating measurements, but by looking at the dependence between two other variables. In this case, the data is usually evaluated by using a graph and a trendline or fit function.

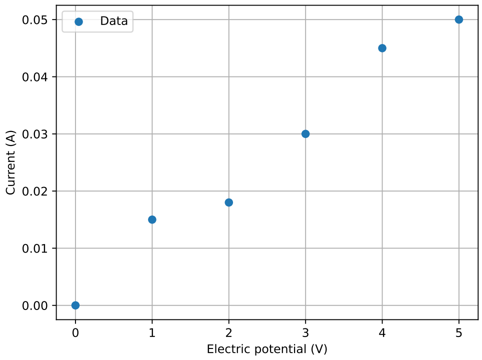
Determining the uncertainty of a fit function [★ 📖]

Suppose one wants to determine the resistance of a resistor R. To do so, one can vary and measure the electric potential U and measure the resulting current I. The different measurements can be plotted in a (U, I)-diagram, see Fig. 8a. A linear **trendline (fit)** of the shape $y = ax + b$ can be fitted to the data, see Fig. 8b. Most computer programs (Excel, Qti-Plot, Google Sheets, . . .) use the method of least squares. The result of this procedure gives the slope a and (when desired) the y-axis offset b.

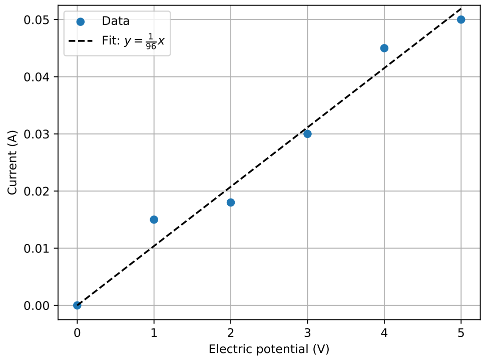
Using Ohm's law ($I = U/R$), one can calculate the resistance using the slope a which is equal to $1/R$. In this case, the resistance is calculated to be 96.3 Ω.

However, this graph does not show whether the fit truly "fits" the data. This can only be determined by looking at the measurement uncertainty, see Fig. 8c. This figure shows that the fit goes through all the uncertainty boxes (the gray shaded areas that span the uncertainty intervalls in the potential and current) and thus is a good fit to the data. At this point, one could say that the resistance of the resistor has a best estimation of $R = 96.3\,\Omega$.

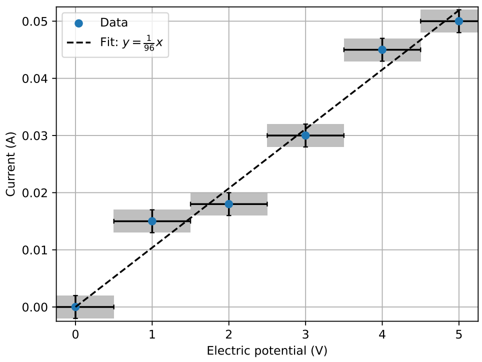
To determine the uncertainty of the fit function, one can refer back to the method of least squares to calculate this. This would yield $u_R = 3.7\,\Omega$, so $R = (96.3 \pm 3.7)\,\Omega$. The exact procedure, however, goes beyond the scope of this unit.



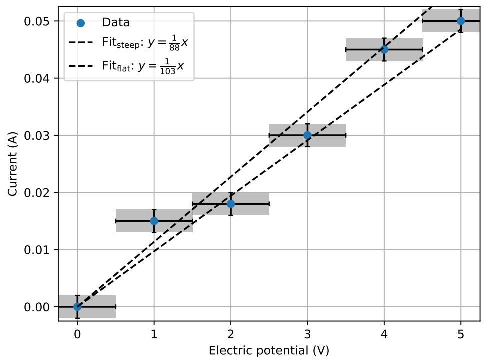
(a) A (U, I)-diagram of some measurement data.



(b) The (U, I)-diagram with a linear fit that is forced through the origin. The fit's slope is equal to $1/R$. No conclusion of the correctness of the linear fit can be drawn.



(c) The same (U, I)-diagram but now with uncertainties (indicated with boxes). This plot shows that the fit is in agreement with all measurements.



(d) The same (U, I)-diagram but now with maximal and minimal slope that go through all uncertainty boxes. These slopes can be used to determine the uncertainty of the slope.

Figure 8: Different plots of measurement data from the electric potential over, and the current going through a resistor.

Alternative graphical uncertainty determination [★ 🪄]

There is a more simplified, graphical, method to determine the uncertainty. This is done by drawing the steepest and the flattest fit function that still goes through all the uncertainty boxes and determine these slopes, see Fig. 8d. This yields a slope with resistances of $R = 88\,\Omega$ and $R = 103\,\Omega$. Using the same process that was used for the maximum deviation (see Eq. (9)), this yields an uncertainty of $u_R = 8\,\Omega$ such that $R = (96 \pm 8)\,\Omega$.

In the latter example, this method clearly overestimates the uncertainty. In other instances, usually when the uncertainties per data point are small, this method will result in too small uncertainties. Sometimes, the method appears to not work at all since there exists no line that goes through all uncertainty boxes. In these cases, it is important to remember that even the method of least squares will not go through all uncertainty boxes but still give a result. One should draw a line that goes through the uncertainty boxes as well as possible. The steepest and flattest slopes should be drawn in the same manner.

The big advantage of this method is that no complex statistical methods are required and it can even be done using paper and pencil.

Fitting a model [★ 💡]

A practical example of a low-cost high school experiment where two model functions are compared with data can be found here: [1]. The two models correspond to two functions and in the experiment and there is a check whether the data fits one of these functions. This fit would indicate which of the two models describes the data best.

References

[1] Kok, K., & Boczianowski, F. (2021). Acoustic Standing Waves: A Battle Between Models. *The Physics Teacher*, *59*(3), 181–184. https://doi.org/10.1119/10.0003659

3.4 Uncertainty Propagation [★]

Sometimes the measurand cannot be measured directly, but rather calculated from other measured quantities or reference values. In these cases, there is an equation that describes how input variables are related to the output variable (i.e., the measurement model). To determine the uncertainty of the output variable, the uncertainties of the input variables need to be propagated correctly.

Rules for propagating uncertainties [★ 📘]

Consider the following example: one wants to determine the average speed of a runner. To do so, a certain distance is measured and then the time the runner takes to run this distance is measured. The equation to calculate the speed is the distance divided by the time: $v = \frac{s}{t}$.

The distance and time each have their own uncertainty, which together result in an uncertainty for the speed. This uncertainty has to be calculated by **propagating** the uncertainty of the distance and time. The rules for propagating uncertainties for some common functions are shown in Tab. 4. In this case, the uncertainty for the speed is given by: $u_v = |v|\sqrt{\left(\frac{u_s}{s}\right)^2 + \left(\frac{u_t}{t}\right)^2}$.

The measurement of the distance is $s = (50.00 \pm 0.50)\,\mathrm{m}$ and a time measurement is $t = (20.0 \pm 1.0)\,\mathrm{s}$. This results in a speed of $v = (2.50 \pm 0.13)\,\mathrm{m/s}$.

Approximations of uncertainty propagation [★ 🪄]

The right column of Tab. 4 also shows some approximations for the determination of the uncertainty (approximations taken from Hellwig p. 275–277 [1]).

Table 4: The rules for the propagation of uncertainties for two independent variables x and y. The table shows how to propagate the uncertainties using Gaussian uncertainty propagation and the approximate procedures. There a and b are numbers without uncertainty, σ is the standard deviation and u is the (approximated) uncertainty.

Function	Gaussian Propagation	Approximation
$z = ax$	$\sigma_z = \lvert a\rvert\sigma_x$	$u_z \approx \lvert a\rvert u_x$
$z = ax \pm by$	$\sigma_z = \sqrt{a^2\sigma_x^2 + b^2\sigma_y^2}$	$u_z \approx au_x + bu_y$
$z = xy, z = \frac{x}{y}$	$\sigma_z = \lvert z\rvert\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$	$u_z \approx \lvert z\rvert\left(\frac{u_x}{\lvert x\rvert} + \frac{u_y}{\lvert y\rvert}\right)$
$z = ax^b$	$\sigma_z = \left\lvert z\frac{b\sigma_x}{x}\right\rvert$	$u_z \approx \left\lvert z\frac{bu_x}{x}\right\rvert$
$z = a\log_n(bx)$	$\sigma_z = \left\lvert a\frac{\sigma_x}{x\ln(n)}\right\rvert$	—
$z = a^{bx}$	$\sigma_z = \lvert zb\ln(a)\sigma_x\rvert$	—
$z = a\sin(bx)$	$\sigma_z = \lvert ab\cos(bx)\sigma_x\rvert$	—
$z = a\cos(bx)$	$\sigma_z = \lvert ab\sin(bx)\sigma_x\rvert$	—
$z = a\tan(bx)$	$\sigma_z = \lvert ab\sec^2(bx)\sigma_x\rvert$	—

For the speed measurement from before, the uncertainty can be calculated as follows:

$$\begin{aligned} u_z &= |z|\left(\frac{u_x}{|x|} + \frac{u_y}{|y|}\right) \\ &= |v|\left(\frac{u_s}{|s|} + \frac{u_t}{|t|}\right) \\ &= 2.50\,\mathrm{m/s}\left(\frac{0.50\,\mathrm{m}}{50.00\,\mathrm{m}} + \frac{1.0\,\mathrm{s}}{20.0\,\mathrm{s}}\right) = 0.15\,\mathrm{m/s}. \end{aligned}$$

Resulting in a measurement result of $v = (2.50 \pm 0.15)\,\mathrm{m/s}$. Which is a slight overestimation of the uncertainty.

For some functions, no standard approximations exist. There is, however, an even easier procedure to approximate the uncertainty propagation that can always be used. In these cases, the uncertainty of the calculated quantity is determined by combining the values of the quantities and their respective uncertainties to calculate the smallest and largest possible outcome, this range is then used as the uncertainty (see Hellwig p.277–278 [1]).

For the speed measurement, this would be the following procedure. The speed is calculated as: $v = \frac{s}{t} = \frac{50.00\,\mathrm{m}}{20.0\,\mathrm{s}} = 2.50\,\mathrm{m/s}$.

The largest possible value would be obtained by combining the measurements and their uncertainties as follows:

$$\begin{aligned} v &= \frac{s + u_s}{t - u_t} \\ &= \frac{50.5\,\mathrm{m}}{19\,\mathrm{s}} = 2.66\,\mathrm{m/s}. \end{aligned}$$

Table 5: All the quantities, their value, uncertainty, and relative uncertainty for a speed measurement.

Quantity	Value	Uncertainty	Relative uncertainty
input:			
Distance	50.00 m	0.50 m	1 %
Time	20.0 s	1.0 s	5 %
output:			
Speed	2.50 m/s	0.13 m/s	5.2 %

For the smallest possible value:

$$v = \frac{s - u_s}{t + u_t} = \frac{49.5\,\mathrm{m}}{21\,\mathrm{s}} = 2.36\,\mathrm{m/s}.$$

The uncertainty is now determined with the procedure of the maximum deviation (see Eq. (9)): the two deviations are: $|2.50\,\mathrm{m/s} - 2.66\,\mathrm{m/s}| = 0.16\,\mathrm{m/s}$ and $|2.50\,\mathrm{m/s} - 2.36\,\mathrm{m/s}| = 0.14\,\mathrm{m/s}$. The largest of the two deviations is, conservatively, chosen as the uncertainty. The result is now $v = (2.50 \pm 0.16)\,\mathrm{m/s}$, which is slightly larger than the approximation from before.

Online uncertainty calculator [★ 💡]

Although there are indications that calculating the uncertainty by hand is advantageous for students [2]. There are instances where this calculational routine takes away the focus of the data analysis and, hence, the interpretation of the result [3]. Many uncertainty calculators can be found online, one that works quite intuitively can be found here: https://nicoco007.github.io/Propagation-of-Uncertainty-Calculator.

References

[1] Hellwig, J. (2012). *Messunsicherheiten verstehen : Entwicklung eines normativen Sachstrukturmodells am Beispiel des Unterrichtsfaches Physik* [Doctoral Thesis]. Ruhr-Universität. https://nbn-resolving.org/urn:nbn:de:hbz:294-36561.

[2] Zangl, H., & Hoermaier, K. (2017). Educational aspects of uncertainty calculation with software tools. *Measurement*, *101*, 257–264. https://doi.org/10.1016/j.measurement.2015.11.005

[3] Séré, M.-G., Journeaux, R., & Larcher, C. (1993). Learning the statistical analysis of measurement errors. *International Journal of Science Education*, *15*(4), 427–438. https://doi.org/10.1080/0950069930150406

3.5 Relative uncertainty [★]

Sometimes one wants to improve the precision of the measurement. In the example of the speed measurement from before, the measurement uncertainty of the distance and the time measurement influence the uncertainty of the measurement. To find out which change will mostly affect the uncertainty of the speed the **relative uncertainty** or percentage uncertainty can be calculated, see Tab. 5. This can be calculated by first making an overview of all input and output quantities, their value, and their uncertainty from which the percentage uncertainty can be calculated.

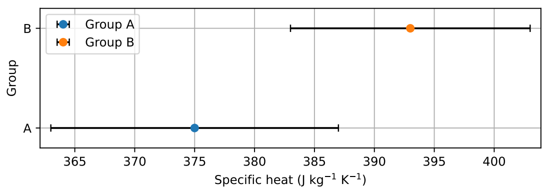
By looking at the relative uncertainties and how they propagate, one can quickly identify the quantity that has the biggest impact on the uncertainty of the calculated quantity. In this case, decreasing the uncertainty of the time measurement (5 % relative uncertainty) will decrease the uncertainty of the speed much more than measuring the distance to a higher precision (1 % relative uncertainty).

Further reading [★ 💡]

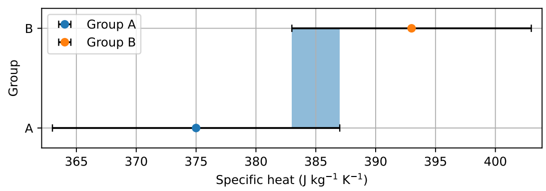
It can sometimes be useful to estimate the uncertainties before an experiment. Setting this goal for the uncertainty will affect the experimental procedure: [1].

References

[1] Priemer, B. (2023). Wie präzise soll's denn sein – Eine einfache Abschätzung von Messunsicherheiten vor einem Experiment. *MNU Journal*, *76*(4), 320–323.



(a) The results of groups A and B with their respective uncertainty bars.



(b) The overlap region of the uncertainty intervals is indicated with blue.

Figure 9: The results of the specific heat measured by groups A and B.

4 DATA COMPARISON

Thus far, the focus has been on the sources of measurement uncertainties, how they affect the data, how they can be quantified, and how they can be propagated. This part will be concerned with *using* the measurement uncertainty.

Measurement uncertainties are not just a burden to be calculated at the end of an experiment—that is, they should not be! Measurement uncertainties are useful because they allow one to compare measurement results. Only by coincidence, two mean values of two datasets are exactly the same. This, however, does not mean that two results are only then compatible with one another. For a correct comparison, the uncertainty intervals have to be compared.

This section is concerned with some practical rules for comparing measurement results in general and how students reason in these comparisons.

4.1 Simple Comparison [📘 ✎]

Comparing two measurement results is a common practice in science. This comparison cannot be done without taking the measurement uncertainty into account.

One way to compare two mean values in science is by calculating a t-test. The mathematical procedure of how this test is calculated and how the significance of this test is determined is beyond the scope of this unit or the high school level. But what the test ultimately does is to compare the two mean values, take their respective variance as well as the number of repeated measurements into account, and determine the degree of overlap between them.

Comparing uncertainty intervals [✎]

To compare datasets or measurement results in a high school setting, a t-test is probably over-complicated. Hence, one has to simplify the method of comparison. This can be done by comparing uncertainty intervals (see Eq. (8) in Sec. 2.7). One rule of thumb is that: overlapping uncertainty intervals indicate compatible results.

Suppose two groups have measured the specific heat of two similar-looking metal objects. Group A has measured $c_A = (375 \pm 12)\,\mathrm{J\,kg^{-1}\,K^{-1}}$ and group B $c_B = (393 \pm 10)\,\mathrm{J\,kg^{-1}\,K^{-1}}$, see Fig. 9a. The groups now ask themselves, could these objects be made of the same metal, i.e., can the metal in the objects have the same specific heat? The uncertainty interval indicates the range in which the measurand can be expected. Hence, the overlapping in uncertainty intervals indicate a shared range of possible specific heat values that can be associated with both objects, see Fig. 9b. Thus, the results are compatible and they have to acknowledge the possibility that the specific heat of both objects is the same.

This does not mean that they have proven that the specific heat is the same! They have one piece of evidence that it could be the same. Depending on the degree of overlap and the range of the uncertainty interval, this possibility can be evaluated.

Students' ideas about data comparison [✎]

When asked about the compatibility between a measurement result and a reference value, some students will refer to a (percentage) difference. Although this is an aspect of the quality of a result, it is not an indication of compatibility. This can be illustrated to students with the results in Tab. 6.

Table 6: Experimental results in determining the gravitational acceleration. Measurement result A, although having the smallest uncertainty, is not compatible with the reference value.

Reference value:	
g_{ref}	$9.81\,\mathrm{m/s^2}$
Experimental results:	
g_{A}	$(9.83 \pm 0.01)\,\mathrm{m/s^2}$
g_{B}	$(9.78 \pm 0.05)\,\mathrm{m/s^2}$
g_{C}	$(9.83 \pm 0.15)\,\mathrm{m/s^2}$

Although groups A and C both have a difference of $0.02\,\mathrm{m/s^2}$ with the reference value, group A's result is not compatible with the reference value. Groups B and C both are compatible with the reference value. And, even though the difference with the reference value of group B is larger than that of group C, its uncertainty is three times smaller.

Further reading [💡]

For a didactical simplification of the *t*-test see: [1].

An online *t*-test calculator can be found here: https://www.graphpad.com/quickcalcs/ttest1.cfm.

Some more explanation about the statistics of the *t*-tests can be found on Wikipedia: https://en.wikipedia.org/wiki/Student's_t-test.

References

[1] Neumann, S. (2021). Bin ich wirklich schneller als mein Sitznachbar? *Plus Lucis*, (4), 36–38 subtitle: Der Nutzen von Streuungsmaßen bei der Auswertung von Experimenten.

4.2 Indicators [📖]

The task of comparing measurement results is a good probe to gauge students' understanding of measurement uncertainties. These data comparison problems should not be reduced to a multiple choice in agreement/not in agreement question. Rather, this should be supported by a justification.

Looking at students' justifications in a data comparison problem, these usually consist of two components: a certain **quantity** that is compared, followed by the check of a certain **criterion** that is met (or not). Looking at the quantity and criterion that students mention in their justification, gives insight into students' conceptual understanding [1].

The following quantities have been identified for students comparing datasets:

uncertainty interval: students compare the whole uncertainty interval either as a range or as a mean ± uncertainty.

complete dataset: students compare the dataset as a whole.

deviations: students subtract two datasets and compare the resulting differences.

mean value: students compare the mean value.

pairwise: students compare the results one pair at a time.

uncertainty: students compare the uncertainty only.

single measurements: students compare single, isolated measurements (e.g., a recurring value, an extreme value).

The following criteria have been identified:

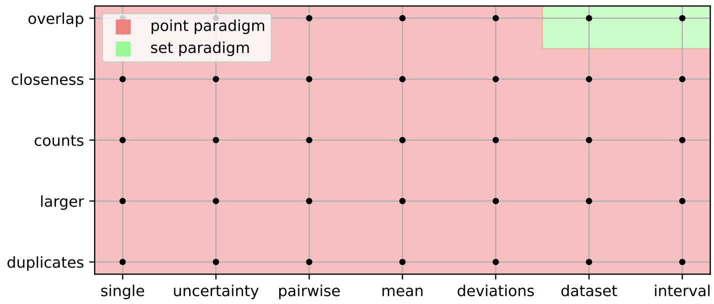


Figure 10: The combination of the quantity that is being compared and the criterion that is checked in a justification can be associated with the point and set paradigms.

overlap: students look for overlap (or the absence thereof) in their datasets or measurement results.

closeness: students look at how close or similar two datasets or measurement results are.

counts: students count the occurrence of certain events e.g., one value being larger/smaller than the other.

larger/smaller: students compare quantities and check whether one is larger or smaller than the other.

duplicates: students look for duplicate or recurring values.

The combination of this compared quantity and checked criterion can be associated with the point and set paradigms, see Fig. 10. If the combination is in a red area, the justification is associated with the point paradigm, if it lands in a green area, the justification can be associated with the set paradigm (for point and set paradigms see, Sec. 2.1).

Typical student responses [ ]

Suppose two groups A and B measure the falling time of an object. The data is given in Tab. 7. Based on the data, students are asked whether the falling times are the same.

One typical student response could be:

> *"The longest falling time* 1.583 s *occurs in group A. Therefore, the falling time of group A's object must be larger."*

This person looks for the largest (larger/smaller criterion) value in the series (single value comparison) and concludes that the group to which this time belongs must have the longest falling time. This person has looked at a single value, and the criterion is that this is larger than other values in the series. This single measurement determines the conclusion of the experiment. This is strongly associated with the point paradigm in which single isolated values are representative of a whole series.

To support these learners, first, they could be asked to consider why a series of measurements is taken and what causes them to fluctuate. Thinking along these lines, learners should start to realize, that the sources for measurement uncertainties affect the results of both groups. Then these learners can be helped by letting them look at the spread of the whole series of measurements.

Another typical student response could be:

> *"The falling time of group A's object is longer because most of the time the falling time of group A is longer than that of group B."*

This student has compared the measurements pairwise (pairwise comparison) one at a time and has looked at which of the two values is larger or smaller than the other (counts criterion). This is more sophisticated than the previous response since all values are considered. However, the isolated comparison of measurements, one at a time is also a clear example of point paradigm reasoning.

Table 7: Two datasets that can be compared for compatibility. The bottom row indicates the mean value.

Falling times group A (s)	Falling times group B (s)
1.530	1.548
1.573	1.534
1.522	1.520
1.548	1.571
1.583	1.523
1.538	1.526
1.549	1.537

There are several strategies to help these students reflect their reasoning. One could re-arrange the measurements and ask the students what their conclusion is. Students will quickly realize that, with their current reasoning, their conclusion should change. Oftentimes, this is enough for students to recognize their flawed reasoning. Students should then be guided to look at the measurement series as a whole. One helpful strategy is to have them plot their results in a diagram. This will help them see the scatter of the dataset as a whole.

Another somewhat more sophisticated response could be:

> *"The two mean values,* 1.549 s *and* 1.537 s *are very close, and therefore the two times can be considered the same."*

From this response, it becomes clear that this person has compared mean values (mean comparison) and that the criterion for agreement is that the values are close enough (closeness criterion). Although the comparison of the mean value could be considered set paradigm reasoning, students often do so as an automated routine. The mean value itself is then the only value that is considered, rendering it point paradigm reasoning. The closeness of the two values is also considered point paradigm reasoning. This is because, for the evaluation of this closeness to be relevant or not, the measurement uncertainty has to be considered.

To help these students, guiding questions could be considered. For instance to ask them about what they mean by "close", and how close is close enough for them. As long as the sum of the uncertainties is smaller than the difference between mean values, the results can still be considered incompatible.

Furthermore, it would be good to ask what role the mean value plays for the students. Is it, an estimation of centrality for the measurement series (set paradigm reasoning) or do students consider it "the final outcome" of the experiment (point paradigm reasoning)? If the latter is the case, one could follow up by asking what would happen to the mean value if another measurement were to be added to the series. A changing "true value" is troublesome from a point paradigm perspective, but a changing best estimation within the uncertainty interval is no issue from a set paradigm perspective.

References

[1] Kok, K., & Priemer, B. (2023). Assessment tool to understand how students justify their decisions in data comparison problems. *Physical Review Physics Education Research*, *19*(2), 020141. https://doi.org/10.1103/PhysRevPhysEducRes.19.020141

4.3 Comparing Multiple Results [★]

There are cases where more than two datasets or measurement results are compared. Two cases are described here, comparing the results of more than two measurement results and analyzing the relation between two quantities by means of graphs (see also Sec. 3.3).

Comparing three measurement results [★ 📖]

In Sec. 4.1 the process of comparing two datasets or measurement results was done. One can also use this same procedure to compare the results of three measurement results. However, one has to be very careful in interpreting the outcomes.

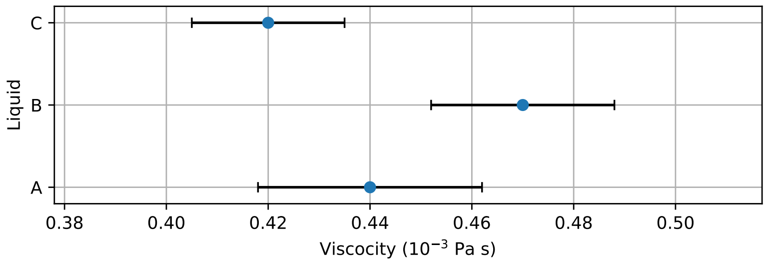
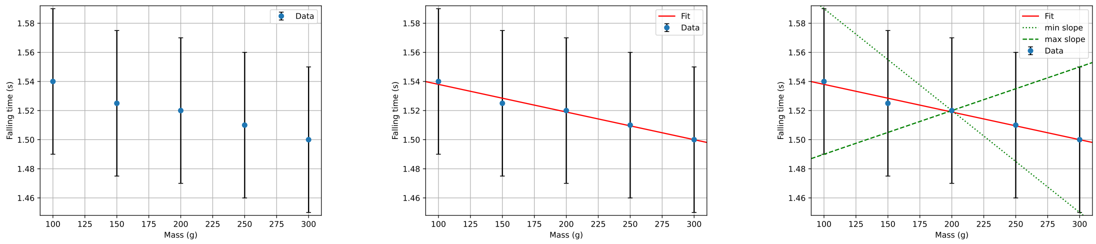


Figure 11: The measurement results for the viscosity of three liquids.



(a) Five time measurements of the falling time of balls with different masses.

(b) The red line indicates a linear fit to the data.

(c) The green dashed lines indicate the minimum and maximum slope of the fit.

Figure 12: The results of an experiment in which balls of different masses are dropped from a certain height.

Suppose that the viscosity of three different liquids A, B, and C is measured, see Fig. 11. When comparing the viscosity of the liquids, one has to conclude that the viscosity of liquid A is compatible with C as well as with B. However, one cannot conclude that the result of liquid B is compatible with liquid C—although they are both compatible with liquid A.

Moreover, based on the results, one has to conclude that liquid C has a lower viscosity than liquid B. For liquid A, the conclusion is limited to saying that its viscosity is compatible with that of both B and C. If one wants to make a more elaborate statement about the viscosity of liquid A, the uncertainties will have to be reduced.

Analyzing fit functions [★ 📕]

In cases where two quantities depend on one another, multiple measurements are taken. In the data analysis, one evaluates whether the two quantities fit a certain model or function. This data can be evaluated in tabular form, but students find it easier to interpret the results in graphical form [1, 2]. In Sec. 3.3 the procedure for drawing fit functions and their uncertainties were described. This section looks into the interpretation of these fit functions.

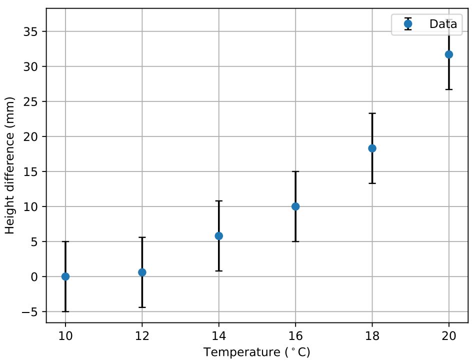
To answer the question of whether heavier objects fall faster than lighter one could do an experiment. Figure 12a shows the data of the falling time of balls of different masses. The question now arises whether the mass influences the falling time or not.

At first glance, the data seems to be showing a decrease in falling time for increasing masses. This is also indicated by the fit function, see Fig. 12b. However, when looking at the maximum and minimum slope, see Fig. 12c, one can clearly see that both positive as well as negative slopes fit the data. The conclusion is that due to the possibility of positive and negative slopes, no influence of mass on the falling time can be assumed. Note that this conclusion could only be drawn after evaluation of the uncertainty.

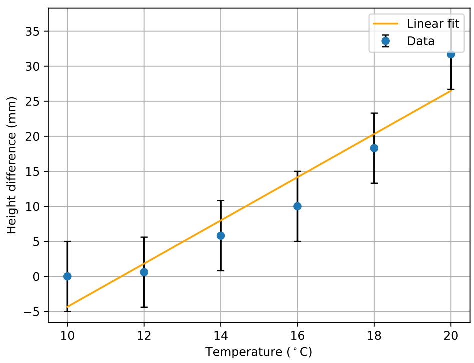
If one aims to show the influence of friction on smaller masses, which would result in longer falling times for smaller masses, the uncertainties would have to be drastically reduced.

Suppose one is interested in the thermal expansion of water and wants to know the relation between the volume change and the temperature. To investigate this, one can perform an experiment in which the height of a column of water is measured for different temperatures, see Fig. 13a. The starting height is placed at 0 mm for a starting temperature of 10 °C and the temperature is slowly increased in steps of 2 °C to a temperature of 20 °C, see Fig. 13b.

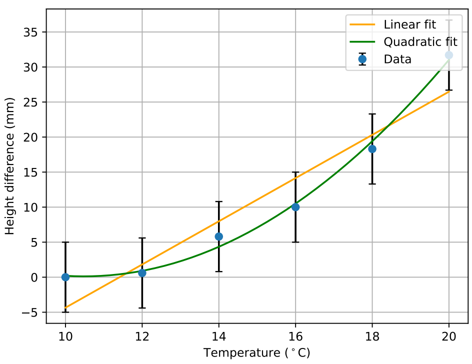
(a) Measuring the height of the column of water for different temperatures.



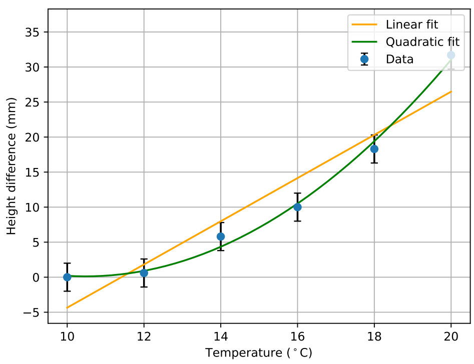
(b) Sample data for the expansion of water.



(c) A linear fit to the data.



(d) A quadratic fit to the data.



(e) More precise measurement data.

Figure 13: Some experimental data for the expansion of a volume of water for different temperatures.

As a first step, one starts to fit a linear fit function, see Fig. 13c. This fit function goes through all uncertainty bars and, therefore, fits the data. Although a quadratic function fits the data better, see Fig. 13d, a linear dependence cannot be excluded. When refining the experiment and decreasing the uncertainties, see Fig. 13e, one is finally able to falsify a linear dependence and make a quadratic dependence more likely.

Further reading [💡]

A more detailed description and analysis of the thermal expansion of water is found here: [3]

References

[1] Kramer, R. S. S., Telfer, C. G. R., & Towler, A. (2017). Visual Comparison of Two Data Sets: Do People Use the Means and the Variability? *Journal of Numerical Cognition*, *3*(1), 97–111. https://doi.org/10.5964/jnc.v3i1.100

[2] Susac, A., Bubic, A., Martinjak, P., Planinic, M., & Palmovic, M. (2017). Graphical representations of data improve student understanding of measurement and uncertainty: An eye-tracking study. *Physical Review Physics Education Research*, *13*(2), 020125. https://doi.org/10.1103/PhysRevPhysEducRes.13.020125

[3] Wagner, S., Maut, C., & Priemer, B. (2021). Thermal expansion of water in the science lab—advantages and disadvantages of different experimental setups. *Physics Education*, *56*(3), 035022. https://doi.org/10.1088/1361-6552/abeac4