Handbuch Messunsicherheiten

Karel Kok*, Burkhard Priemer
Humboldt-Universität zu Berlin

25. Oktober 2024

Inhaltsverzeichnis

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

1 EINFÜHRUNG

Willkommen bei diesem digitalen Handbuch zum Thema Messunsicherheiten. Dies ist das Lehrkräftehandbuch, das sowohl den konzeptionellen als auch den didaktischen Hintergrund der digitalen Lernumgebung (DLE) für Lernende vermittelt, welche Sie hier finden: https://lernen.physik.hu-berlin.de/measurementuncertainties. Nach dem Durcharbeiten dieses Handbuchs werden Sie:

- die zugrunde liegenden Konzepte von Messunsicherheiten kennen;
- einige Alltagsvorstellungen von Messunsicherheiten kennen;
- Stragegien zur Adressierung dieser Vorstellungen kennen;
- über einige didaktische Strategien zur Einführung in das Thema Messunsicherheiten verfügen;
- einige praktische Beispiele zum Umgang mit Messunsicherheiten kennen;
- einige Konzepte gesehen haben, die über die DLE hinausgehen.

Dieses Handbuch ist konzeptionell ausgerichtet und konzentriert sich nicht auf Zahlen und Definitionen, sondern auf die zugrunde liegenden Prinzipien, die Bedeutung und die Interpretation der Konzepte. Die Konzepte werden anhand von Beispielen und Daten aus der Praxis vorgestellt.

Das Handbuch besteht aus drei Teilen, die in mehrere Schritte unterteilt sind. Jeder Schritt besteht aus mehreren thematischen Blöcken. Diese Blöcke sind wiederum mit Symbolen beschriftet, wobei diese Symbole angeben, welche Art von Inhalt in dem jeweiligen Block behandelt wird:

- Fachinhaltliche Grundlagen. Dies ist der Bereich, in dem Definitionen mit ihrer Bedeutung und Interpretation ergänzt werden.
- Didaktiktische Grundlagen. Hier finden Sie Sichtenweisen der Lernenden, Ideen für Unterrichtsstrategien und Ideen zur didaktischen Reduktion, die Ihnen helfen werden, das Thema Messunsicherheiten in Ihrem Unterricht zu behandeln.
- Vertiefung und Beispiele. Hier finden Sie Links zu praktischen Beispielen oder weiterführenden Informationen, die an die Inhalte des Schritts anknüpfen.
- Fortgeschrittene Konzepte. Dies sind einige optionale Konzepte, die über die Konzepte der DLE der Lernenden hinausgehen, aber für bestimmte Experimente essenziell sein können. Fortgeschrittene Konzepte sind ebenfalls mit einem oder gekennzeichnet, um zu zeigen, ob es sich um die Grundlagen oder den didaktischen Teil handelt. Da diese Konzepte in der DLE der Lernenden nicht behandelt werden, müssen Sie diese Konzepte selbst einführen, wenn Sie möchten, dass Ihre SchülerInnen mit diesen Konzepten arbeiten können!

Im ersten Teil werden die grundlegenden Konzepte zu Messunsicherheiten vorgestellt. Sie lernen die fachinhaltlichen Konzepte sowie die didaktischen Ansichten und Alltagsvorstellungen der Lernenden zu diesen Konzepten kennen.

Im zweiten Teil wird die Bestimmung der Messunsicherheit diskutiert. Hier lernen Sie die Berechnung der Messunsicherheit, die grafische Darstellung, die Fortpflanzung und die Analyse der Unsicherheitsbilanz kennen. Ergänzt wird diese Thematik durch eine didaktische Analyse, in der Sie einige didaktische Strategien und Ideen zur didaktischen Reduktion kennen lernen werden.

Der letzte Teil beschäftigt sich mit der Anwendung von Messunsicherheiten im Sinne eines Datenvergleichs. Dies beinhaltet sowohl die formalen Regeln für den Datenvergleich als auch ein didaktisches Werkzeug, um die Sichtweise der Lernenden auf Messunsicherheiten in Datenvergleichsproblemen zu ermitteln.

Tabelle 1: Einige kontrastierende Charakteristika des Point- und des Set-Paradigmas.

Point-Paradigma	**Set-Paradigma**
Jedes Experiment hat einen “wahren Wert”, der mit dem richtigen Verfahren und den richtigen Messinstrumenten bestimmt werden kann.	In einem Experiment kann eine bestimmte Messgröße bestimmt werden. Diese Messgröße ist eine möglichst genaue Beschreibung dessen, was gemessen werden soll.
Unsicherheiten können beseitigt, vermieden oder auf Null reduziert werden. Sie sind ein Anzeichen dafür, dass etwas schief gelaufen ist.	Unsicherheiten sind in jedem Experiment allgegenwärtig. Ziel ist es, sie zu kontrollieren, sie auf ein gewünschtes Maß zu reduzieren und sie zu quantifizieren. Sie sind ein Indikator für die Qualität eines Experiments.
Wiederholte Messungen werden durchgeführt, um ein Ergebnis zu bestätigen, das Messen zu üben oder einen Mittelwert berechnen zu können.	Wiederholte Messungen werden durchgeführt, um die Streuung der Messungen abzuschätzen und die Messunsicherheit zu quantifizieren.
Der Mittelwert stellt den “wahren Wert” eines Experiments dar	Jedes Messergebnis - unabhängig davon, ob es aus einer einzigen oder aus mehreren Messungen stammt - enthält einen Bestwert (oft den Mittelwert) und ein Unsicherheitsintervall (den Wertebereich) um diesen Wert herum.

2 KONZEPTE

2.1 Point und Set Paradigmen [ ]

Die Forschung über das Verständnis von Lernenden bezüglich Messunsicherheiten hat zwei Paradigmen identifiziert, siehe Tab. 1: das Point- und Set-Paradigma [1]. Diese Paradigmen beschreiben zwei Extreme in der Sichtweise der Lernenden im Bezug auf Messdaten. Dieser Teil beginnt mit einer Beschreibung der beiden Paradigmen, die dann als Grundlage verwendet wird, um auf Schwierigkeiten der Lernenden im Bezug auf Messunsicherheiten zu fokussieren.

Point Paradigma [ ]

Im Point-Paradigma betrachten die Lernenden Messungen als einzelne isolierte Ereignisse, die mit den richtigen Instrumenten und der richtigen Vorgehensweise zu dem “wahren Wert” führen, der in einem Experiment gemessen werden kann. Messunsicherheiten sind das Ergebnis von Unzulänglichkeiten in den Messinstrumenten und/oder -verfahren, die im Prinzip beseitigt werden können.

Set Paradigma [ ]

Im Gegensatz dazu betrachten die Lernenden beim Set-Paradigma einen Datensatz als Ganzes. Die Schwankungen in den Messungen werden zur Bestimmung der Messunsicherheit herangezogen und sind somit ein Hinweis auf die Datenqualität. Die Lernenden lassen die Vorstellung von einem wahren Wert hinter sich und versuchen stattdessen, das Unsicherheitsintervall zu bestimmen, das als eine Spanne von Werten angesehen werden kann. Das Ergebnis einer Messung kann also niemals ein einzelner Wert sein, sondern ist immer ein Wertebereich.

Ziel [ ]

Die Forschung hat wiederholt gezeigt, dass die überwiegende Mehrheit der Lernenden, auch auf Hochschulebene, im Point-Paradigma verortet ist. Ziel vieler Laborkurse für Studienanfänger ist es, das Denken der Lernenden vom Point- zum Set-Paradigma zu bringen. Die Forschung hat auch gezeigt, dass dies am besten gelingt, wenn zunächst die Grundlagen der Messunsicherheit behandelt werden [2, 3], bevor man sich statistischen Verfahren zuwendet. Dieser Ansatz wird auch in der Digitalen Lernumgebung (DLE) verfolgt, die für Lernende der Sekundarstufe I und II entwickelt wurde.

Die Tabelle 1 zeigt einige Vorstellungen, die mit dem Point-Paradigma und dem Set-Paradigma der Lernenden verbunden sind. Die Konzepte in dieser Tabelle werden in den folgenden Unterabschnitten weiter ausgeführt.

Relevanz der Messunsicherheit [ ]

Abhängig von der Zielsetzung des Experiments ist die Abschätzung der Messunsicherheit eine Notwendigkeit für den Erfolg des Experiments. Messunsicherheiten müssen immer dann berücksichtigt werden, wenn mit den Daten der Messungen Schlussfolgerungen gezogen werden. In anderen Fällen ist diese Abschätzung nicht erforderlich, z.B. wenn ein Zusammenhang nur illustriert werden soll. Für weitere Informationen siehe Kok et al. [4].

Für Lehrstrategien zur Messunsicherheit siehe Holz und Heinicke.
https://www.youtube.com/watch?v=uMvT02mHkss

Literatur

[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 Wahrer Wert [ ]

Angenommen, man interessiert sich für die Fallzeit eines Objekts, das aus einer Höhe von 1 m fallen gelassen wird. Diese Zeit lässt sich leicht mit der Gleichung: $h = \frac{1}{2}gt^2 \rightarrow t = \sqrt{\frac{2h}{g}} = \sqrt{\frac{2 \cdot 1\,\mathrm{m}}{9{,}81\,\mathrm{m\,s^{-2}}}} = 0{,}451\,523\,6\ldots\,\mathrm{s}$ berechnen.

Im Prinzip kann diese Fallzeit genau berechnet werden. Bei der experimentellen Überprüfung dieser exakten Zahl treten jedoch Probleme auf. Die menschliche Reaktionszeit beeinflusst die Messungen, und es gibt keine perfekten Messinstrumente. Man kann sich fragen, ob das theoretische Ergebnis experimentell bestätigt werden kann.

Messen mit absoluter Präzision [ ]

Um diese Fallzeit zu verifizieren, muss ein Experiment durchgeführt werden. Der Aufbau des Experiments muss so sein, dass das Objekt 1 m fallen kann. Das ist leichter gesagt als getan. Mit einem Maßband kann die Höhe nur bis zu einer Genauigkeit von 1 cm gemessen werden. Das bedeutet, dass man nur sicher sein kann, dass die Höhe zwischen 1,005 m und 0,995 m liegt. Man kann dieses Intervall verringern, indem man ein präziseres Messinstrument verwendet, aber die Genauigkeit wird immer eine Grenze haben, die durch die Auflösung des Instruments bedingt ist. Außerdem setzt dies voraus, dass das Messinstrument exakt kalibriert wurde. Dies kann natürlich niemals der Fall sein, da man dazu ein anderes perfektes Messgerät zum Vergleich heranziehen müsste.

Nicht nur das Messgerät, sondern auch die Umgebung beeinflusst die Messung. Angenommen, in der Versuchsanordnung wird ein Ball von einem Tisch fallen gelassen. Eine Temperaturänderung führt dazu, dass sich die Stahlbeine des Tisches ausdehnen oder zusammenziehen. Für alltägliche Messungen ist dieser Effekt vernachlässigbar klein, aber um absolute Präzision zu erreichen, muss dieser Effekt berücksichtigt werden.

Letztendlich kann man nie hundertprozentig sicher sein, dass der Versuchsaufbau eine Höhe von *exakt* $1.\overline{000}\,\mathrm{m}$ hat.

Und das sind nur die Probleme für den Versuchsaufbau. Auch die Fallzeit muss mit absoluter Genauigkeit gemessen werden, man braucht einen absoluten Wert für g, man muss den Luftwiderstand berücksichtigen etc.

Die Messgröße [ ]

Natürlich ist dieses Beispiel etwas übertrieben. Es veranschaulicht jedoch die Schwierigkeiten, etwas mit *absoluter Genauigkeit* zu messen. Irgendwann werden sich Messunsicherheiten immer auf Messungen in der realen Welt auswirken. Deshalb müssen diese Bedingungen in die Definition der zu messenden Größe einbezogen werden: die Fallzeit eines Objekts aus einer Höhe von 1 m, bei einer Temperatur von 20 °C, bei einem Längengrad x und einem Breitengrad y, ... Diese Beschreibung der zu messenden Größe und der Umgebungsbedingungen, die ihren Wert beeinflussen, wird als **Messgröße** bezeichnet. Da es nicht möglich ist, die Messgröße, d.h. alle experimentellen Bedingungen (die ihrerseits mit Messunsicherheiten behaftet sein können), absolut genau zu definieren, bedeutet dies, dass es einen **„wahren Wert“** der Messgröße—ein idealisiertes Ergebnis mit null Messunsicherheit—nicht geben kann! Man beachte, dass diese Nicht-Existenz der wahren Größe über die Unmöglichkeit hinausgeht, den 'wahren Wert' nicht bestimmen zu können. In der Wissenschaft des Messens—der Metrologie—wird der Begriff des wahren Wertes deshalb nicht mehr verwendet.

Gute wissenschaftliche Praxis erfordert die Reproduzierbarkeit von Ergebnissen. Daher muss bei der Berichterstattung über ein Experiment die Messgröße klar definiert werden. Diese Beschreibung erfolgt häufig in Form von Messgleichungen, sollte aber auch beinhalten, was gemessen wurde und welche Bedingungen das Messergebnis beeinflusst haben könnten. Außerdem sollte die Messunsicherheit, eine Quantifizierung der Variabilität der Messergebnisse, angegeben werden. Diese gibt die Genauigkeit des Messergebnisses an (eine geringe Messunsicherheit deutet auf ein genaues Ergebnis hin) und ist ein Hinweis auf die Qualität des Experiments.

Es gibt Fälle, in denen bestimmte Informationen über die Messgröße weggelassen werden können. Im obigen Beispiel könnte die Lufttemperatur die Fallzeit des Objekts beeinflusst haben. Denn bei der Verwendung alltäglicher Messinstrumente ist die Messunsicherheit bei Höhen- und Zeitmessungen viel größer als die Messunsicherheit aufgrund von Temperaturschwankungen. Daher ist der Einfluss von Temperaturschwankungen vernachlässigbar und kann aus der Messgröße herausgerechnet werden.

Bedingungen, die das Messergebnis beeinflussen können, sollten angegeben werden. So sollte beispielsweise bei der Angabe der Tischhöhe die Unsicherheit mit angegeben werden. Wenn die Zeitmessungen von Hand durchgeführt werden, sollte eine Schätzung der Reaktionszeit angegeben und in die Abschätzung der Messunsicherheit einbezogen werden.

Alltagsvorstellungen über den “wahren Wert” [ ]

Die Vorstellung von der Existenz eines “wahren Wertes” hält sich bei Lernenden sehr hartnäckig [1, 2, 3, 4]. Lernende streben oft nach einer “exakten Zahl” und wollen wissen, ob ihr Ergebnis richtig ist. Dieser Glaube an eine einzige und exakte Zahl entspricht dem Point-Paradigma-Denken.

Obwohl einige Lernende verstehen, dass ihr Messergebnis mit einer Messunsicherheit behaftet ist, denken sie vielleicht immer noch, dass Referenzwerte “wahre Werte” sind, mit denen sie ihr eigenes Ergebnis vergleichen können. Leider wird die Unsicherheit dieser Referenzwerte in den Schulbüchern sehr oft nicht angegeben. Auch wenn die Unsicherheit nicht angegeben wird, bedeutet das nicht, dass die Referenzwerte keine Unsicherheit aufweisen. Manchmal werden die angegebenen Werte bis zu einer Dezimalstelle angegeben, die von der Unsicherheit unberührt bleibt. Beispielsweise hat der Wert $G = 6{,}674 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$ weniger Nachkommastellen als der bekannte Wert von $G = (6{,}674\,30 \pm 0{,}000\,15) \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$. Trotz praktischer Gründe, die Unsicherheit wegzulassen, sollten sich die Lernenden bewusst sein, dass alle Referenzwerte [abgesehen von sieben fundamentalen Konstanten, die als absolute Größen definiert sind 5] gemessen wurden oder auf

Messungen beruhen und somit eine Unsicherheit aufweisen. Einige hilfreiche Fragen zur Diskussion mit Lernenden könnten sein:

- Wie wurde der Referenzwert bestimmt?
- War diese Messung mit Messunsicherheiten behaftet?
- Wie sicher bist du dir bei deiner Messung?
- Du sagst, U hat einen Wert von 5 V, wie sicher bist du, dass er nicht 5,1 V ist?
- Beschreibe mir genau, was du in deinem Experiment gemessen haben.
- Gibt es irgendwelche Quellen von Unsicherheiten in deinem Aufbau?
- Diskutiere mit deinem Sitznachbar*in, wie sinnvoll es ist, über den wahren Wert deiner Körpergröße zu sprechen.
- Kannst du die Genauigkeit deiner Messinstrumente einschätzen?
- Beschreibe möglichst detailliert, wie du die Zeit, die du für den Schulweg brauchst, möglichst genau messen kannst.

Mit diesen Impulsen können die Lernenden zu einem Verständnis des Set-Paradigmas geführt werden. Sie beginnen, ihr Experiment kritisch zu betrachten. Schließlich beginnen sie damit, die Messgröße genauer zu beschreiben. Außerdem beginnen sie, die Grenzen ihres Messergebnisses in Form eines Unsicherheitsintervalls auszudrücken.

Wie lang ist eine Banane? [ ]

Ein praktisches Beispiel, wie man den Lernenden das Konzept der Messgröße vorstellen kann, findet sich unter: Musold und Kok [6].

Literatur

[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* (9. Aufl.). 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 (Hrsg.), *Messunsicherheiten im Physikunterricht* (1. Aufl.). Berlin Universities Publishing. https://cloud.physik.hu-berlin.de/s/Yc87gQ334SFk78y

2.3 Messwiederholungen [ ]

Messungen werden wiederholt, damit man eine Vorstellung von der Variabilität der Messungen bekommt. Bei genügend wiederholten Messungen folgen diese Messungen in der Regel der Normalverteilung mit einem Mittelwert und einer Standardabweichung. Es stellt sich die Frage, wie viele Messungen dafür erforderlich sind.

Gründe für wiederholte Messungen [ ]

Einige Wissenschaftler*innen haben ein Gespür dafür entwickelt, wann man mit messen aufhören kann, andere verwenden Faustregeln. Wieder andere messen weiter, um die Genauigkeit zu erhöhen. Wenn man an der Variabilität der Messungen interessiert ist, möchte man den Mittelwert und die Standardabweichung kennen. Da die Standardabweichung der zugrundeliegenden Verteilung nicht von der Anzahl der Messungen abhängt (siehe 3.1), kann man mit den Messungen aufhören, sobald die Verteilung der Normalverteilung hinreichend entspricht. Letzteres geschieht in der Regel mit einem statistischen Test, der den Rahmen dieser Einheit sprengen würde.

Alternativ kann man mit weiteren Messungen nach Veränderungen im Wert der Standardabweichung und des Mittelwerts suchen. Wenn sich die ersten beiden signifikanten Ziffern der Standardabweichung (und die entsprechenden Ziffern des Mittelwerts) bei zusätzlichen Messungen nicht ändern, ist dies ein Hinweis auf eine hinreichend genau bestimmte Normalverteilung.

Alltagsvorstellungen über Messwiederholungen [ ]

Die wissenschaftlichen Gründe für die Wiederholung von Messungen können nur verstanden werden, wenn ein korrektes Verständnis der Messunsicherheiten vorhanden ist. Es ist daher nicht überraschend, dass die von den Lernenden angegebenen Gründe, warum sie ihre Messungen wiederholen müssen, stark von den oben beschriebenen abweichen.

Berücksichtigt man den Glauben der Lernenden an einen 'wahren Wert', so wird deutlich, dass die Lernenden keine Notwendigkeit sehen, Messungen zu wiederholen. Wenn man richtig misst, erhält man den richtigen Wert. Einige Lernende messen noch einmal, um ihr Ergebnis zu bestätigen [1]. Wenn die zweite Messung numerisch nicht identisch ist, sind einige Lernende verwirrt oder haben das Gefühl, etwas falsch gemacht zu haben. Andere messen weiter, bis sie wiederkehrende Werte finden.

Für einige Lernende ist die Reihenfolge der Messungen wichtig. Sie schenken der ersten Messung mehr Vertrauen, da das Gerät nach der Messung [2] verschleißen könnte. Obwohl dies tatsächlich der Fall sein kann, bedeutet dies, dass das Gerät nach jeder Messung repariert werden müsste.

Ein weiterer beobachteter Ansatz ist, dass Lernende wiederholte Messungen als eine Möglichkeit sehen, ihre Messtechnik zu üben [3]. Diese Lernenden geben dann den zuletzt gemessenen Wert an, da sie bei diesem die meiste Übung im Messen hatten. An sich ist es eine gute Strategie, einige Messungen zu Übungszwecken durchzuführen—man weiß, was einen erwartet und wie man die Messungen richtig durchführt. Wenn es sich bei diesen Messungen jedoch wirklich um Übungsreihen handelt, sollten sie nicht in den Datensatz aufgenommen werden, da mit Fehlern zu rechnen ist.

Andere Lernende führen routinemäßig Wiederholungsmessungen durch: Sie machen sechs Messungen, weil es ihnen gesagt wurde, weil sie es beim letzten Mal gemacht haben, um einen Mittelwert berechnen zu können oder weil "mehr einfach besser ist" [2, 4, 5].

All diese Überlegungen können mit dem Point-Paradigma in Verbindung gebracht werden. Die Lernenden verlassen sich auf einzelne Werte und sehen daher keinen Grund für wiederholte Messungen.

Die Gründe für wiederholte Messungen und die Frage, wann sie beendet werden sollten, sind sehr komplex. Wahrscheinlich sind sie für Lernende der Sekundarstufe I zu komplex. Die Lernenden können jedoch ein konzeptuelles Verständnis dafür haben, warum wiederholte Messungen durchgeführt werden: um eine Vorstellung von der Variabilität der Messungen zu bekommen. Für die Frage, wann man aufhören sollte, können einfache Faustregeln verwendet werden.

Einige Faustregeln [ ]

Die einfachste Faustregel besteht darin, die Anzahl der aufzuzeichnenden Messungen anzugeben. Im Allgemeinen können sechs bis zehn Messungen einen recht guten Eindruck von der Variabilität der Daten vermitteln und eine Annäherung an die Normalverteilung ermöglichen. Simulationen verschiedener alternativer Unsicherheitsquantifizierungen [siehe 6, 7] haben gezeigt, dass acht Wiederholungsmessungen eine günstige Anzahl darstellen. Die Quantifizierung "Extrema ausschließen", siehe Abschnitt 3.1, nähert sich bei dieser Anzahl von Wiederholmessungen dem Wert der Standardabweichung an.

Eine andere Faustregel besagt, dass man die Messungen stoppen sollte, wenn drei aufeinanderfolgende Messungen nicht kleiner/größer als die kleinste und größte Messung im Datensatz sind.

Tabelle 2: Beispiel-Datensatz: mit der Faustregel kann man sich entscheiden, die Messung nach sieben Messungen zu beenden.

Messung	Zeit [s]
1	2,18
2	3,02
3	2,82
4	3,61
5	2,47
6	3,31
7	2,24

Tabelle 2 zeigt einen Beispieldatensatz mit sieben Wiederholungsmessungen. Nach der vierten Messung liegen die nächsten drei aufeinander folgenden Messungen (5–7) zwischen dem Minimum (2,18 s) und dem Maximum (3,61 s) aus den ersten vier Werten der Reihe. Es kann entschieden werden, die Messung nach der siebten Messung zu beenden.

Ein wesentliches Problem dieser Regel besteht darin, dass die Reihenfolge der Messungen zu einem (explizit) wichtigen Aspekt wird. Trotz dieses Problems wird der Dialog darüber, warum diese Regel verwendet wird, den Weg zum Set-Paradigma Verständniss: Man möchte die Varianz sehen und wissen, wann diese hinreichend erfasst wurde. Sobald die Lernenden die Konzepte des Mittelwerts und der Unsicherheit vollständig verstanden haben, können sie beginnen, die Gründe für die zugrunde liegende Argumentation bezüglich wiederholter Messungen zu verstehen.

Literatur

[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.

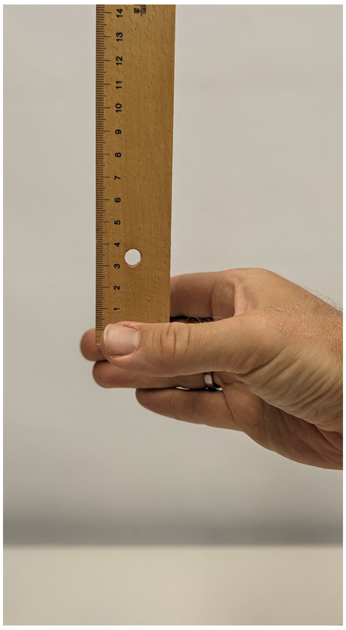
[7] Kok, K., & Priemer, B. (2023). Messunsicherheiten Quantifizieren: Welche Maße Gibt Es Dafür? *MNU Journal*, *76*(4), 330–333. https://doi.org/10.18452/27043

2.4 Mittelwert [ ]

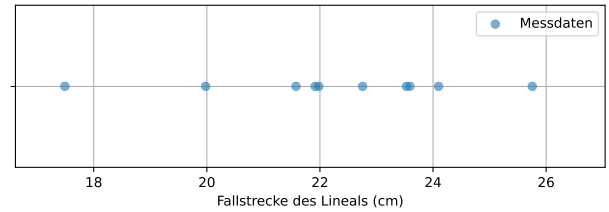
Das arithmetische Mittel einer Reihe von Messungen kann wie folgt berechnet werden:

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

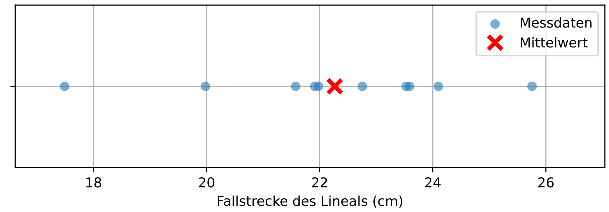
wobei N die Anzahl der wiederholten Messungen ist und x_i die einzelnen Messwerte sind.



(a) Messung der Fallstrecke eines Lineals zur Bestimmung der Reaktionszeit.



(b) Zehn wiederholte Messungen der Fallstrecke eines Lineals.



(c) Der Mittelwert (rotes Kreuz) ist der zentralste Punkt in diesem Datensatz.

Abbildung 1: Messdaten eines Experiments, bei dem die Reaktionszeit durch Messung der Fallstrecke eines Lineals bestimmt wird.

Dieser Mittelwert gilt als der Bestwert des Wertes der Messgröße aus einer Reihe von Messungen—der zentralste Punkt der Messreihe. Man kann es sich auch so vorstellen: Wenn man raten müsste, wie die nächste Messung ausfallen würde, liegt dieser Mittelwert—im Durchschnitt—am nächsten an dieser neuen Messung.

Dies deckt sich mit der Vorstellung, dass man mit dem Messen aufhören kann, wenn sich dieser Mittelwert (und die Unsicherheit) stabilisiert hat. Zu diesem Zeitpunkt stimmt dieser Mittelwert mit dem zentralsten Wert der Messungen überein, die das Experiment hervorbringt.

Das arithmetische Mittel wird meist als der Bestwert der Messgröße verwendet. Es gibt jedoch auch alternative Mittelwerte. Eine Möglichkeit ist die Verwendung des **Medians**: der mittlere Wert eines sortierten Datensatzes. Ein anderes Maß ist der **Modus**: die Messung, die am häufigsten auftritt. In dem Fall, dass wiederholte Messungen alle genau das gleiche Ergebnis liefern, ist der Bestwert schließlich nur dieser eine Wert.

Angenommen, man möchte die Reaktionszeit einer Person messen. Dazu kann man die Fallstrecke eines Lineals messen, siehe Abb. 1a. Mit Hilfe der Gleichung $h = \frac{1}{2}gt^2$ kann die Reaktionszeit bestimmt werden. Die gemessenen Daten werden in Abb. 1b grafisch dargestellt. Der Mittelpunkt dieser Messungen ist der Mittelwert, der in Abb. 1c als rotes Kreuz dargestellt ist.

Alltagsvorstellungen über den Mittelwert [ ]

Die Berechnung des Mittelwerts stellt für die Lernenden normalerweise kein Problem dar. Allerdings bleiben die Überlegungen der Lernenden darüber, was der Mittelwert repräsentiert, oft oberflächlich [1, 2].

Manchmal wird der Mittelwert als Grund für die Durchführung wiederholter Messungen genannt. Da man natürlich mehrere Messungen durchführen muss, um einen Mittelwert zu berechnen [3], ist dies oft ein routinemäßiges und automatisiertes Verfahren, und der Mittelwert wird anschließend als isolierter Wert betrachtet. Einige Lernende gehen sogar noch weiter und betrachten diesen Wert als "wahren Wert" [4]. Obwohl der Prozess der Berechnung eines Mittelwerts aus einer Reihe von wiederholten Messungen mit dem Set-Paradigma in Verbindung gebracht werden kann, wird die Betrachtung des Mittelwerts als isolierter Wert als Point-Paradigma betrachtet [5]. Um zu einer adäquaten Argumentation im Set-Paradigma zu gelangen, müssen die Lernenden den

Mittelwert um die Messunsicherheit ergänzen.

Eine weitere Vorstellung über den Mittelwert ist die, dass er "präziser" ist [6]. Obwohl ein (stabilisierter) Mittelwert die beste Schätzung des Wertes der Messgröße ist, sagt dies nichts über die Präzision aus, die durch die Messunsicherheit angegeben wird.

Literatur

[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 (Hrsg.), *Research in Science Education - Past, Present, and Future* (S. 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 Unsicherheit [📘 ✐]

Die Unsicherheit oder Messunsicherheit einer Messung ist eine Quantifizierung der Variabilität der Werte, die mit dem Ergebnis der Messung verbunden sein können. In den meisten Fällen folgen die Werte von wiederholten Messungen in einem Experiment der Normalverteilung. Daher ist es nicht verwunderlich, dass die Standardabweichung (oder die Standardabweichung des Mittelwerts) die gängigste wissenschaftliche Quantifizierung der Messunsicherheit ist.

Natürlich gibt es Messungen, bei denen keine Variabilität auftritt (zum Beispiel die Messung der Länge eines Tisches mit einem Maßband). Hier wird die Unsicherheit durch das Messinstrument bestimmt. Mehr Informationen dazu im Abschnitt 3.2.

Standardabweichung der Verteilung [📘]

Die Standardabweichung der Verteilung ist der wissenschaftliche Standard zur Beschreibung der Variabilität einer Messgröße um ihren Mittelwert. Sie kann wie folgt berechnet werden:

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

Dabei sind N die Anzahl der Messungen, x_i die einzelnen Messwerte und $\bar{x}$ der Mittelwert der Messung.

Wenn die Daten normalverteilt sind, siehe Abb. 2, deckt die Fläche unter der Normalverteilung zwischen $-\sigma$ und $+\sigma$ 68 % der Gesamtfläche ab. Das bedeutet, dass 68 % der Messungen im 1σ-Unsicherheitsintervall (d.h. im Wertebereich zwischen $\bar{x} - \sigma$ und $\bar{x} + \sigma$) liegen werden. Für das 2σ-Intervall erhöht sich dieser Prozentsatz auf 95 %.

Betrachtet man wieder die Daten der Fallstrecke des Lineals, siehe Abb. 3a, so sind auch diese Daten mit einer Unsicherheit behaftet. Mit Hilfe der Gleichung (2) kann diese Unsicherheit berechnet werden, sie ist in Abb. 3b dargestellt, das Unsicherheitsintervall ist in Abb. 3c dargestellt. Da

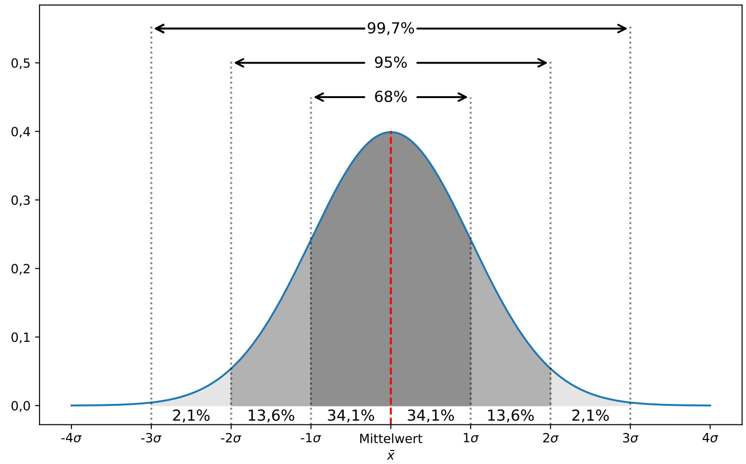


Abbildung 2: Eine Normalverteilung mit einem Mittelwert bei Null (rote gestrichelte Linie) und einer Standardabweichung σ. Die gestrichelten Linien zeigen die Grenzen von einem, zwei und drei Sigma an. Die prozentualen Angaben beschreiben den Anteil an Messwerten in dem entsprechenden Intervall.

das Unsicherheitsintervall 68 % der Messungen umfasst, liegen nicht alle Messwerte innerhalb des Unsicherheitsintervalls. Abbildung 3d zeigt auch den gesamten Wertebereich, der alle Messungen umfasst.

Standardabweichung des Mittelwerts [★]

Die Standardabweichung des Mittelwerts gibt, wie der Name bereits andeutet, die Messunsicherheit des Mittelwerts an (und nicht die Messunsicherheit der einzelnen Messungen).

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

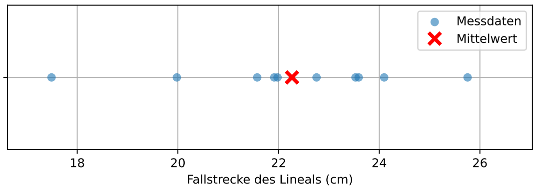
Die Standardabweichung des Mittelwerts wird manchmal auch als Standardfehler bezeichnet. Aufgrund der unpassenden Konnotation des Wortes Fehler wird die Verwendung des Begriffs Standardabweichung des Mittelwerts oder Standardunsicherheit bevorzugt.

Im Gegensatz zur Standardabweichung einer Verteilung wird die Standardabweichung des Mittelwerts mit zunehmender Zahl der Messungen kleiner. Konzeptionell ist dies zu verstehen, weil mit zunehmender Anzahl von Messungen, auch wenn sie mit großen Messunsicherheiten behaftet sind, die zugrunde liegende Verteilung immer deutlicher wird. Aus dieser Verteilung kann das Zentrum (der Mittelwert) mit größerer Sicherheit bestimmt werden.

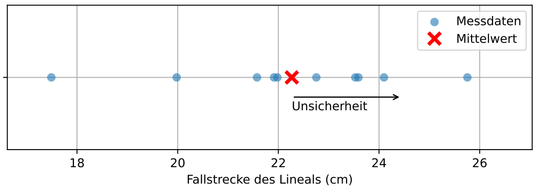
Vorstellungen der Lernenden über die Messunsicherheiten [ ]

Viele Lernende sehen Messunsicherheiten als etwas an, das in einem Experiment eliminiert werden kann [1, 2, 3]. Dies hängt stark mit dem Glauben an die Existenz eines “wahren Wertes” zusammen, siehe Abschnitt 2.2. Einige Lernende glauben, dass die Messunsicherheit mit genügend Übung auf Null reduziert werden kann, andere glauben, dass dies nur mit “echten” Messinstrumenten möglich ist, wieder andere denken, dass dies nur von “professionellen” Wissenschaftler*innen erreicht werden kann. Wie auch immer diese Vorstellungen sind: Messunsicherheiten werden immer in jedem Experiment auftauchen. Sie sind in *jedem* Experiment vorhanden und ungleich null.

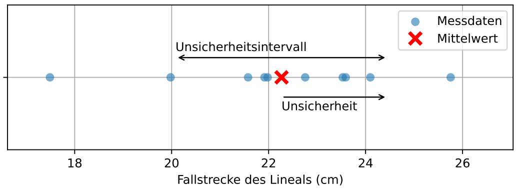
Einige Lernende und sogar Lehrbücher bezeichnen Messunsicherheiten als “Fehler”. Es ist bekannt, dass die Verwendung dieses Wortes die Lernenden zu der Annahme verleitet, dass sie etwas falsch gemacht haben [3, 4, 5, 6, 7, 8]. Auch der Guide to the Uncertainty of Measurements [kurz GUM 9] verzichtet ausdrücklich auf die Verwendung des Begriffs “Fehler”. Das Wort “Fehler” steht im Zusammenhang mit Messunsicherheiten für einen tatsächlichen Fehler: einen (nicht behebba-



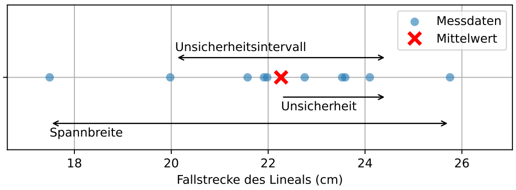
(a) Die Messungen der Fallstrecke eines Lineals.



(b) Die Unsicherheit dieses Datensatzes.



(c) Die Unsicherheit und das Unsicherheitsintervall dieses Datensatzes.



(d) Der Spannbreite, der Unsicherheit und dem Unsicherheitsintervall.

Abbildung 3: Messdaten der Fallstrecke eines Lineals, wie in Abb. 1 jetzt mit der Unsicherheit.

ren) Fehler oder etwas, das schief gelaufen ist. Messunsicherheiten hingegen gehören zu jedem wissenschaftlichen Experiment und lassen sich nicht vermeiden.

Die Allgegenwart von Messunsicherheiten und die Vorstellung, dass alle Wissenschaftler*innen damit umgehen müssen—und die daraus resultierende Ungewissheit—kann ein beruhigender Gedanke für die Lernenden sein. Es zeigt ihnen, dass sie nicht etwas falsch machen, sondern sich in einer authentischen wissenschaftlichen Praxis bewegen. Dadurch entwickelt sich gleichzeitig eine angemessene Sicht auf das Wesen der Naturwissenschaft [10, 11, 12].

Hierbei ist wichtig zu beachten, dass die Standardabweichung in der Sekundarstufe I zu komplex oder zeitaufwändig sein könnte (Alternativen werden später in Abschnitt diskutiert 3.1). Folgendes häufiges Missverständnis soll hier noch richtiggestellt werden:

die Auffassung dass mehr Messungen zu einer geringeren Standardabweichung führen [13]. Dies ist jedoch nicht der Fall. Mehr Messungen (x_i) führen nicht zu einem kleineren Wert der Standardabweichung (σ), siehe Gleichung (2). Vielmehr erhält man mit mehr Messungen eine bessere Beschreibung der Verteilung der Messwerte, d.h. die Form der Normalverteilung wird deutlicher. Die Breite der Verteilung ändert sich jedoch nicht. Im gegensatz dazu gilt: die Standardabweichung des Mittelwertes ($\sigma_{\bar{x}}$) wird kleiner. Diese Quantifizierung ist den Lernenden jedoch nicht immer bekannt, oder sie können begrifflich nicht zwischen den beiden unterscheiden.

Literatur

[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, Dezember). *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* (Techn. Ber. Nr. 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 "Messfehlernïm 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 (Hrsg.), *Handbook of Research on Science Education* (S. 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 (Hrsg.), *Physics Education* (S. 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 Fehler (Abweichung) [ ]

Das Wort Fehler wird manchmal ungünstigerweise als Bezeichnung für die Unsicherheit einer Messung verwendet. Fehler im Sinne von Verfahrensfehlern oder Irrtümern sind jedoch etwas, das vermieden werden sollte. Messunsicherheiten hingegen lassen sich nicht vermeiden.

Eine andere Verwendung des Wortes Fehler geschieht manchmal im Sinne von Abweichung: die (zu große) Differenz zwischen dem Bestwert und einem Referenzwert. Um Verwechslungen zu vermeiden, wird das Wort Fehler nur im Zusammenhang mit (korrigierbaren) Verfahrensfehlern oder Irrtümern verwendet.

Definition [ ]

Die Messabweichung ist die Differenz zwischen dem Bestwert und einem Referenzwert:

$$\text{Abweichung} = |\text{Bestwert} - \text{Referenzwert}|. \tag{4}$$

Wenn bei einem Experiment alles gut läuft, sollte der Referenzwert innerhalb des Unsicherheitsintervalls liegen. Man sagt, dass die Abweichung klein ist und innerhalb der Messunsicherheit liegt.

Allerdings kann es **systematische Effekte** geben, die zu einer Verschiebung des Bestwertes mit seinem Unsicherheitsintervall gegenüber dem Referenzwert führen. Zum Beispiel kann ein Messgerät falsch kalibriert sein ("nicht auf Null gestellt"), oder der Startpunkt eines bewegten Objekts wurde verschoben, usw. In diesen Fällen haben *alle* Messungen die gleiche Verschiebung ihres Wertes. Wenn diese systematischen Effekte erkannt werden, kann man diese Messabweichung gegebenenfalls korrigieren.

Abweichungen können also stets der Ausgangspunkt für die Suche nach den Ursachen sein. Gibt es einen systematischen Effekt? Werden die Unsicherheiten generell unterschätzt? Ist die Messgröße richtig definiert? Ist der Referenzwert angemessen? Usw.

Alltagsvorstellungen über Fehler [ ]

Wie im Abschnitt 2.7 diskutiert, bezeichnen einige Lernende Messunsicherheiten als Fehler, was dazu führen kann, dass sie glauben, einen Fehler gemacht zu haben [1, 2, 3, 4, 5, 6]. Es wird empfohlen, das Wort Fehler im Zusammenhang mit Messdaten im Allgemeinen zu vermeiden. Wenn man sich auf die Varianz von Messungen bezieht, ist der Begriff Messunsicherheit vorzuziehen, und wenn man sich auf die Differenz zwischen einem Bestwert und einem Referenzwert bezieht, ist der Begriff Abweichung vorzuziehen.

$$\left.\begin{array}{l} I = \mathtt{0{,}2274}\color{red}{\mathtt{59}}\ \mathrm{A} \\ u_I = \mathtt{0{,}0050}\ \mathrm{A} \end{array}\right\} \quad I = (0{,}2275 \pm 0{,}0050)\ \mathrm{A}$$

Abbildung 4: Der Bestwert des Stroms wird entsprechend der Unsicherheit des Stroms gerundet, d. h. auf vier Dezimalstellen.

Eine weitere weit verbreitete Vorstellung ist, dass eine kleine Abweichung immer besser ist. Eine geringe Abweichung ist zwar ein Hinweis auf eine gute Genauigkeit, bedeutet aber nicht unbedingt, dass es sich um ein gutes Messergebnis handelt. Angenommen, zwei Gruppen versuchen den Widerstand eines $R = 100\,\Omega$-Widerstands zu überprüfen. Gruppe A misst $R_\mathrm{A} = (99 \pm 10)\,\Omega$ und Gruppe B misst $R_\mathrm{B} = (103 \pm 5)\,\Omega$. Die Abweichung der Gruppe A, $d_\mathrm{A} = 1\,\Omega$, ist kleiner als die der Gruppe B, $d_\mathrm{B} = 3\,\Omega$. Die Messunsicherheit der Gruppe B ist jedoch nur halb so groß wie die der Gruppe A, was auf eine bessere Präzision hindeutet. Da beide Messergebnisse mit dem Referenzwert kompatibel sind (mehr über den Vergleich von Messergebnissen in Abschnitt 4.1), könnte man argumentieren, dass das Ergebnis der Gruppe B besser ist.

Literatur

[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, Dezember). *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 Messergebnis [ ]

Da Messunsicherheiten die Qualität eines Experiments anzeigen, ist kein Messergebnis ohne die Angabe der Unsicherheit vollständig. Im Folgenden werden die verschiedenen Möglichkeiten zur Angabe des Messergebnisses aufgezeigt, gefolgt von den Regeln für die Anzahl der signifikanten Ziffern, die angegeben werden müssen.

Dokumentation von Messergebnissen [ ]

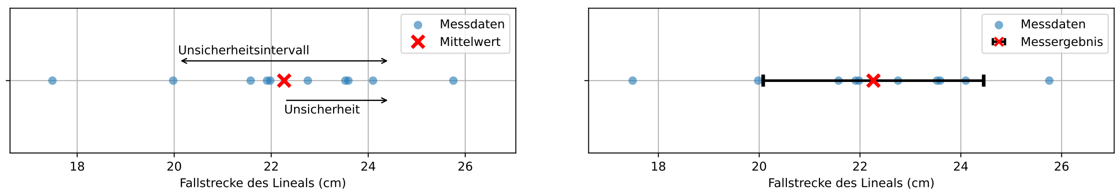
Ein vollständiges Messergebnis gibt immer den Bestwert und die Messunsicherheit an. Meist wird dies wie folgt angegeben:

$$\text{Messergebnis} = \text{Bestwert} \pm \text{Messunsicherheit}. \tag{5}$$

Der Bestwert ist meist der Mittelwert $\bar{x}$, und die Messunsicherheit die Standardabweichung σ.

Die Rundung des Bestwertes hängt von der Anzahl der Ziffern der Unsicherheit ab. Der Bestwert erhält genauso viele Dezimalstellen wie die Unsicherheit. Ein Beispiel: Ein Strom wird gemessen: $I = 0{,}227\,459\,\mathrm{A}$ mit einer Unsicherheit $u_I = 0{,}0050\,\mathrm{A}$, siehe Abb. 4. Die Unsicherheit hat vier Dezimalstellen, daher wird der Bestwert auf vier Dezimalstellen gerundet:

$$I = (0{,}2275 \pm 0{,}0050)\,\mathrm{A} \tag{6}$$



(a) Die Unsicherheit und das Unsicherheitsintervall.

(b) Die grafische Darstellung des Messergebnisses.

Abbildung 5: Messdaten der Fallstrecke eines Lineals, wie in Abb. 1 jetzt mit volständiges Messergebniss.

Tabelle 3: Die Anzahl der signifikanten Ziffern in blau für bestimmte Größen.

Größe	signifikante Ziffern
$l = 17{,}2\,\mathrm{cm}$	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

Eine weitere Möglichkeit, Unsicherheit zu kennzeichnen, besteht darin, sie in Klammern zu setzen. Dies lässt sich am besten anhand des obigen Beispiels veranschaulichen. Durch die Schreibweise in Klammern wird dies zu:

$$I = 0{,}2275(50)\,\mathrm{A}. \tag{7}$$

Diese Schreibweise ist für Lernende etwas weniger intuitiv. Die DIN bevorzugt diese Notation für die Industrie, da sie sich von der Toleranz abgrenzt, die die gleiche Notation wie (6) hat. Die Toleranz gibt die maximale Abweichung zwischen Messungen und einem Referenzwert an, die z.B. in Produktionsprozessen *erlaubt* ist.

Die Messunsicherheit ist ein Hinweis auf die Genauigkeit eines Messergebnisses und damit ein Hinweis auf die Qualität des Experiments. Die Messunsicherheit erstreckt sich über ein Unsicherheitsintervall um den Bestwert. Das Unsicherheitsintervall reicht von:

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

Das Unsicherheitsintervall kann als der Bereich von Werten angesehen werden, in dem die Messgröße (mit einem bestimmten Grad des Vertrauens) zu erwarten ist.

Betrachtet man wieder die Messdaten der Fallstrecke eines Lineals, siehe Abb. 5a, so kann dieses Messergebnis grafisch dargestellt werden. Die Standardmethode zur Darstellung dieses Ergebnisses ist die Verwendung von Unsicherheitsbalken (manchmal auch als Fehlerbalken bezeichnet), wie in Abb. 5b gezeigt. Es ist zu beachten, dass in diesem Fall die Unsicherheit für die Variable auf der x-Achse dargestellt wird, während für die Unsicherheiten auf der y-Achse die Unsicherheitsbalken vertikal verlaufen.

Anzahl der signifikanten Ziffern [ ]

Im Beispiel von vorhin wurde die Messunsicherheit des Stroms mit zwei signifikanten Ziffern angegeben. Die Anzahl signifikante Ziffern ist die Anzahl der Ziffern ohne führende Nullen, siehe Tab. 3. Die Anzahl der signifikanten Ziffern der Messunsicherheit bestimmt die Anzahl der signifikanten Ziffern der besten Schätzung, die entsprechend der Messunsicherheit gerundet wird, siehe Abb. 4. Es ist zu beachten, dass die Anzahl der signifikanten Ziffern des Bestwertes immer größer oder gleich der Anzahl der signifikanten Ziffern der Messunsicherheit ist. Es gibt drei Regeln für die Anzahl der signifikanten Ziffern der Messunsicherheit:

- Die Messunsicherheit wird mit einer signifikanten Ziffer ausgedrückt.
  Beispiele:
  $u = 3{,}581\,\mathrm{cm} = 4\,\mathrm{cm}$
  $u = 149\,\mathrm{m} = 1 \times 10^2\,\mathrm{m} = 0{,}1\,\mathrm{km}$

$u = 0{,}005\,01\,\mathrm{A} = 0{,}005\,\mathrm{A} = 5\,\mathrm{mA}$
$u = 0{,}029\,48\,\mathrm{L} = 0{,}03\,\mathrm{L}$

- Die Messunsicherheit wird mit zwei signifikanten Ziffern ausgedrückt.
  Beispiele:
  $u = 3{,}581\,\mathrm{cm} = 3{,}6\,\mathrm{cm}$
  $u = 149\,\mathrm{m} = 1{,}5 \times 10^2\,\mathrm{m} = 0{,}15\,\mathrm{km}$
  $u = 0{,}005\,01\,\mathrm{A} = 0{,}0050\,\mathrm{A} = 50\,\mathrm{mA}$
  $u = 0{,}029\,48\,\mathrm{L} = 0{,}029\,\mathrm{L}$

- Die Unsicherheit wird mit einer signifikanten Ziffer angegeben, es sei denn, die erste signifikante Ziffer ist eine 1 oder eine 2. In diesem Fall werden zwei signifikante Ziffern angegeben.
  Beispiele:
  $u = 3{,}581\,\mathrm{cm} = 4\,\mathrm{cm}$
  $u = 149\,\mathrm{m} = 1{,}5 \times 10^2\,\mathrm{m} = 0{,}15\,\mathrm{km}$
  $u = 0{,}005\,01\,\mathrm{A} = 0{,}005\,\mathrm{A} = 5\,\mathrm{mA}$
  $u = 0{,}029\,48\,\mathrm{L} = 0{,}029\,\mathrm{L}$

Der GUM [siehe 1] bevorzugt keine spezifische Regel für die Anzahl der signifikanten Ziffern der Unsicherheit. Jedoch wird angemerkt, dass maximal zwei Ziffern zu verwenden sind.

Für die **Rundung** der Messunsicherheit gibt es zwei Möglichkeiten: normales Runden (wie oben) oder Aufrunden. Ein konsequentes Aufrunden (z. B. $0{,}005\,01\,\mathrm{A} = 0{,}0051\,\mathrm{A}$ oder sogar $0{,}005\,01\,\mathrm{A} = 0{,}006\,\mathrm{A}$) ist ein sehr konservatives Verfahren, das zu Überschätzungen der Unsicherheit führen kann. Der GUM ist hier nicht eindeutig, scheint aber die normale Rundung zu bevorzugen. Es wird betont, dass es angemessen ist, Unsicherheiten mit gesundem Menschenverstand aufzurunden. Dies würde bedeuten, $u = 0{,}029\,48\,\mathrm{L}$ auf $0{,}030\,\mathrm{L}$ aufzurunden, aber $u = 0{,}005\,01\,\mathrm{A}$ auf $0{,}0050\,\mathrm{A}$ abzurunden.

Vorstellungen der Lernenden über das Messergebnis [ ]

Auf die Frage, wie ein Messergebnis angegeben werden sollte, geben viele Lernende (selbst auf Universitätsebene) an, dass der Mittelwert angegeben werden sollte [2]. Für ein vollständiges Messergebnis sollte dieser jedoch durch die Unsicherheit ergänzt werden. Um die Lernenden zum Nachdenken darüber anzuregen, könnte man sie fragen, wie ihr Messergebnis die Qualität ihres Experiments widerspiegelt.

Literatur

[1] Joint Committee for Guides in Metrology. (2008). *Evaluation of Measurement – Guide to the Expression of Uncertainty in Measurement* (Techn. Ber. Nr. 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* (Techn. Ber. Nr. Project PL 95-2005). Centre for Studies in Science and Mathematics Education. Leeds.

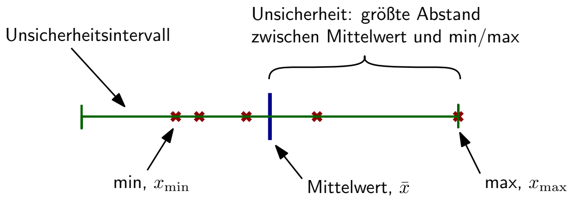


Abbildung 6: Grafische Darstellung der Berechnung der maximalen Abweichung. Die roten Kreuze sind die einzelnen Messungen, der blaue Balken zeigt den Mittelwert an, und die Messunsicherheit ist der größte Abstand zwischen dem Mittelwert und entweder dem Mindest- oder dem Höchstwert des Datensatzes. Der grüne Balken zeigt das gesamte Unsicherheitsintervall um den Mittelwert an.

3 BESTIMMUNG DER UNSICHERHEIT

Im vorangegangenen Schritt wurden alle grundlegenden Konzepte und die Terminologie behandelt. In diesem Schritt geht es um die Quantifizierung der Messunsicherheit. Dies geschieht für wiederholte Messungen, Einzelmessungen und die grafische Auswertung mit Diagramme. Danach werden einige Regeln für die Unsicherheitsfortpflanzung und die relative Messunsicherheit behandelt. Jeder Teil beginnt mit den Grundlagen, gefolgt von didaktischen Überlegungen.

3.1 Typ A [ ]

Wenn Messungen Schwankungen unterworfen sind, sollten wiederholte Messungen durchgeführt werden, um das Messergebnis zu ermitteln (siehe 2.3). Die Messunsicherheit kann in diesen Fällen mit statistischen Mitteln bestimmt werden. Diese Art der Unsicherheitsbewertung wird als **Typ A**-Unsicherheitsbewertung bezeichnet.

Verringerung der Komplexität [ ]

Die Standardabweichung und die Standardabweichung des Mittelwerts (siehe Abschnitt 2.7) sind die häufigsten Quantifizierungen der Messunsicherheit in der wissenschaftlichen Praxis. Oft haben Lernende Schwierigkeiten, diese zu berechnen und zu interpretieren [siehe z.B.:, 1, 2, 3]. Um ihnen zu helfen, könnte man den Prozess der Unsicherheitsberechnung automatisieren. Allerdings hat man festgestellt, dass die Lernenden das Gefühl für die Glaubwürdigkeit des Ergebnisses dadurch verlieren [3].

Im Gegensatz dazu könnte man die Messungen auch grafisch darstellen. Dadurch wird die Variabilität der Messungen visualisiert, was die kognitive Belastung für die Lernenden verringert [4, 5]. Jedoch fehlt bei diesem Ansatz eine numerische Quantifizierung der Unsicherheit, was von Forscher*innen bevorzugt wird und auch in der Schule notwendig ist [1, 3].

Alternative Quantifizierungen [ ]

Eine weitere Möglichkeit besteht darin, andere Quantifizierungen zu verwenden, die leichter zu berechnen und konzeptionell zu verstehen sind. Im Folgenden werden vier alternative Quantifizierungen definiert. Die Forschung hat gezeigt, dass beim Vergleich dieser Quantifizierungen mit der Standardabweichung mit zunehmender mathematischer Komplexität auch die statistische Qualität zunimmt [6].

Die einfachste Quantifizierung ist die **maximale Abweichung**. Um diese zu berechnen, sortiert man alle Messungen vom kleinsten (x_1) zum größten (x_N) und berechnet den Mittelwert ($\bar{x}$). Die größte Differenz zwischen dem Mittelwert und dem kleinsten bzw. größten Wert der Reihe ist die Unsicherheit:

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

Dieses Maß ist in Abb. 6 grafisch dargestellt. Der Vorteil dieser Quantifizierung ist ihre mathematische Einfachheit. Der Nachteil ist, dass es eine starke Überschätzung der Unsicherheit (im Vergleich zur Standardabweichung) darstellt, die sehr anfällig für Ausreißer ist.

Um Ausreißer zu kompensieren, kann dieses Verfahren so angepasst werden, dass der kleinste (x_1) und der größte (x_N) Wert der Reihe ausgeschlossen werden und das Verfahren für die übrigen

Messungen wiederholt wird. Dies wird als **ausschließen von Extremwerten** Quantifizierung bezeichnet:

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

Diese Quantifizierung ist fast so einfach wie die **maximale Abweichung**, aber weniger empfindlich für Ausreißer. Mit zunehmender Anzahl von Messwiederholungen wird diese Quantifizierung jedoch ebenso empfindlich für Ausreißer.

Bei größeren Stichprobenumfängen könnte man, das Verfahren für die mittlere Hälfte der Messungen anwenden, dies wird als **mittlere 50%** Verfahren bezeichnet. Dazu werden nach und nach Paare von Extremwerten ausgeschlossen, bis die verbleibende Anzahl von Messungen weniger als die Hälfte der Gesamtzahl der Messungen beträgt. So wird beispielsweise bei 4-7 Messungen ein Paar der Extremwerte ausgeschlossen, bei 8-11 Messungen zwei Paare usw. In Form einer Gleichung kann dies wie folgt geschrieben werden:

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

Der Nachteil dieser Maßnahme ist, dass sie für die Lernenden unbefriedigend sein könnte. Sie könnten sich fragen, warum überhaupt wiederholte Messungen durchgeführt wurden, wenn die Hälfte ihres Datensatzes nicht verwendet wird.

Die letzte Alternative ist die **mittlere absolute Abweichung (MAA)**. Dies ist der Durchschnitt der Abweichungen aller Einzelmessungen vom Mittelwert:

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

Da die Berechnung der MAA die Berechnung eines Mittelwerts beinhaltet, ist sie konzeptionell leicht zu verstehen. Diese Unsicherheit ergibt einen Wert, der systematisch kleiner ist als die Standardabweichung.

Weitere Lektüre [💡]

Weitere Informationen zu den verschiedenen Unsicherheitsquantifizierungen und einige Beispiele finden Sie unter Kok und Priemer [7].

Für einige Ideen für Experimente, bei denen die Analyse der Messunsicherheit eine Notwendigkeit ist, um eine korrekte Schlussfolgerung zu ziehen, siehe: [8, 9, 10, 11, 12, 13, 14]

Literatur

[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 (Hrsg.), *Messunsicherheiten im Physikunterricht* (1. Aufl.). Berlin Universities Publishing. https://cloud.physik.hu-berlin.de/s/Yc87gQ334SFk78y
[13] Nagel, C. (2021). Sicher ist sicher! Fachliche Klährung für die didaktische Rekonstruktion von Messunsicherheiten im Unterricht. *Plus Lucis*, (4), 7–11.
[14] 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 Typ B [★]

Wenn wiederholte Messungen immer den gleichen Messwert ergeben oder wenn nur eine Messung durchgeführt werden kann, ist eine Unsicherheitsbewertung vom Typ A nicht geeignet. Im Falle wiederholter Messungen bedeutet dies wahrscheinlich, dass die Messinstrumente nicht genau genug sind, um die Abweichung zu messen. Das bedeutet jedoch nicht, dass es keine Messunsicherheit gibt! In diesen Fällen muss die Unsicherheit auf andere Weise mit Hilfe einer Liste von quantifizierte Unsicherheitsquellen geschätzt werden. Dies wird als eine Unsicherheitsanalyse vom **Typ B** bezeichnet.

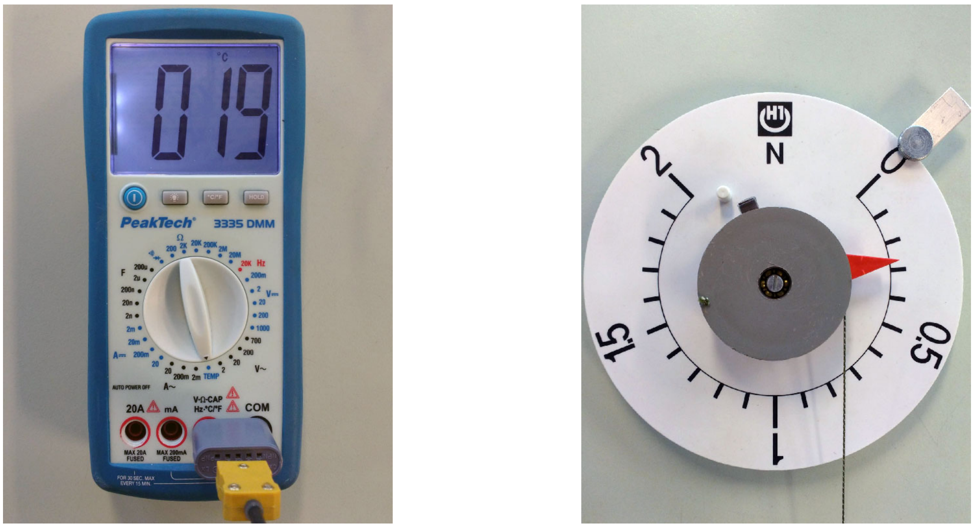
Unsicherheitsquellen [★ 📕]

Um eine Liste von Unsicherheitsquellen zu erstellen, betrachtet man zunächst das Experiment und sucht nach Ursachen von Unsicherheiten. Diese Unsicherheiten können in mehrere Kategorien eingeteilt werden: der experimentelle Prozess, die Umgebungsbedingungen, die Messinstrumente, mathematische Rundungen und die experimentierende Person [1, 2]. Diese Unsicherheiten müssen abgeschätzt und aufsummiert werden, um die "endgültige" Unsicherheit zu erhalten.

Die Unsicherheit des Messgerätes selbst kann in drei Komponenten unterteilt werden: die Eichunsicherheit (in der Regel auf dem Gerät angegeben), die Linearitätsunsicherheit (wie genau die Markierungen auf einer Skala sind oder wie genau die Digitalisierung des Gerätes ist) und die Ableseunsicherheit (die endliche Anzahl von Markierungen auf der Skala) [3]. Für die Ablesungenauigkeit gibt es Faustregeln für digitale und analoge Geräte:

Digitale Messgeräte: Die Messunsicherheit für ein digitales Messgerät ist ein Skalenteil. Zum Beispiel wird eine Temperatur von $T = 19\,°\mathrm{C}$ mit einem digitalen Thermometer gemessen, siehe Abb. 7a. Das Gerät hat eine Skalenteilung von $1\,°\mathrm{C}$, sodass die Ableseunsicherheit $u_{\mathrm{scale}} = 1\,°\mathrm{C}$ beträgt. Der Grund für die Wahl eines ganzen Skalenteils als Unsicherheit ist, dass man nicht wissen kann, wie das Gerät die Werte rundet (wird der Wert 7,6 °C auf 7 °C abgeschnitten oder korrekt auf 8 °C gerundet). Dadurch wird das Messergebnis zu $T = (19 \pm 1)\,°\mathrm{C}$.

Analoge Messgeräte: Die Unsicherheit für ein analoges Messgerät ist ein halber Skalenteil. Zum Beispiel misst ein Federkraftmesser eine Kraft F, siehe Abb. 7b. Die Anzeige liegt zwischen 0,2 N und 0,3 N, so dass der beste Schätzwert 0,25 N wäre. Das Messgerät hat Markierungen in Abständen von 0,1 N. Da es sich um ein analoges Messgerät handelt, beträgt die Messunsicherheit $u_{\mathrm{scale}} = 0{,}05\,\mathrm{N}$. Der Grund für diese Wahl ist, dass man selbst (richtig) rundet. Das Messergebnis ist somit $F = (0{,}25 \pm 0{,}05)\,\mathrm{N}$.



(a) Ein Thermometer misst $(19 \pm 1)\,°\mathrm{C}$.

(b) Ein Federkraftmesser misst $(0{,}25 \pm 0{,}05)\,\mathrm{N}$.

Abbildung 7: Zwei Messgeräte.

Didaktische Reduktion [★]

Bei einer Unsicherheitsbewertung vom Typ B werden meist praktische Überlegungen angestellt. Bei der Messung der Zeit, die eine Person benötigt, um eine Strecke von 100 m zurückzulegen, könnte beispielsweise eine geschätzte Reaktionszeit von 0,5 s berücksichtigt werden.

Bei Messungen wie in Abb. 7 könnte man beschließen, nur die Ableseunsicherheit einzubeziehen, sich aber nicht um die Unsicherheiten der Eichung und Linearisierung zu kümmern (die viel schwieriger zu schätzen sind).

Schließlich könnte man einen Wertebereich schätzen, in dem man die Messung plausibel erwarten kann. Nehmen wir zum Beispiel an, dass beim Messen der Länge eines Klassenzimmers mit einem 1 m-Lineal mit 1 cm-Markierungen das Lineal achtmal verschoben werden muss. Die Verschiebung des Lineals führt zu einer zusätzlichen Unsicherheit zur 0,5 cm Skalenunsicherheit. Man könnte beschließen, für jede Verschiebung eine Unsicherheit von 1 cm zu berücksichtigen. Dies führt zu einer Gesamtunsicherheit von $8 \cdot 1\,\mathrm{cm} + 0{,}5\,\mathrm{cm} = 8{,}5\,\mathrm{cm}$.

Weitere Lektüre [★]

Weitere Hintergrundinformationen und einige praktische Beispiele zur Unsicherheit von Messinstrumenten findet man unter: Nagel [3].

Literatur

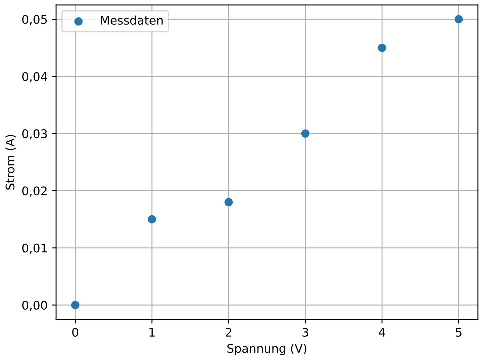
[1] Hellwig, J. (2012). *Messunsicherheiten Verstehen : Entwicklung Eines Normativen Sachstrukturmodells Am Beispiel Des Unterrichtsfaches Physik* [Doctoral Thesis]. Ruhr-Universität. https://nbn-resolving.org/urn:nbn:de:hbz:294-36561.

[2] Hennes, M. (2007). Konkurrierende Genauigkeitsmaße–Potential Und Schwächen Aus Der Sicht Des Anwenders. *Allgemeine Vermessungs-Nachrichten*, *7*, 136–146.

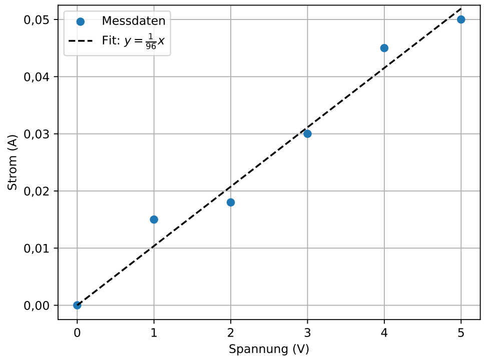
[3] Nagel, C. (2021). Sicher ist sicher! Fachliche Klährung für die didaktische Rekonstruktion von Messunsicherheiten im Unterricht. *Plus Lucis*, (4), 7–11.

3.3 Graphen [★]

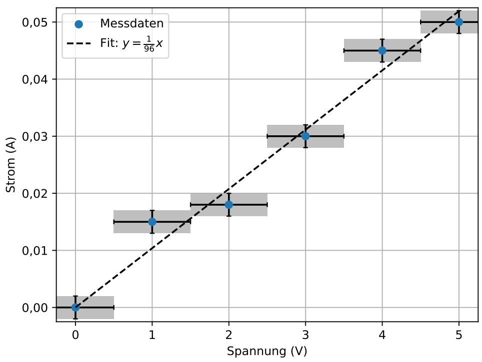
Manchmal erfolgt die Bestimmung einer Größe nicht durch wiederholte Messungen, sondern durch die Betrachtung der Abhängigkeit zwischen zwei anderen Variablen. In diesem Fall können die Daten mit Hilfe eines Diagramms und einer Trendlinie oder einer Fitfunktion ausgewertet werden.



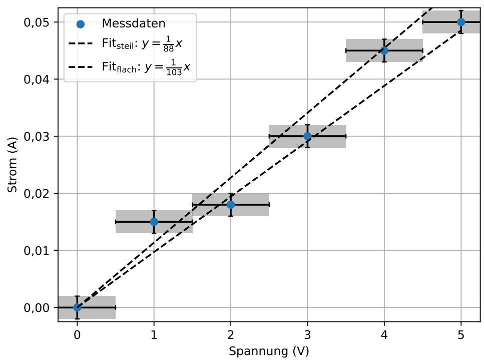
(a) Ein (U, I)-Diagramm von einigen Messdaten.



(b) Das (U, I)-Diagramm mit einem linearen Fit, welcher durch den Ursprung gezwungen wird. Die Steigung des Fits ist gleich $1/R$. Es kann kein Schluss auf die Angemessenheit der linearen Passung bzgl. der Messdaten gezogen werden.



(c) Dasselbe (U, I)-Diagramm, aber jetzt mit Messunsicherheiten (durch Kästchen gekennzeichnet). Diese Darstellung zeigt, dass die Anpassung mit allen Messungen übereinstimmt.



(d) Das gleiche (U, I)-Diagramm, aber jetzt mit maximaler und minimaler Steigung, die durch alle Unsicherheitsfelder gehen. Diese Steigungen können verwendet werden, um die Unsicherheit der Steigung zu bestimmen.

Abbildung 8: Verschiedene Darstellungen von Messdaten der Spannung am und des Stroms durch einen Widerstand.

Bestimmung der Unsicherheit einer Fitfunktion [★ 📕]

Angenommen, man möchte den Widerstand eines Widerstands R bestimmen. Dazu kann man die Spannung U variieren und messen und den resultierenden Strom I messen. Die verschiedenen Messungen können in einem (U, I)-Diagramm aufgetragen werden, siehe Abb. 8a. Eine lineare **Trendlinie (Fit)** der Form $y = ax + b$ kann an die Daten angepasst werden, siehe Abb. 8b. Die meisten Computerprogramme (Excel, Qti-Plot, Google Sheets, ...) verwenden die Methode der kleinsten Quadrate. Das Ergebnis dieses Verfahrens liefert die Steigung a und (wenn gewünscht) den y-Achsenabschnitt b.

Unter Verwendung des Ohm'schen Gesetzes ($I = U/R$) kann man den Widerstand anhand der Steigung a berechnen, die gleich $1/R$ ist. In diesem Fall wird der Widerstand zu $96{,}3\,\Omega$ berechnet.

Dieses Diagramm zeigt jedoch nicht, ob der Fit wirklich zu den Daten „passt“. Dies lässt sich nur feststellen, wenn man die Messunsicherheit betrachtet, siehe Abb. 8c. Diese Abbildung zeigt, dass die Anpassung durch alle Unsicherheitsfelder (die grau schattierten Bereiche, die durch die Unsicherheit der Spannung und des Stroms aufgespannt werden) geht und somit eine gute Anpassung an die Daten darstellt. An diesem Punkt könnte man sagen, dass der Widerstand des Widerstands einen Bestwert von $R = 96{,}3\,\Omega$ hat.

Um die Messunsicherheit der Fitfunktion zu bestimmen, kann man auf die Methode der kleinsten Quadrate zurückgreifen und diese berechnen. Dies würde $u_R = 3{,}7\,\Omega$ ergeben, also $R = (96{,}3 \pm 3{,}7)\,\Omega$. Die genaue Berechnung würde jedoch den Rahmen dieser Einheit sprengen.

Alternative grafische Unsicherheitsermittlung [★]

Es gibt eine vereinfachte, grafische Methode zur Bestimmung der Unsicherheit. Dazu zeichnet man die steilste und die flachste Fitfunktion, die noch durch alle Unsicherheitsfelder gehen und bestimmt deren Steigungen, siehe Abb. 8d. Daraus ergibt sich eine Steigung mit Widerständen von $R = 88\,\Omega$ und $R = 103\,\Omega$. Unter Verwendung desselben Verfahrens, das für die maximale Abweichung verwendet wurde (siehe Gl. (9)), ergibt sich eine Messunsicherheit von $u_R = 8\,\Omega$, sodass $R = (96 \pm 8)\,\Omega$.

In diesem Beispiel wird die Messunsicherheit durch diese Methode eindeutig überschätzt. In anderen Fällen, wenn die Messunsicherheiten pro Datenpunkt klein sind, liefert diese Methode zu kleine Messunsicherheiten. Manchmal scheint die Methode überhaupt nicht zu funktionieren, da es keine Linie gibt, die durch alle Unsicherheitsfelder geht. In diesen Fällen ist es wichtig, sich daran zu erinnern, dass auch die Methode der kleinsten Quadrate nicht durch alle Unsicherheitsfelder geht, aber dennoch ein Ergebnis liefert. Man sollte eine Linie zeichnen, die so gut wie möglich durch die Unsicherheitsboxen geht. Die steilsten und flachsten Steigungen sollten auf die gleiche Weise gezeichnet werden.

Ein weiterer Vorteil dieser Methode ist, dass keine komplexen statistischen Methoden erforderlich sind und sie auch mit Papier und Bleistift durchgeführt werden können.

Anpassen eines Modells [★]

Ein praktisches Beispiel für ein kostengünstiges Schülerexperiment, bei dem zwei Modellfunktionen mit Daten verglichen werden, finden Sie hier: Kok und Boczianowski [1]. Die beiden Modelle entsprechen zwei Funktionen und im Experiment wird geprüft, ob die Daten zu einer dieser Funktionen passen. Diese Anpassung würde zeigen, welches der beiden Modelle die Daten am besten beschreibt.

Literatur

[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 Unsicherheitsfortpflanzung [★]

Manchmal kann die Messgröße nicht direkt gemessen werden, sondern muss aus anderen Messgrößen oder Referenzwerten berechnet werden. In diesen Fällen gibt es eine Gleichung, die beschreibt, wie die Eingangsgrößen mit der Ausgangsgröße zusammenhängen: die Messgleichung (auch: Messmodel). Um die Unsicherheit der Ausgangsgröße zu bestimmen, müssen die Unsicherheiten der Eingangsgrößen korrekt fortgepflanzt werden.

Regeln für Unsicherheitsfortpflanzung [★]

Betrachte folgendes Beispiel: Man möchte die Durchschnittsgeschwindigkeit eines Läufers bestimmen. Dazu misst man eine bestimmte Strecke und anschließend die Zeit, die der Läufer für diese Strecke benötigt. Die Messgleichung für die Geschwindigkeit ist die Strecke dividiert durch die Zeit: $v = \frac{s}{t}$.

Strecke und Zeit sind jeweils mit einer Unsicherheit behaftet, die zusammen eine Unsicherheit für die Geschwindigkeit ergeben. Diese Unsicherheit muss durch **Fortpflanzung** der Unsicherheiten der Strecke und der Zeit berechnet werden. Die Regeln für die Fortpflanzung der Unsicherheiten für einige gängige Funktionen sind in Tab. 4 dargestellt. In diesem Fall ist die Unsicherheit für die Geschwindigkeit gegeben durch: $u_v = |v|\sqrt{\left(\frac{u_s}{s}\right)^2 + \left(\frac{u_t}{t}\right)^2}$.

Das Messergebnis der Strecke ist $s = (50{,}00 \pm 0{,}50)\,\mathrm{m}$ und das Ergebnis der Zeitmessung ist $t = (20{,}0 \pm 1{,}0)\,\mathrm{s}$. Daraus ergibt sich eine Geschwindigkeit von $v = (2{,}50 \pm 0{,}13)\,\mathrm{m/s}$.

Tabelle 4: Die Regeln für die Fortpflanzung von Unsicherheiten für zwei unabhängige Variablen x und y. Die Tabelle zeigt, wie die Unsicherheiten mit Hilfe der Gauß'schen Unsicherheitsfortpflanzung und den angenäherten Verfahren fortgepflanzt werden. Hier sind a und b Zahlen ohne Unsicherheit, σ ist die Standardabweichung der Verteilung und u ist die (angenäherte) Unsicherheit.

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

Annäherung der Fortpflanzung [★ ✎]

Die rechte Spalte von Tab. 4 zeigt auch einige Näherungswerte für die Bestimmung der Unsicherheit [Näherungswerte aus 1, S. 275–277].

Für die Geschwindigkeitsmessung von vorher kann die Unsicherheit wie folgt berechnet werden:

$$\begin{aligned} u_z &= |z|\left(\frac{u_x}{|x|} + \frac{u_y}{|y|}\right) \\ &= |v|\left(\frac{u_s}{|s|} + \frac{u_t}{|t|}\right) \\ &= 2{,}50\,\mathrm{m/s}\left(\frac{0{,}50\,\mathrm{m}}{50{,}00\,\mathrm{m}} + \frac{1{,}0\,\mathrm{s}}{20{,}0\,\mathrm{s}}\right) = 0{,}15\,\mathrm{m/s}. \end{aligned}$$

Daraus ergibt sich ein Messergebnis von $v = (2{,}50 \pm 0{,}15)\,\mathrm{m/s}$. Dies stellt eine leichte Überschätzung der Unsicherheit dar.

Für einige Funktionen gibt es keine Standardnäherung. Es gibt jedoch eine noch einfachere Annäherung für die Fortpflanzung der Unsicherheit, die immer verwendet werden kann. In diesen Fällen wird die Unsicherheit der berechneten Größe bestimmt, indem die Werte der Größen und ihre jeweiligen Unsicherheiten kombiniert werden, um das kleinste und das größte mögliche Ergebnis zu berechnen; dieser Bereich wird dann als Unsicherheit verwendet.

Für die Geschwindigkeitsmessung wäre dies das folgende Verfahren. Die Geschwindigkeit wird berechnet als: $v = \frac{s}{t} = \frac{50{,}00\,\mathrm{m}}{20{,}0\,\mathrm{s}} = 2{,}50\,\mathrm{m/s}$.

Der größtmögliche Wert ergibt sich aus der Kombination der Messungen und ihrer Unsicherheiten wie folgt:

$$\begin{aligned} v &= \frac{s + u_s}{t - u_t} \\ &= \frac{50{,}5\,\mathrm{m}}{19\,\mathrm{s}} = 2{,}66\,\mathrm{m/s}. \end{aligned}$$

Für den kleinstmöglichen Wert:

$$\begin{aligned} v &= \frac{s - u_s}{t + u_t} \\ &= \frac{49{,}5\,\mathrm{m}}{21\,\mathrm{s}} = 2{,}36\,\mathrm{m/s}. \end{aligned}$$

Die Unsicherheit wird nun nach der Methode der maximalen Abweichung bestimmt (siehe Gl. (9)): Die beiden Abweichungen sind: $|2{,}50\,\mathrm{m/s} - 2{,}66\,\mathrm{m/s}| = 0{,}16\,\mathrm{m/s}$ und $|2{,}50\,\mathrm{m/s} - 2{,}36\,\mathrm{m/s}| = 0{,}14\,\mathrm{m/s}$. Die größere der beiden Abweichungen wird konservativ als Unsicherheit gewählt. Das Ergebnis ist nun $v = (2{,}50 \pm 0{,}16)\,\mathrm{m/s}$, was etwas größer ist als die vorherige Näherung.

Online Rechner für Unsicherheiten [★ 💡]

Es gibt Hinweise darauf, dass die manuelle Berechnung der Unsicherheit für die Lernenden von Vorteil ist [2]. Es gibt Fälle, in denen diese Berechnungsroutine von der Analyse der Daten und der Interpretation der Ergebnisse ablenkt [3]. Viele Unsicherheitsrechner sind online verfügbar, einer, der recht intuitiv funktioniert, ist hier zu finden: https://nicoco007.github.io/Propagation-of-Uncertainty-Calculator.

Literatur

[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 Unsicherheit [★]

Manchmal möchte man die Präzision der Messung verbessern. Im Beispiel der Geschwindigkeitsmessung von vorhin beeinflussen die Messunsicherheit der Entfernungs- und der Zeitmessung die Unsicherheit der Messung. Um herauszufinden, welche Änderung sich am stärksten auf die Unsicherheit der Geschwindigkeit auswirkt, kann die **relative Unsicherheit** oder die prozentuale Unsicherheit berechnet werden, siehe Tab. 5. Dazu wird zunächst eine Übersicht über alle Eingangs- und Ausgangsgrößen, ihren Wert und ihre Unsicherheit erstellt, aus der die prozentuale Unsicherheit berechnet werden kann.

Betrachtet man die relativen Unsicherheiten und wie sie sich fortpflanzen, kann man schnell die Größe identifizieren, die den größten Einfluss auf die Unsicherheit der berechneten Größe hat. In diesem Fall wird eine Verringerung der Unsicherheit der Zeitmessung (5 % relative Unsicherheit) die Unsicherheit der Geschwindigkeit viel stärker verringern als eine Messung der Entfernung mit höherer Genauigkeit (1 % relative Unsicherheit).

Weitere Lektüre [★ 💡]

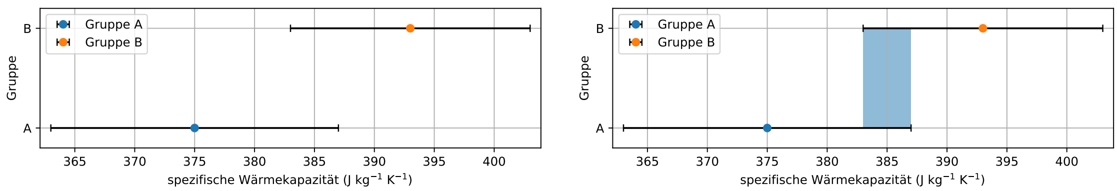
Manchmal kann es sinnvoll sein, die Unsicherheit *vor* einem Experiment abzuschätzen. Die Festlegung dieses Unsicherheitsziels hat Auswirkungen auf die Versuchsplanung: Priemer [1].

Tabelle 5: Alle Größen, ihr Betrag, ihre Unsicherheit und die relative Unsicherheit für eine Geschwindigkeitsmessung.

Größe	Betrag	Unsicherheit	Relative Unsicherheit
Eingangsgrößen:			
Weg	50,00 m	0,50 m	1 %
Zeit	20,0 s	1,0 s	5 %
Ausgangsgrößen:			
Geschwindigkeit	2,50 m/s	0,13 m/s	5,2 %

Literatur

[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) Die Messergebnisse der Gruppen A und B mit ihren jeweiligen Unsicherheitsbalken.

(b) Der Überlappungsbereich der Unsicherheitsintervalle ist blau markiert.

Abbildung 9: Die Messergebnisse der spezifischen Wärmekapazität, gemessen von den Gruppen A und B.

4 ERGEBNISSE VERGLEICHEN

Bisher lag der Schwerpunkt auf den Quellen der Messunsicherheit, ihrer Auswirkung auf die Daten, ihrer Quantifizierung und ihrer Fortpflanzung. In diesem Abschnitt geht es um die Anwendung der Messunsicherheit.

Messunsicherheiten sind keine Belastung, die am Ende eines Experiments berechnet werden müssen—zumindest sollten sie es nicht sein! Messunsicherheiten sind nützlich, weil sie den Vergleich von Messergebnissen ermöglichen. Es wäre Zufall, wenn zwei Mittelwerte von zwei verschiedenen Messreihen genau gleich sind. Das bedeutet aber nicht, dass zwei Ergebnisse nur dann miteinander verträglich sind. Für einen korrekten Vergleich müssen die Unsicherheitsintervalle berücksichtigt werden.

Dieser Teil behandelt einige praktische Regeln für den Vergleich von Messergebnissen im Allgemeinen und wie Lernende dabei vorgehen können.

4.1 Einfacher Vergleich [ ]

Der Vergleich von zwei Messergebnissen ist Praxis in der Wissenschaft. Hierbei kann nicht ohne Berücksichtigung der Messunsicherheit vorgegangen werden.

Eine Möglichkeit, zwei Mittelwerte zu vergleichen, besteht in der Berechnung eines t-Tests. Das mathematische Verfahren dieser Test würde aber den Rahmen dieser Einheit sprengen und für den Unterricht ungeeignet. Der Test vergleicht jedoch die beiden Mittelwerte, berücksichtigt ihre jeweilige Varianz sowie die Anzahl der wiederholten Messungen und bestimmt den Grad der Überschneidung zwischen ihnen.

Vergleich von Unsicherheitsintervallen [ ]

Für den Vergleich von Datensätzen oder Messergebnissen im schulischen Kontext ist ein t-Test oftmals zu kompliziert. Daher muss die Vergleichsmethode vereinfacht werden. Dies kann durch den Vergleich von Unsicherheitsintervallen geschehen (siehe Gl. (8) in Abschnitt 2.7). Eine Faustregel besagt, dass sich überlappende Unsicherheitsintervalle auf verträgliche Ergebnisse hinweisen.

Angenommen zwei Gruppen haben die spezifische Wärmekapazität von zwei gleich aussehenden Metallobjekten gemessen. Gruppe A hat $c_A = (375 \pm 12)\,\mathrm{J\,kg^{-1}\,K^{-1}}$ gemessen und Gruppe B $c_B = (393 \pm 10)\,\mathrm{J\,kg^{-1}\,K^{-1}}$, siehe Abb. 9a. Die Gruppen fragen sich nun: Könnte das Metall der beiden Objekte das gleiche sein, also die gleiche spezifische Wärmekapazität haben? Das Unsicherheitsintervall gibt den Bereich an, in dem die Messgröße zu erwarten ist. Die Überlappung der Unsicherheitsintervalle deutet also auf einen gemeinsamen Bereich möglicher spezifischer Wärmekapazitäten hin, die beiden Objekten zugeordnet werden können, siehe Abb. 9b. Daher sind die Ergebnisse verträglich und die Gruppen müssen die Möglichkeit berücksichtigen, dass die spezifische Wärmekapazität der beiden Objekte gleich ist.

Das bedeutet nicht, dass sie bewiesen haben, dass die spezifische Wärmekapazität dieselbe ist! Sie haben einen Hinweis für die Möglichkeit, dass sie dieselbe sein könnte. Je nach Grad der Überschneidung und dem Bereich des Unsicherheitsintervalls kann diese Möglichkeit bewertet werden.

Tabelle 6: Experimentelle Ergebnisse bei der Bestimmung der Erdbeschleunigung. Das Messergebnis A hat zwar die geringste Unsicherheit, ist aber nicht verträglich mit dem Referenzwert.

Referenzwert:	
g_{Ref}	$9{,}81\,\mathrm{m/s^2}$
Messergebnisse:	
g_{A}	$(9{,}83 \pm 0{,}01)\,\mathrm{m/s^2}$
g_{B}	$(9{,}78 \pm 0{,}05)\,\mathrm{m/s^2}$
g_{C}	$(9{,}83 \pm 0{,}15)\,\mathrm{m/s^2}$

Alltagsvorstellungen zum Vergleich von Daten [ ]

Auf die Frage nach der Verträglichkeit zwischen einem Messergebnis und einem Referenzwert verweisen einige Lernende auf die (prozentuale) Differenz. Dies ist zwar ein Aspekt der Qualität eines Ergebnisses, aber kein Hinweis auf die Verträglichkeit. Dies kann den Lernenden anhand der Ergebnisse in Tab. 6 verdeutlicht werden.

Obwohl die Gruppen A und C beide eine Differenz von $0{,}02\,\mathrm{m/s^2}$ zum Referenzwert aufweisen, ist das Ergebnis der Gruppe A nicht verträglich mit dem Referenzwert. Die Gruppen B und C sind beide mit dem Referenzwert verträglich. Und obwohl die Differenz zum Referenzwert bei Gruppe B größer ist als bei Gruppe C, ist ihre Unsicherheit dreimal kleiner.

Weitere Lektüre [ ]

Für eine didaktische Reduktion des *t*-Tests siehe: [1].

Ein Online-Rechner für den *t*-Test findet sich hier: https://www.graphpad.com/quickcalcs/ttest1.cfm.

Weitere Erklärungen zur Statistik des *t*-Tests findet man bei Wikipedia: https://en.wikipedia.org/wiki/Student's_t-test.

Literatur

[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 Indikatoren [ ]

Der Vergleich von Messergebnissen ist auch ein guter Test, um das Verständnis der Lernenden über Messunsicherheiten zu überprüfen. Solche Datenvergleichsaufgaben sollten nicht auf eine Multiple-Choice-Frage (stimme zu/ stimme nicht zu) reduziert werden. Vielmehr sollte dies durch eine Begründung unterstützt werden.

Wenn man sich die Begründungen der Lernenden bei einer Datenvergleichsaufgabe ansieht, bestehen diese meistens aus einer bestimmten **Vergleichsmaß**, die verglichen wird, gefolgt von der Überprüfung eines bestimmten **Vergleichskriteriums**, das erfüllt (oder nicht erfüllt) ist. Die Betrachtung des Vergleichsmaßes und des Vergleichskriteriums, die von den Lernenden in ihren Begründungen genannt werden, gibt Aufschluss über das konzeptuelle Verständnis der Lernenden [1].

Für Lernende, die Datensätze vergleichen, wurden die folgenden Vergleichsmaße identifiziert

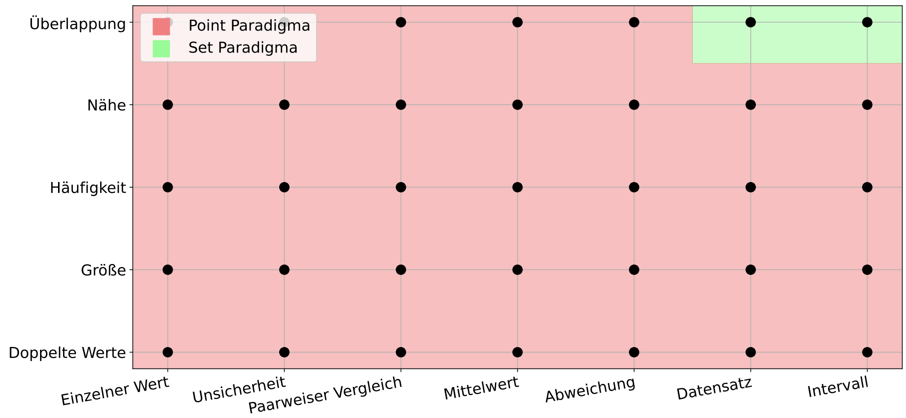


Abbildung 10: Die Kombination aus der zu vergleichenden Größe und dem Kriterium, das in einer Begründung geprüft wird, kann mit dem Point-Paradigma (rot) und dem Set-Paradigma (grün) in Verbindung gebracht werden.

Intervall: Die Lernenden vergleichen das gesamte Unsicherheitsintervall entweder als Bereich oder als Bestwert ± Unsicherheit.

Datensatz: Die Lernenden vergleichen die Datensätze als Ganzes.

Abweichung: Die Lernenden subtrahieren zwei Datensätze und vergleichen die resultierenden Abweichungen.

Mittelwert: Die Lernenden vergleichen den Mittelwert.

Paarweiser Vergleich: Die Lernenden vergleichen die Ergebnisse jeweils paarweise.

Unsicherheit: Die Lernenden vergleichen nur die Unsicherheit.

Einzelner Wert: Die Lernenden vergleichen einzelne, isolierte Messungen (z. B. einen wiederkehrenden Wert oder einen Extremwert).

Die folgenden Vergleichskriterien wurden identifiziert:

Überlappung: Lernende suchen nach Überschneidungen (oder dem Fehlen solcher) in ihren Datensätzen oder Messergebnissen.

Nähe: Lernende untersuchen, wie nahe oder ähnlich sich zwei Datensätze oder Messergebnisse sind.

Häufigkeit: Die Lernenden zählen das Auftreten bestimmter Ereignisse, z. B. wenn ein Wert größer/kleiner als der andere ist.

Größe: Die Lernenden vergleichen Größen und prüfen, ob die eine größer oder kleiner ist als die andere.

Doppelte Werte: Die Lernenden suchen nach doppelten oder wiederkehrenden Werten.

Die Kombination dieses Vergleichsmaßes und des Vergleichskriteriums kann dem Point- und dem Set-Paradigma zugeordnet werden, siehe Abb. 10. Liegt die Kombination im roten Bereich, ist die Begründung dem Point-Paradigma zuzuordnen, liegt sie im grünen Bereich, kann die Begründung dem Set-Paradigma zugeordnet werden (zu Point- und Set-Paradigmen siehe Abschnitt 2.1).

Typische Antworten der Lernenden [ ]

Angenommen, zwei Gruppen A und B messen die Fallzeit eines Objekts. Die Daten sind in Tab. 7 angegeben. Anhand der Daten soll die Frage beantwortet werden, ob die Fallzeiten gleich sind.

Eine typische Antwort der Lernenden könnte lauten:

> *"Die längste Fallzeit* 1,583 s *tritt in Gruppe A auf. Daher muss die Fallzeit des*

Tabelle 7: Zwei Datensätze, die auf Verträglichkeit verglichen werden können. Die untere Zeile zeigt den Mittelwert.

Fallzeit Gruppe A (s)	Fallzeit Gruppe 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

> *Objekts der Gruppe A größer sein."*

Diese Person sucht nach dem größten (Vergleichskriterium: Größe) Wert in der Reihe (Vergleichsmaß: Einzelner Wert) und schließt daraus, dass die Gruppe, zu der dieser Wert gehört, die längste Fallzeit haben muss. Diese Person hat einen einzelnen Wert betrachtet, und das Kriterium ist, dass dieser Wert größer als die anderen Werte in der Reihe ist. Dieser einzelne Messwert bestimmt die Schlussfolgerung des Experiments. Dies steht in engem Zusammenhang mit dem Point-Paradigma, bei dem einzelne Werte repräsentativ für eine ganze Reihe sind.

Um diese Lernenden zu unterstützen, könnte man sie zunächst auffordern, darüber nachzudenken, warum eine Messreihe durchgeführt wird und was die Ursache für Schwankungen ist. Auf diese Weise sollten die Lernenden erkennen, dass die Quellen der Messunsicherheit die Ergebnisse beider Gruppen beeinflussen. Dann kann man den Lernenden helfen, indem man sie die Streuung der gesamten Messreihe betrachten lässt.

Eine andere typische Antwort der Lernenden könnte lauten:

> *'Die Fallzeit des Objekts der Gruppe A ist länger, weil die Fallzeit der Gruppe A in der Regel länger ist als die der Gruppe B."*

Diese Person hat die Messungen paarweise verglichen (Vergleichsmaß: paarweiser Vergleich) und geschaut, welcher der beiden Werte größer oder kleiner als der andere ist (Vergleichskriterium: Häufigkeit). Dies ist elaborierter als die vorherige Antwort, da alle Werte berücksichtigt werden. Der isolierte Vergleich von Messwertpaaren ist jedoch auch ein anschauliches Beispiel für Point-Paradigma-Schlussfolgerungen.

Es gibt mehrere Strategien, die den Lernenden helfen, ihre Überlegungen zu reflektieren. Man könnte die gleichen Messungen in einer anderen Reihenfolge anordnen und die Lernenden fragen, was ihre Schlussfolgerung ist. Die Lernenden werden schnell erkennen, dass sich ihre Schlussfolgerung mit ihrer Argumentation ändern muss. Oft reicht dies aus, damit die Lernenden ihre fehlerhafte Argumentation erkennen. Die Lernenden sollten dann angeleitet werden, die Messreihe als Ganzes zu betrachten. Eine hilfreiche Strategie ist es, die Ergebnisse in einem Diagramm darzustellen. Dies hilft ihnen, die Streuung des Datensatzes als Ganzes zu erkennen.

Eine weitere Antwort könnte lauten:

> *"Die beiden Mittelwerte,* 1,549 s *und* 1,537 s*, liegen sehr nahe beieinander, und daher können die beiden Zeiten als gleich angesehen werden."*

Aus dieser Antwort wird deutlich, dass diese Person Mittelwerte verglichen hat (Vergleichsmaß: Mittelwert) und dass das Kriterium für die Übereinstimmung darin besteht, dass die Werte nahe genug beieinander liegen (Vergleichskriterium: Nähe). Obwohl der Vergleich des Mittelwerts als Set-Paradigma-Argumentation betrachtet werden könnte, tun Lernende dies oft als automatisierte Routine. Der Mittelwert selbst ist dann der einzige Wert, der berücksichtigt wird, sodass es sich um Point-Paradigma-Argumentation handelt. Die Annäherung der beiden Werte wird ebenfalls als Point-Paradigma-Argumentation betrachtet. Dies liegt daran, dass die Messunsicherheit berücksichtigt werden muss, damit eine Bewertung dieser Nähe hinsichtlich der Verträglichkeit erfolgen kann.

Um die Lernenden zu unterstützen, könnten Leitfragen in Betracht gezogen werden. Zum Beispiel könnte man sie fragen, was sie unter "nah" verstehen und was nah genug für die konkret

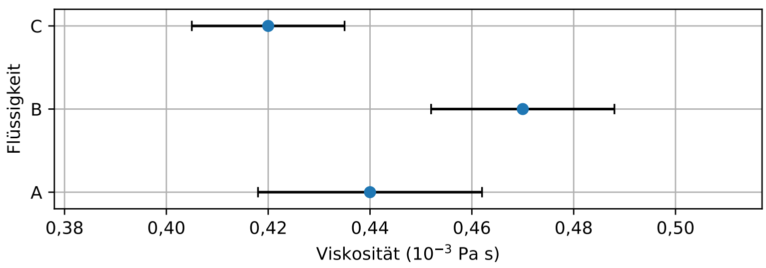


Abbildung 11: Die Messergebnisse für die Viskosität von drei Flüssigkeiten.

bedeutet. Solange die Summe der Unsicherheiten beider Messgrößen kleiner ist als die Differenz zwischen den Mittelwerten, können die Ergebnisse immer noch als unverträglich angesehen werden.

Eine weitere Frage ist, welche Rolle der Mittelwert für die Lernenden spielt. Ist er ein Schätzwert für die Zentralität der Messreihe (Set-Paradigma-Argumentation) oder betrachten die Lernenden ihn als 'Endergebnis' des Experiments (Point-Paradigma-Argumentation)? Wenn letzteres der Fall ist, könnte anschließend gefragt werden, was mit dem Mittelwert passieren würde, wenn eine weitere Messung zur Reihe hinzugefügt würde. Ein sich ändernder 'wahrer Wert' ist aus Sicht des Point-Paradigmas problematisch, aber ein sich ändernder Bestwert innerhalb des Unsicherheitsintervalls ist aus Sicht des Set-Paradigmas unproblematisch.

Literatur

[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 Vergleich mehrerer Ergebnisse [★]

Es gibt Fälle, in denen mehr als zwei Datensätze oder Messergebnisse verglichen werden. Hier werden zwei Fälle beschrieben, in denen die Ergebnisse von mehr als zwei Messergebnissen verglichen und die Beziehung zwischen zwei Größen anhand von Graphen analysiert werden (siehe auch 3.3).

Vergleich von drei Messergebnissen [★ 📘]

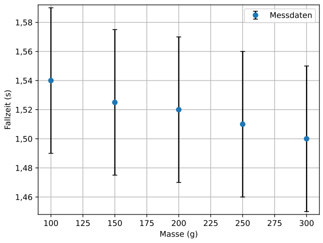
In Abschnitt 4.1 wurde der Prozess des Vergleichs zweier Datensätze oder Messergebnisse durchgeführt. Man kann dasselbe Verfahren auch anwenden, um die Ergebnisse von drei Messungen zu vergleichen. Allerdings muss man bei der Interpretation der Ergebnisse sehr vorsichtig sein.

Angenommen, die Viskosität von drei verschiedenen Flüssigkeiten A, B und C wird gemessen (siehe Abb. 11). Wenn man die Viskosität der Flüssigkeiten vergleicht, muss man zu dem Schluss kommen, dass die Viskosität von Flüssigkeit A sowohl mit C als auch mit B verträglich ist. Man kann jedoch nicht zu dem Schluss kommen, dass das Ergebnis von Flüssigkeit B mit Flüssigkeit C verträglich ist—obwohl beide mit Flüssigkeit A verträglich sind.

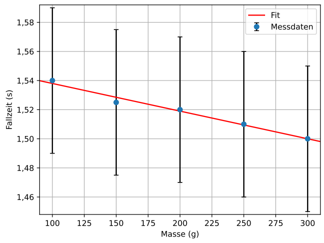
Darüber hinaus kann aus den Ergebnissen geschlossen werden, dass die Viskosität von Flüssigkeit C niedriger ist als die Viskosität von Flüssigkeit B. Für die Flüssigkeit A beschränkt sich die Schlussfolgerung darauf, dass ihre Viskosität sowohl mit der von B als auch mit der von C verträglich ist. Will man eine ausführlichere Aussage über die Viskosität der Flüssigkeit A treffen, müssen die Unsicherheiten verringert werden.

Analyse von Fitfunktionen [★ 📘]

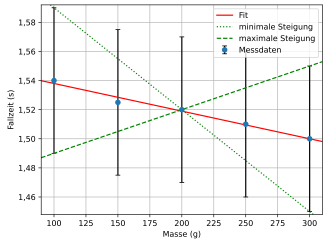
Oft werden mehrere Messungen durchgeführt, um die Zusammenhang zwischen zwei Größen zu untersuchen. Bei der Datenanalyse wird beurteilt, ob die beiden Größen einem bestimmten Modell oder einer bestimmten Gleichung entsprechen. Diese Daten können in tabellarischer Form



(a) Fünf Zeitmessungen der Fallzeit von Kugeln mit unterschiedlichen Massen.



(b) Die rote Linie zeigt eine lineare Fitfunktion für die Daten.



(c) Die grün gestrichelten Linien zeigen die minimale und maximale Steigung der Fitfunktion an.

Abbildung 12: Die Messergebnisse eines Experiments, bei dem Kugeln mit unterschiedlicher Masse aus einer bestimmten Höhe fallen gelassen werden.

ausgewertet werden, aber gegebenenfalls ist es einfacher, die Ergebnisse in grafischer Form zu interpretieren [1, 2]. Im Abschnitt 3.3 wurde das Verfahren zum Zeichnen von Anpassungsfunktionen und deren Unsicherheiten beschrieben. In diesem Abschnitt geht es um die Interpretation dieser Anpassungsfunktionen.

Um die Frage zu beantworten, ob schwerere Objekte schneller fallen als leichtere, könnte man ein Experiment durchführen. In Abb. 12a sind die Fallzeiten von Kugeln unterschiedlicher Masse dargestellt. Die Frage ist nun, ob die Masse die Fallzeit beeinflusst oder nicht.

Auf den ersten Blick scheinen die Daten eine Abnahme der Fallzeit mit zunehmender Masse zu zeigen. Dies geht auch aus der Fitfunktion hervor, siehe Abb. 12b. Betrachtet man jedoch die maximale und minimale Steigung, siehe Abb. 12c, so wird deutlich, dass sowohl positive als auch negative Steigungen mit den Daten übereinstimmen. Die Schlussfolgerung ist, dass aufgrund der Möglichkeit positiver und negativer Steigungen kein Einfluss der Masse auf die Fallzeit angenommen werden kann. Es ist zu beachten, dass diese Schlussfolgerung erst nach der Unsicherheitsanalyse gezogen werden konnte.

Wenn man den Einfluss der Reibung auf kleinere Massen zeigen will, was zu längeren Fallzeiten für kleinere Massen führen würde, müssten die Unsicherheiten drastisch reduziert werden.

Wenn man sich für die thermische Ausdehnung von Wasser interessiert, kann man die Abhängigkeit der Volumenänderung von der Temperatur untersuchen. Dazu kann man ein Experiment durchführen, bei dem die Höhe einer Wassersäule bei verschiedenen Temperaturen gemessen wird, siehe Abb. 13a. Die Anfangshöhe beträgt 0 mm bei einer Anfangstemperatur von 10 °C und die Temperatur wird langsam in Schritten von 2 °C bis zu einer Temperatur von 20 °C erhöht, siehe Abb. 13b.

In einem ersten Schritt wird eine lineare Fitfunktion erstellt, siehe Abb. 13c. Diese Fitfunktion geht durch alle Unsicherheitsbalken und passt daher zu den Daten. Obwohl eine quadratische Funktion besser zu den Daten passt, siehe Abb. 13d, kann eine lineare Abhängigkeit nicht ausgeschlossen werden. Wenn man das Experiment verfeinert und die Unsicherheiten verringert, siehe Abb. 13e, ist man schließlich in der Lage, eine lineare Abhängigkeit zu falsifizieren und eine quadratische Abhängigkeit wahrscheinlicher zu machen.

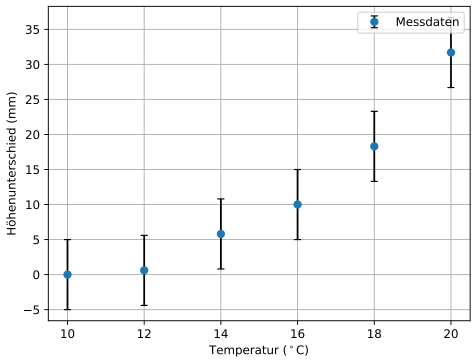
Weitere Lektüre [💡]

Eine ausführlichere Beschreibung und Analyse der thermischen Ausdehnung von Wasser findet sich hier: Wagner et al. [3].

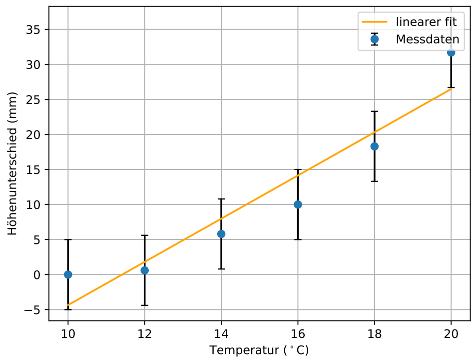
Literatur

[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

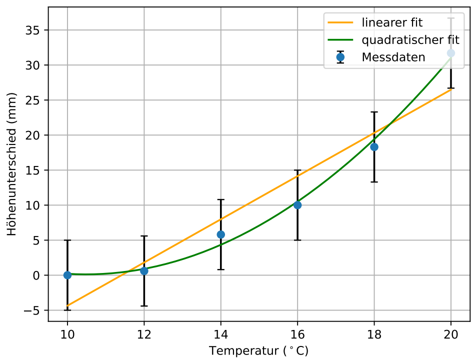
(a) Messung der Höhe der Wassersäule bei verschiedenen Temperaturen.



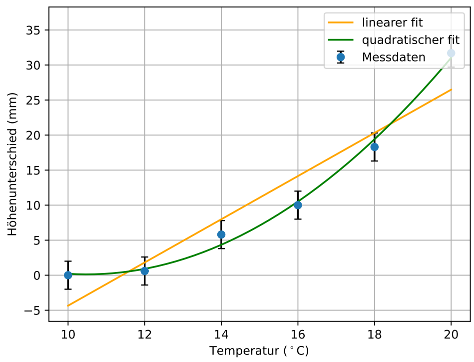
(b) Messergebnisse für die Ausdehnung von Wasser.



(c) Eine lineare Fitfunktion für die Daten.



(d) Eine quadratische Fitfunktion für die Daten.



(e) Präzisere Messergebnisse.

Abbildung 13: Einige experimentelle Daten zur Ausdehnung eines Wasservolumens bei verschiedenen Temperaturen.

[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