

MASTERARBEIT | MASTER'S THESIS

Titel | Title

P-prims bei Schüler:innenvorstellungen zum Thema Zufall im
Kontext von Messunsicherheiten

verfasst von | submitted by
Jakob Wanko BEd

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Education (MEd)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 199 520 523 02

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Lehramt Sek (AB) Unterrichtsfach
Mathematik Unterrichtsfach Physik

Betreut von | Supervisor:

Univ.-Prof. Dr. Martin Hopf

Mitbetreut von | Co-Supervisor:

Mag. Dr. Clemens Nagel

Danksagung

Ein herzliches Dankeschön an...

... meinem Betreuer *Clemens Nagel*, der mich während der gesamten Arbeit stets mit großem Engagement unterstützt hat. Jede Besprechung mit ihm war bereichernd – voller neuer Ideen, Impulse und hilfreicher Perspektiven. Besonders dankbar bin ich ihm auch für die Vermittlung des Kontakts zu den vier Bachelorstudent:innen.

... *Anna Fleissner, Julia Hartmann, Christian Lechmann und Hannah Schaubmaier*, die im Zuge ihrer Bachelorarbeit die Interviewstudien durchgeführt, transkribiert und ausgewertet haben – ohne euch wäre diese Arbeit im jetzigen Ausmaß niemals möglich gewesen.

... meine Ehefrau *Susi*, die mir während des gesamten Studiums – und insbesondere in der Entstehung dieser Arbeit – mit unermüdlicher Unterstützung, Geduld und einem offenen Ohr zur Seite stand.

... meinen Eltern *Martina und Alfred*, die mir mein Studium durch ihre Unterstützung ermöglicht haben.

... meinem Bruder *Simon*, für seine Ratschläge und die angeregten Diskussionen, die mir immer wieder neue Perspektiven eröffnet haben.

Abstract

Diese Arbeit beschäftigt sich mit den Vorstellungen von Schüler:innen hinsichtlich Messunsicherheiten im Zusammenhang mit zufälligen Ereignissen und Messfehlern. Messunsicherheiten sind ein zentrales Thema in den Naturwissenschaften, insbesondere im Physikunterricht. Ziel der Untersuchung war es, zu erfassen, welche intuitiven Konzepte – sogenannte „p-prims“ – Schüler:innen bei der Analyse von Messungen und der Bewertung von Unsicherheiten zugrunde liegen.

Dazu wurde eine umfangreiche Interviewstudie mit leitfadengestützten Gesprächen durchgeführt, in deren Rahmen die Schüler:innen auch ein eigenes Experiment durchführen mussten. Neben diesem Experiment wurde auch ein Video eines Versuches abgespielt und es wurden bereits erstellte Messreihen verglichen. Mithilfe dieser Interviews sollten die Gedankenprozesse der Schüler:innen im Umgang mit Messwerten und deren Streuungen sichtbar gemacht werden.

Die Suche nach neuen p-prims im Kontext zufälliger Ereignisse erwies sich als sprichwörtliche „Suche nach der Stecknadel im Heuhaufen“, da zunächst zahlreiche bereits bekannte Schülervorstellungen bestätigt wurden. Die Ergebnisse zeigen, dass viele Schüler:innen den Zufall in Messprozessen nicht als natürlichen Bestandteil akzeptieren, sondern Messunsicherheiten häufig als Fehler in der Durchführung interpretieren.

Ein weiteres zentrales Ergebnis der Studie ist die Tendenz der Schüler:innen, den Mittelwert einer Messreihe als den „wahren“ Wert anzusehen. Die Aussagen „Die Mitte ist am besten“ und „(digitale) Technik ist fehlerlos“ erweisen sich dabei als aussichtsreiche Kandidaten für p-prims zum Thema Zufall im Kontext von Messunsicherheiten.

Abstract

This work deals with students' perceptions of measurement uncertainties in connection with random events and measurement errors. Measurement uncertainties are a central topic in the natural sciences, especially in physics lessons. The aim of the study was to determine which intuitive concepts - so-called „p-prims“ - students use as a basis when analyzing measurements and evaluating uncertainties.

For this purpose, an extensive interview study with guided discussions was conducted, in the course of which the students also had to conduct their own experiment. In addition to this experiment, a video of an experiment was also played and previously created measurement series were compared. With the help of these interviews, the students' thought processes in dealing with measured values and their scattering were to be made visible.

The search for new p-prims in the context of random events proved to be a proverbial „search for a needle in a haystack“, as numerous already known student ideas were initially confirmed. The results show that many students do not accept randomness in measurement processes as a natural component, but often interpret measurement uncertainties as errors in the implementation.

Another key finding of the study is the tendency of students to regard the mean value of a series of measurements as the „true“ value. The statements „The middle is best“ and „(digital) technology is error-free“ prove to be promising candidates for p-prims on the topic of chance in the context of measurement uncertainties.

Inhaltsverzeichnis

1	Theoretischer Hintergrund der Arbeit	3
1.1	Messunsicherheiten in der Naturwissenschaft	3
1.1.1	Der Unterschied zwischen Fehler und Unsicherheit	3
1.1.2	Bestimmung der Messunsicherheit	4
1.1.3	Typ-A-Messunsicherheit	4
1.1.4	Messunsicherheit nach der Methode Typ B	5
1.1.5	Der wahre Wert	6
1.2	Phenomenological primitives	7
1.2.1	p-prims im Bezug auf Ereignisse im Kontext von Messunsicherheiten	9
1.2.2	p-prims im Bezug auf Zufall	9
1.3	Lernendenvorstellungen	10
1.3.1	Lernendenvorstellungen im Bezug auf Messunsicherheiten	11
1.3.2	Lernendenvorstellungen im Bezug auf zufällige Ereignisse	12
1.4	Wie hängen p-prims und Lernendenvorstellungen zusammen?	13
1.5	Aktueller Forschungsstand	14
2	Methodik der Arbeit	16
2.1	Grundlagen qualitativer Forschung	16
2.2	Leitfadengestütztes Interview	16
2.3	Think aloud Interview	17
2.4	Transkribieren	18
2.5	MAXQDA	18
2.6	Kategorienbildung	19
2.6.1	Deduktive Kategorien	19
2.6.2	Induktive Kategorien	19
2.7	Codierleitfaden	20
3	Qualitätssicherung	21
3.1	Erkenntnistheoretische Qualitätskriterien	21
3.1.1	Objektivität	21
3.1.2	Validität	21
3.1.3	Reliabilität	21
3.2	Forschungsethische Qualitätskriterien	22
3.3	Pragmatische Qualitätskriterien	23
4	Datenauswertung	25
4.1	Definition der Kategorien	25
4.1.1	Genauere Betrachtung des Fallversuch-Videos	34
4.2	Quantitative Analyse der Kategorien	35

4.2.1	Detailanalyse zur Identifikation möglicher neuer p-prims	39
4.3	Interpretation der Ergebnisse	40
5	Diskussion und Ausblick	44
Anhang		49
Transkripte		49

Einleitung, Motivation und Forschungsfrage

Messunsicherheiten sind ein Thema, welches vorwiegend im Studium thematisiert wird. Im Unterricht in der Sekundarstufe 1 und 2 ist es fest im Lehrplan (Bundesminister für Unterricht und Kunst, 1985) verankert. Im Schulalltag ist es allerdings so, dass den Lehrpersonen mit den zugrundeliegenden Konzepten nicht ausreichend vertraut sind, dass sie sich zutrauen diese Thematik schülergerecht aufzubereiten. Somit kommt es vor, dass erwähnt wird, dass es eine Messunsicherheit gibt. Weitere Konzepte wie Messunsicherheiten und die Unterschiede zwischen Unsicherheit und Fehler werden kaum thematisiert. (Nagel, Lux & Steindl, 2021). Typ-A und Typ-B Messunsicherheiten werden kaum bis gar nicht erwähnt, was man auch an der Universität merkt, wenn im Zuge der Ausbildung die Interpretation von Messwerten und die Auswertung von Messdaten eine größere Rolle spielt. Erhaltene Ergebnisse sollen mit einer realistischen Messtoleranz angegeben werden, wofür ein Grundwissen über die Typ-A und Typ-B Messunsicherheit erforderlich ist.

Die Auseinandersetzung mit Messunsicherheiten spielt auch im täglichen Leben eine Rolle. Bei der Planung einer Aktivität, zum Beispiel eine Wanderung mit der Schulklasse, verfolgt man den Wetterbericht schon einige Tage im Voraus, um gegebenenfalls ein Alternativprogramm buchen zu können. Doch wie vertrauenswürdig sind Wetterberichte? Vor allem wenn es sich um einen Ausblick auf die kommenden Tage handelt? Hier spielt die Messunsicherheit eine große Rolle, denn die Meteorologen arbeiten mit gemessenen Werten und versuchen daraus, mit einer gewissen Toleranz, eine Vorhersage zu treffen. Gut erkennen kann man diese Toleranzen und Schwankungsbreiten an Wahlsonntagen. Hier werden am Abend, wenn die ersten Ergebnisse und Hochrechnungen veröffentlicht werden, Schwankungsbreiten angegeben.

Ein zentrales Missverständnis im Umgang mit Messunsicherheiten besteht in der weit verbreiteten Annahme, es gäbe bei jeder Messung einen eindeutig bestimmbar „wahren Wert“. Diese Vorstellung ist unter Schüler:innen weit verbreitet, zeigt sich jedoch auch vereinzelt bei Studierenden zu Beginn ihrer universitären Ausbildung. Der Glaube an einen exakt messbaren Wert kann dazu führen, dass Abweichungen vom Erwartungswert ausschließlich als Fehler der eigenen Durchführung interpretiert werden, anstatt als Ausdruck der natürlichen Streuung im Messprozess. In der Realität ist der wahre Wert eines Messobjekts jedoch nicht direkt erfassbar, sondern kann lediglich durch wiederholte Messungen und die Angabe einer Messunsicherheit angenähert werden.

Die zufälligen Abweichungen von Messwerten sind allgegenwärtig und führen häufig zu Problemen. Der Umgang mit diesen Abweichungen – der bis hin zur Ablehnung des Zufalls reichen kann – sowie die Gründe für diesen Umgang sollen im Folgenden untersucht werden. Wenn möglich, sollen dabei p-prims identifiziert werden, die diese Herangehensweise erklären.

Im Zuge zweier Masterarbeiten (Bärenthaler-Pachner, 2022) (Loidl, 2021) wurden bereits Unterrichtskonzepte erstellt, mit welchen das Thema der Messunsicherheiten für Schülerinnen und Schüler der Sekundarstufe 1 und der Sekundarstufe 2 greifbar gemacht werden soll. Weiters wurde eine Behaltensstudie

mit einem verzögerten Wissenstest in der Masterarbeit von Eva Abl (Abl, 2024) durchgeführt. Das Ziel der folgenden Arbeit ist es herauszufinden welche Vorstellungen die Schülerinnen und Schüler im Bezug auf Zufall bei Messunsicherheiten haben und vor allem welche p-prims dahinterstehen könnten. Es sollen bekannte Modelle bestätigt oder widerlegt werden. Darüber hinaus sollen, wenn möglich neue p-prims definiert werden, welche zu einem besseren Verständnis über die Gedankenprozesse der Schüler:innen beitragen sollen.

Zentral ist dabei die Frage, ob und wie p-prims diese Vorstellungen beeinflussen. Es soll geprüft werden, ob Schüler:innen den Zufall intuitiv als störende Abweichung, als erklärbares Phänomen oder als unvermeidliche Komponente physikalischer Messung verstehen, und inwiefern diese Erklärsmuster durch p-prims erklärbar sind. Daraus ergeben sich folgende leitende Forschungsfragen:

- Spielen p-prims eine Rolle in den Vorstellungen von Schüler:innen zum Thema Zufall im Kontext von Messunsicherheiten?
- Werden Schüler:innenvorstellungen zum Zufall durch p-prims induziert, und wenn ja, welche lassen sich identifizieren oder sogar neu definieren?

Diese Hauptfragen werden durch eine Reihe untergeordneter Fragestellungen ergänzt, die spezifische Aspekte des Schüler:innenverhaltens in experimentellen Kontexten beleuchten:

- Wie interpretieren Schüler:innen die Ergebnisse von Messreihen, und nach welchen Kriterien bestimmen sie einen „richtigen“ Wert?
- Wie verändert sich ihre Einschätzung, wenn der Stichprobenumfang erhöht wird?
- Gibt es Unterschiede zwischen der Wahrnehmung bei der eigenen Durchführung eines Experiments und der Beobachtung durch Dritte?

Ziel der Arbeit ist es nicht nur, bestehende theoretische Modelle zu bestätigen oder kritisch zu hinterfragen, sondern – sofern möglich – auch neue p-prims zu identifizieren, die bisher nicht systematisch erfasst wurden. Diese könnten wesentlich zu einem tieferen Verständnis der Denkprozesse von Lernenden beitragen und somit Impulse für die Weiterentwicklung didaktischer Konzepte zum Thema Messunsicherheit liefern.

Im ersten Kapitel wird der theoretische Hintergrund und die bereits vorhandene Literatur zum Thema behandelt. Anschließend wird die Methodik, also das Leitfadengestützte Think-Aloud-Interview näher thematisiert. Kapitel 3 beschäftigt sich mit der Qualitätssicherung und schließlich werden in Kapitel 4 die Ergebnisse präsentiert. In Kapitel 5 werden schließlich die erhaltenen Ergebnisse interpretiert und versucht, einen Ausblick zu geben.

1 Theoretischer Hintergrund der Arbeit

Ein wesentlicher Punkt beim wissenschaftlichen Arbeiten ist die bereits vorhandene Literatur. In diesem Kapitel soll der theoretische Hintergrund der Arbeit behandelt werden. Dabei geht es sowohl darum, was man unter Messunsicherheiten versteht bzw. wie phenomenological primitives definiert sind, aber auch um bereits existierende Forschungen.

1.1 Messunsicherheiten in der Naturwissenschaft

In der wissenschaftlichen Praxis ist es notwendig Messunsicherheiten anzugeben, um über Ergebnisse bzw. Messwerte Aussagen treffen zu können. Nur mit diesen Angaben können unterschiedliche Messwerte verglichen werden und nur so können Aussagen über die Präzision der Messungen getätigt werden. Im Alltag werden Unsicherheiten oft als Fehler bezeichnet, was in der Form jedoch nicht als richtig erachtet werden kann.

1.1.1 Der Unterschied zwischen Fehler und Unsicherheit

Fehler bei Messungen sind vermeidbar beziehungsweise im Nachhinein korrigierbar. Man unterscheidet prinzipiell zwei Arten von Fehlern:

Zum einen gibt es den systematischen Fehler, welcher beispielsweise ein Fehler der Kalibrierung des Messgeräts sein kann. Die Abweichung vom Referenzwert kann bestimmt und in weiterer Folge korrigiert werden.

Zum anderen gibt es grobe Fehler, welche vorwiegend an der Bedienung bzw. der Verarbeitung der Daten liegen. Beispiele für solche Fehler sind falsche Auswertungen der erhaltenen Daten, Rechenfehler, falsche Bedienung der Messinstrumente, Ablesefehler oder auch eine grundsätzlich falsche Überlegung bei der Konzeption des Messverfahrens.

Im Gegensatz dazu steht die Unsicherheit, welche durch nicht vermeidbar sind. Bei der Durchführung einer Messung spielen immer Umweltfaktoren eine Rolle, welche zu einer Abweichung zwischen Modell und Realität führen. Die Messunsicherheit ist also eine zufällige Abweichung der Messwerte voneinander bzw. von der Theorie, welche jedoch auf keinen Fehler zurückzuführen sind.

Messunsicherheiten werden nach dem INC-1 in zwei verschiedene Typen unterschieden, welche Typ A-Unsicherheit bzw. Typ B-Unsicherheit genannt.

1.1.2 Bestimmung der Messunsicherheit

1.1.3 Typ-A-Messunsicherheit

Die Typ-A Messunsicherheit beschreibt bei einer Messung die Unsicherheit, die sich aus der statistischen Auswertung wiederholter Messungen ergibt. Man arbeitet also nicht mit Einzelwerten, sondern mit der gesamten Messreihe. Die Grundlage ist die induktive Anwendung statistischer Methoden auf die aus Experimenten resultierende Häufigkeitsverteilung. Mit Hilfe eines geeigneten mathematischen Wahrscheinlichkeitsmodells – meistens die Normalverteilung, manchmal aber auch die Poisson Verteilung – kann diese beschrieben werden.

Ein weiteres wichtiges Merkmal der Typ-A Unsicherheit ist, dass sie in keinem Zusammenhang mit anderen Fehlerquellen oder Unsicherheiten steht, sondern sich ausschließlich aus den inneren Streuungen der Messergebnisse ergibt.

Die gebräuchlichste Methode zur Bestimmung eines Messergebnisses ist die Bestimmung des arithmetischen Mittelwertes $\bar{x}$. Dieser ergibt sich aus der Summe der Einzelwerte geteilt durch die Anzahl der Messwerte:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

Je größer der Stichprobenumfang ist, umso näher kommt man dem „wahren Wert“ auf den im Kapitel 1.1.5 näher eingegangen wird.

Damit man die Streuung der einzelnen Messwerte, um diesen berechneten Mittelwert zu beschreiben, wird die Standardabweichung der Stichprobe s_x bestimmt. Sie zeigt an, wie stark die Messwerte vom Mittelwert abweichen:

$$s_x = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1}} \quad (2)$$

Je größer s_x ist, desto weiter liegen die einzelnen Messwerte auseinander. Bei einer kleinen Standardabweichung streuen die Werte also nur sehr wenig um den Mittelwert. Statistisch liegen 68,3% aller Einzelergebnisse innerhalb des Bereichs von $\bar{x} \pm s_x$. Der Vertrauensbereich kann aber auch breiter definiert werden. Wenn man $\bar{x} \pm 2s_x$ wählt, dann beinhaltet dieses Intervall etwa 95,5% der Messwerte, bei $\bar{x} \pm 3s_x$ sind es sogar 99,7% der Werte.

Neben Mittelwert und Standardabweichung kommen häufig auch weitere Kenngrößen zum Einsatz:

Spannweite (Range)

Ein einfaches Maß dafür die Streuung einer Messreihe zu beschreiben ist die Spannweite R . Diese gibt die Differenz zwischen dem größten und dem kleinsten Messwert an:

$$R = x_{\max} - x_{\min} \quad (3)$$

Diese Methode wird zur schnellen Abschätzung der Streuung verwendet.

Abschätzung der maximalen Unsicherheit

Bei indirekten Messgrößen – wenn ein Wert also aus mehreren Messgrößen berechnet wird – lässt sich die maximale Messunsicherheit mit folgender Formel abschätzen:

$$\Delta y_{\max} = \sum_{i=1}^k \left| \frac{\partial y}{\partial x_i} \right| \cdot \Delta x_i \quad (4)$$

Die Unsicherheiten der einzelnen Einflussgrößen Δx_i auf das Ergebnis y werden durch ihre partiellen Ableitungen gewichtet. Mit dieser Methode können die Obergrenzen für die Unsicherheiten bestimmt werden, was zur Abschätzung des tatsächlichen Ergebnisses hilft.

1.1.4 Messunsicherheit nach der Methode Typ B

Diese Art der Unsicherheit tritt bereits bei einer einzelnen Messung auf. Sie hängt von der Funktionsweise, der Herstellung und der Beschaffenheit des Messgeräts ab. Die Typ B Messunsicherheit ist als vorwiegend vom Messgerät, welche in Form von Eichunsicherheiten, Linearitätsunsicherheiten und Ableseunsicherheit (auch Digitalisierungsunsicherheit genannt) auftritt.

Die Eichunsicherheit hat ihren Namen vom Eichnormal. Dies ist ein Normal, welches mit einer möglichst hohen Genauigkeit erzeugt wurde – spricht eine möglichst geringe Unsicherheit aufweist – und einem internationalen Standard entspricht. Der Prozess der Eichung eines Messgerätes ist streng genommen nur durch das Eich- und Vermessungswesen durchzuführen und dient dazu die Messung des Geräts an das Eichnormal anzupassen. Bei linearen Skalen wird dafür vorwiegend die Ein-Punkt-Kalibrierung verwendet, bei welcher das Ende der Skala mit dem Eichnormal gleichgesetzt wird. Die dabei entstehende Unsicherheit - durch eine mögliche Ungenauigkeit beim Eichungsprozess bzw. beim Eichnormal - nennt man eben Eichunsicherheit.

Nachdem beim Eichungsprozess die Skala definiert wurde, muss diese in kleinere Abschnitte unterteilt werden. Diese Unterteilungen können in den Abständen variieren, wobei nicht bekannt ist, in welche Richtung eine Markierung von ihrem Sollwert abweicht. Die dabei entstehenden Unsicherheiten werden Linearitätsunsicherheiten genannt.

Wenn man in der Praxis Werte von einer Skala ablesen muss und der zu bestimmende Wert zwischen zwei Unterteilungen liegt, dann muss sich der Experimentierende für eine der beiden Markierungen entscheiden. Ein ähnliches Problem gibt es auch bei digitalen Messgeräten, bei denen es nur diskrete Schritte beim Auslesen gibt. Der angezeigte Messwert ist also vom Least Significant Bit abhängig.

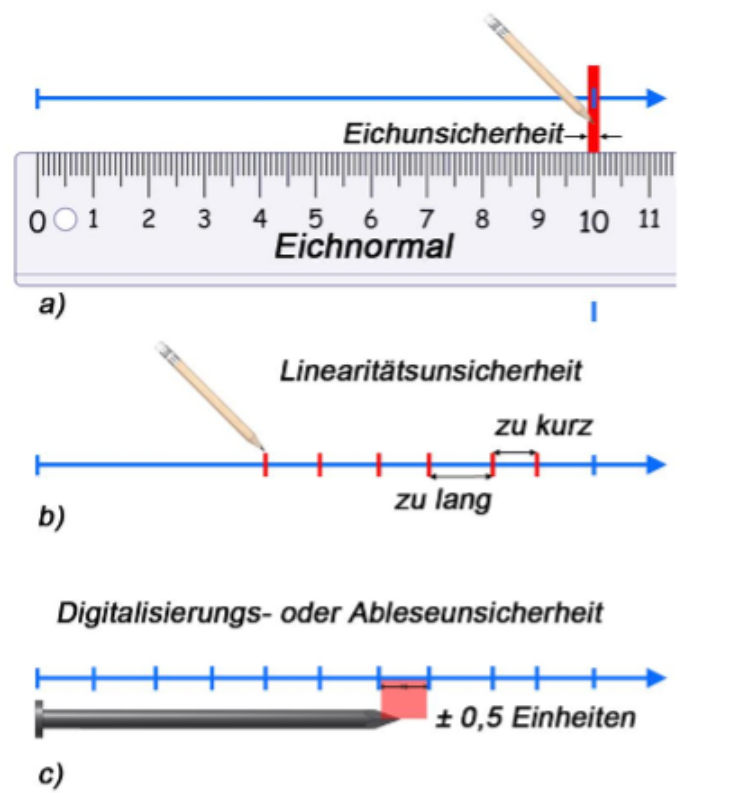


Abbildung 1: Die drei Hauptkomponenten der Typ B Messunsicherheit (Nagel et al., 2021)

Diese Messunsicherheiten werden von dem Hersteller eines Messgeräts entweder auf das Gerät selbst oder zumindest in die Bedienungsanleitung geschrieben. Meistens werden die drei genannten Unsicherheiten zu einem Wert zusammengefasst, manchmal werden diese auch einzeln angegeben, dann müssen diese für die Typ B Messunsicherheit geometrisch addiert werden. Falls keine weiteren Angaben gemacht werden, kann die Auflösung des Messgeräts als Typ B-Messunsicherheit angesehen und für die weiteren Berechnungen verwendet werden.

1.1.5 Der wahre Wert

Dieser Wert ist im Prinzip eine Unbekannte und kann auch nicht exakt bestimmt werden. Bei Beobachtungen der Natur versucht man den wahren Wert einer zu messenden Größe durch eine möglichst exakte Durchführung nahe zu kommen. Auch mathematische Verfahren, mit denen die Messunsicherheiten mitbezogen werden können, helfen dabei bessere Abschätzungen über den wahren Wert treffen zu können.

Wesentlich daran ist, dass dieser Wert jedoch nie bekannt sein kann, weil es grundlegende Prinzipien der empirischen Wissenschaft widersprechen (vgl. Popper, 1934).

Wir können dem „wahren Wert“ nur durch bestmögliche Schätzungen nahekommen – die beste Methode dafür ist das arithmetische Mittel, wie bereits im Kapitel 1.1.3 erwähnt wurde.

Mit Hilfe der Normalverteilung kann man voraussagen, dass mit einer Wahrscheinlichkeit von 68,3% der wahre Wert im Intervall liegen wird, wenn die Richtigkeit der Messung gewährleistet ist (Nagel, 2017).

Die Vorstellung, dass es so einen wahren Wert geben muss, kann aber motivierend wirken, etwa kann dies zur Wiederholung von Messungen oder zur Entwicklung kreativer Messmethoden führen. Es birgt gleichzeitig allerdings die Gefahr, Messabweichungen als Fehler statt als natürliche Streuung zu interpretieren. Fairbrother und Hackling (Fairbrother & Hackling, 1996) kritisieren, dass Schüler:innen bei der Durchführung von Experimenten das Ziel verfolgen, ein „richtiges“ Ergebnis zu erhalten. Jegliche Abweichungen von diesem „richtigen“ Ergebnis wird als persönliches Versagen angesehen.

1.2 Phenomenological primitives

Das Prinzip der Phenomenological primitives, auch p-prims genannt, entspring der Definition von DiSessa (diSessa, 2018). P-prims entstehen aus Beobachtungen im Alltag und können zu Lernendenvorstellungen führen. Es können mehrere p-prims nebeneinander existieren und je nach Situation aktiviert werden. DiSessa sieht in p-prims hypothetische Gedankenstrukturen, die folgende Eigenschaften haben:

Größe und Charakter

Die sehr kleinen Gedankenstrukturen, die jedoch großflächig aktiviert werden. Werden bestimmte Handlungsabläufe erkannt, wird das zuständige p-prim ausgelöst. Für das Individuum benötigen p-prims keine interne Erklärung, sondern werden als immer gültig angesehen und auch so angewandt. Somit werden p-prims für das Individuum zu physikalischen Gesetzen, die sowohl bei bekannten als auch bei vergleichbaren neuen Phänomenen angewandt werden. Die neuen Phänomene werden mit dem p-prim erklärt, während das „Gesetz“ im Gedankenkonstrukt unerklärt bleibt.

Kognitiver Mechanismus

Wie stark ein *p-prim* in den Denkprozessen des Individuums verankert ist, hängt davon ab, wie häufig es aktiviert wird. Stark verankerte *p-prims* sind daher besonders resistent gegenüber Veränderung. DiSessa beschreibt sie als „Knowledge in Pieces“ – also als kleinste, oft implizite Wissenseinheiten, die das intuitive Denken und die Interpretation von Alltagsphänomenen steuern (vgl. diSessa, 1988).

Unterrichtskonzepte wie das „Conceptual Change“ (Posner, Strike, Hewson & Gertzog, 1982) setzen genau an diesem Punkt an: Sie zielen darauf ab, dass Lernende selbst erkennen, dass ihre bisherigen Kon-

zepte zur Erklärung bestimmter Phänomene nicht ausreichen. Erst wenn diese Unzufriedenheit mit dem bisherigen Verständnis entsteht, kann ein tiefergreifender Konzeptwandel erfolgen. Dazu muss das neue Konzept verständlich, plausibel und nützlich erscheinen (vgl. Posner et al., 1982).

Entwicklung

p-prims sind abstrakte Gedankenstrukturen, die sich über die Zeit entwickeln. Zu Beginn sind diese unstrukturiert und nicht oder nur kaum miteinander verknüpft. Mit der Zeit können sich Verknüpfungen bilden bzw. können einzelne p-prims wichtiger und unwichtiger werden. Auch die Rolle von p-prims können sich verändern, je nachdem wie es in das fachliche Gedankenkonstrukt passt. Damit eine Entwicklung der p-prims stattfinden kann, muss durch neue Erkenntnisse das gesamte Gedankenkonstrukt verändert werden.

Systematizität

Dieser Begriff ist schwer abzugrenzen, diSessa (diSessa, 2018) versucht dennoch diesen Zusammenhang zu beschreiben. Dabei nennt er mehrere Aspekte, die für die Systembildung entscheidend sind:

- **Gemeinsame Aktivierung**

Wenn bei einem Prozess oder einer Situation mehrere p-prims gleichzeitig aktiviert werden oder die Aktivierung in Folge eines dynamischen Prozesses von statten geht, entsteht eine gewisse Systematik. Dadurch können diese p-prims als zusammengehörig wahrgenommen werden und gemeinsam entweder als relevanter oder als weniger bedeutsam eingestuft werden.

- **Gemeinsames Vokabular**

Wenn verschiedene p-prims gleiche oder ähnliche Begriffe verwenden, deutet dies auf eine gedankliche Vernetzung hin. Beispielsweise können mehrere p-prims das Konzept von Bewegung oder Kraft beinhalten, was dazu führt, dass diese p-prims einfacher gemeinsam aktiviert werden können.

- **Top-Down-Kohärenz**

Besonders stabile beziehungsweise dominante p-prims können als „Anker“ fungieren, auf welchen weitere p-prims aufgebaut werden. Diese Anker zeichnen sich dadurch aus, dass sie häufig aktiviert werden und somit einen großen Einfluss auf das gesamte Erklärungssystem haben. Sie machen das System resistent gegenüber konzeptuellen Veränderungen.

- **Wechselseitige Plausibilität**

Systematizität zeigt sich auch, wenn sich p-prims gegenseitig bestätigen und stützen. Gleichzeitig verstärken sie damit ihre Glaubwürdigkeit und erscheinen dem Individuum somit als „intuitiv richtig“, egal ob sie aus wissenschaftlicher Sicht korrekt oder falsch sind.

- **Vollständigkeit**

Eine Gruppe von p-prims kann auch eine vollständige Erklärung für ein beobachtetes Phänomen liefern. Verschiedene p-prims decken in den Fall verschiedene Aspekte – zum Beispiel Ursache, Wirkung und Veränderung – desselben Problems ab und ermöglichen es dem Individuum somit das Phänomen vollständig zu beschreiben.

- **Abstraktion und Kontextübertragbarkeit**

Weiters zeichnet sich ein systematisches p-prim Netzwerk dadurch aus, dass die einzelnen p-prims nicht nur auf einen spezifischen Fall anwendbar sind, sondern auf viele Phänomene übertragbar sind.

1.2.1 p-prims im Bezug auf Ereignisse im Kontext von Messunsicherheiten

Schüler:innen gehen, laut Coelho und Séré (Coelho & Séré, 1998) davon aus, dass es für jede Messung einen wahren Wert gibt und man diesen mit einer ausreichend präzisen Messung ermitteln kann. Damit dies funktioniert werden hochwertige und höchstpräzise Messgeräte benötigt, zusätzlich dazu muss die Messung mit großer Sorgfalt durchgeführt werden.

Sorgfalt bei der Messung

Eine Abweichung des Messwertes vom wahren Wert wird vorwiegend auf menschliche Fehler zurückgeführt, sofern ein vermeintlich präzises Messgerät verwendet wird. Die Individuen gehen davon aus, dass diese Abweichungen nur zu Stande kommen, weil die Messung nicht sorgfältig genug durchgeführt wurde. Das Problem liegt also an der falschen Handhabung des Messgeräts oder an der Durchführung der Messung, welche vorwiegend auf mangelnde Konzentration oder fehlende Erfahrung im Umgang mit dem Messaufbau zurückzuführen ist. Das dahinterstehende p-prim „von nichts kommt nichts“ wird bei DiSessa (diSessa, 1993) als „Actuating agency“ bezeichnet.

Häufigkeit als Indikator für den wahren Wert

Nachdem die Existenz eines wahren Wertes eine weit verbreitete Lernendenvorstellung ist, schließen in weiterer Folge sehr viele darauf, dass die Häufigkeit eines Messwertes auf dessen Richtigkeit hinweist. Es wird also jener Wert der Messreihe als wahrer Wert interpretiert, der am häufigsten auftritt. Alle anderen Messwerte werden als Ergebnis einer ungenauen Messung oder eines Fehlers angesehen. Das p-prim „viel hilft viel“ ist bereits als „Ohm´s p-prim“ bekannt (diSessa, 1993).

1.2.2 p-prims im Bezug auf Zufall

Obwohl in den Ausführungen von diSessa der Begriff des Zufalls nicht explizit erwähnt wird und er somit auch keine p-prims dafür definiert hat, können einige p-prims identifiziert werden, welche in zufallsbezogenen Situationen häufig aktiviert werden.

Unkenntnis als Quelle des Zufalls

Ein häufiges Erklärungsmuster lautet: *„Wenn ich keinen Grund erkennen kann, dann ist das Ereignis zufällig.“* (Mittelstraß, 1996). Das bedeutet, dass diese Individuen davon ausgehen, dass ein Ereignis zufällig auftritt, sofern man keine Ursache erkennen kann. Der Zufall füllt in diesem Fall eine Lücke im eigenen Wissen und führt somit zu einer sehr subjektiven Sichtweise des Zufalls, weil dieser von der individuellen Perspektive abhängt.

Komplexität führt zu Unvorhersagbarkeit

Auch wenn es sich um komplexe Systeme oder Prozesse handelt, bei welchen viele gleichzeitig wirkende Faktoren zusammenspielen müssen, wird häufig aufgrund der Unvorhersehbarkeit mit dem Zufall argumentiert (Weßling, 2022). Das dahinterliegende intuitive Konzept kann so beschrieben werden: *„Wenn viele Dinge gleichzeitig wirken, ist das Ergebnis nicht vorhersagbar.“*

Glück und Pech als Erklärungsmodelle

Das wahrscheinlich weit verbreitetste p-prim beziehungsweise die weit verbreitetste Erklärung für Ergebnisse abstrakter Konzepte sind Glück und Pech (Sill & Kurtzmann, 2012). Der Zufall wird mit dieser Sichtweise komplett aus der Gleichung genommen, man reduziert das Ereignis auf ein schicksalhaftes Element: *„Wenn ich gewinne, hatte ich Glück. Wenn nicht, hatte ich Pech.“*

Häufigkeit als Wahrheit

Wie bereits in 1.2.1 erwähnt, werden auch bei Zufallsprozessen intuitiv Häufigkeit und Richtigkeit beziehungsweise Wahrheit gleichgesetzt. Die Individuen tendieren dazu, den am häufigsten beobachteten Wert oder das am öftesten auftretende Ereignis als „richtig“ anzusehen. *„Wenn ein Wert oft auftritt, muss er stimmen.“* ist eine häufig getätigte Aussage in diesem Kontext. Dadurch kann ein Missverständnis gefördert werden, welches beispielsweise die Annahme stützt, dass nach mehreren „Kopf“-Ergebnissen beim Münzwurf „Zahl“ folgen müsse. Gleichzeitig kann in diesem Szenario auch die Fehlvorstellung auftreten, dass „Kopf“ die richtige Lösung sei, beispielsweise weil die Münze gezinkt sei.

Gigerenzer (2015) beschreibt dieses Phänomen als typische Fehlinterpretation von Wahrscheinlichkeiten: Menschen neigen dazu, aus der Häufigkeit eines Ereignisses voreilig auf dessen Richtigkeit oder Gesetzmäßigkeit zu schließen. Solche kognitiven Verzerrungen führen etwa zur irrigen Erwartung eines Ausgleichs nach mehreren gleichen Ergebnissen – ein Denkfehler, der unter dem Begriff *Gambler's Fallacy* bekannt ist (Gigerenzer, 2015).

1.3 Lernendenvorstellungen

Alle Schüler:innen im Unterricht haben ihren eigenen familiären und auch schulischen Werdegang aus welchem unterschiedlichste Vorstellungen hervorgehen. Einige dieser Vorstellungen entstehen aufgrund der Alltagserfahrungen, der Alltagssprache oder auch der Medien, es gibt jedoch auch Vorstellungen, die durch den vorangegangenen Unterricht hervorgerufen werden. Diese Vorstellungen sind für die Ler-

nenden etwas nur schwer veränderliches, weil man es selbst gesehen/beobachtet hat oder von, für die Lernenden, vertrauenswürdigen Quellen gehört hat.

„Lernendenvorstellungen beschreiben Dispositionen, d. h. Tendenzen von Schülerinnen und Schülern, physikalische Begriffe in einer bestimmten Weise zu interpretieren oder Phänomene in einer bestimmten Weise zu beschreiben, [...]“ (Schecker, Wilhelm, Hopf & Duit, 2018)

Im Unterricht muss den Lernenden die Möglichkeit geboten werden diese Vorstellungen herauszufordern und auch verändern zu können.

1.3.1 Lernendenvorstellungen im Bezug auf Messunsicherheiten

Punkt-Paradigma

Lernende gehen davon aus, dass eine einzige, sorgfältig durchgeführte Messung ausreicht, um den wahren Wert zu liefern. Dieses Denkmuster wird als Punkt-Paradigma bezeichnet und zeigt sich vor allem dadurch, dass entweder nur eine einzelne Messung durchgeführt wird oder bei Messreihen der erste Messwert als wahr angenommen wird (Buffler, Allie & Lubben, 2001b).

Set-Paradigma

Im Gegensatz dazu steht das Set-Paradigma, bei dem mehrere Messwerte notwendig sind, um mithilfe von Mittelwert und Streuung eine qualitative Aussage über die Messung treffen zu können (Buffler et al., 2001b).

Präzision der Messgeräte

Eine der am häufigsten auftretenden Lernendenvorstellungen in diesem Kontext ist, dass es Messgeräte gibt, die so exakt sind, dass es zu keinen Unsicherheiten kommt. Jede sorgfältig durchgeführte Messung führt, laut dieser Vorstellung, immer zum gleichen – dem wahren – Wert.

Daraus folgt auch die Annahme, dass weniger empfindliche Messgeräte genauere Ergebnisse liefern, da die Schwankungen wesentlich geringer oder nicht merkbar sind. Beim Vergleich von Mikrometerschraube und Maßband fällt auf, dass beim Maßband immer eine Länge von beispielsweise 7,2 cm gemessen wird, während es bei der Mikrometerschraube kleine Abweichungen in den Nachkommastellen gibt. Diese Abweichungen können auf Basis dieses p-prims als vermeintlich geringere Genauigkeit interpretiert werden.

Naive Vorstellung zur Wiederholung von Messungen

Eine Wiederholung einer Messung erfolgt oft nur, wenn es zu einer Diskrepanz mit Literaturwerten kommt oder ein offensichtlicher Fehler erkannt wird. Dies liegt daran, dass die Vorstellung verbreitet ist, dass bei wiederholten Messungen immer derselbe Wert gemessen wird. Wenn dem nicht so ist, wird dies als Fehler angesehen, der auf menschliches Versagen zurückzuführen ist (Pollard, Hobbs, Stanley, Dounas-Frazer & Lewandowski, 2017).

Bewertung von Datenqualität

Wenn von zwei oder mehr Datensätzen der Mittelwert und die Streuung gegeben sind, gehen Lernende oftmals davon aus, dass der Mittelwert das entscheidende Kriterium für die Messqualität ist. Wenn zwei Datensätze also denselben Mittelwert aufweisen, aber unterschiedliche Streuungen haben, denken die Lernenden, dass diese beiden Messreihen gleich gut sind (Majiet & Allie, 2018).

1.3.2 Lernendenvorstellungen im Bezug auf zufällige Ereignisse

Eine einzelne Messung gibt einen wahren Wert

Schüler:innen die vor allem noch wenige Erfahrung mit Messprozessen und Messdaten haben neigen dazu bereits einen einzelnen Messwert als wahres und „richtiges“ Ergebnis anzusehen. In der Literatur sprechen Höttecke und Hopf vom „Point-Verständnis“. Konträr zum „Set-Verständnis“ bei dem die Schüler:innen den Messprozess bzw. die Messdaten statistisch verstehen, wird angenommen, dass man feststehende Gesetze der Natur bereits mit einzelnen Messungen aufdecken kann. Es wird also der Wahrheitswert einzelner Messungen deutlich überschätzt. Als Quelle für diese Fehlvorstellung sehen Höttecke und Hopf den Begriff des Messfehlers, weil es den Schüler:innen assoziiert, dass man, wenn die Messung fehlerfrei durchgeführt wird, den richtigen – wahren – Wert erhalten kann, egal wie viele Messungen man durchführt. Besser wäre der Begriff der Messunsicherheit, weil damit vermittelt wird, dass es einen gewissen Vertrauensbereich gibt, in welchem gemessen und abgeschätzt wird (Schecker et al., 2018, S. 279-280).

In der Physik gibt es keinen Zufall

Der Begriff des Zufalls ist für Schüler:innen mit Unwissen verbunden. Wüsste man alle Parameter bzw. könnte man den Messprozess komplett kontrollieren, dann gäbe es keine Messunsicherheit und somit auch keinen Zufall. Selbst für Albert Einstein war die Vorstellung von zufälligen Prozessen unsympathisch, allgemein akzeptiert ist jedoch der aktuelle Forschungsstand der Physik, welcher besagt, dass es zufällige Einzelereignisse in physikalischen Prozessen gibt (Schecker et al., 2018, S. 229-230).

Gleiche Ursachen haben gleiche Wirkung

Lernende gehen davon aus, dass es in der Physik immer Ursachen und Wirkungen gibt, somit ist kein Platz für zufällige Ereignisse. Man muss nur die Ursache genau genug kennen, dann kann auch jegliche Wirkung eindeutig vorhergesagt werden. Und wenn es zu Ereignissen kommt, die unerwartet sind, dann kommt dies nur vor, weil man zu wenige Informationen über die Ursache kennt. Es wird darüber hinaus auch zwischen der Theorie und der realen Welt unterschieden, da in der realen Welt mehr Einflussfaktoren vorhanden sind (Schecker et al., 2018, S. 229-231).

Zufällige Ereignisse sind unstrukturiert

Wenn es zu zufälligen Ereignissen kommt, dann müssen diese als unstrukturierte Verteilungen auftreten. Eine Häufung von Messwerten um einen bestimmten Punkt wird nicht als zufälliges Ereignis erkannt. Wenn Muster in einer Messreihe erkannt werden, dann wird diese als nicht zufällig definiert (Schecker et al., 2018, S. 229-231).

1.4 Wie hängen p-prims und Lernendenvorstellungen zusammen?

Als grundlegende gedankliche Strukturen, als welche p-prims bereits im Kapitel 1.2 definiert wurden, entstehen sie aus alltäglichen Beobachtungen und Erfahrungen. Man könnte sie als eine Art intuitives Wissen beschreiben, welche sich Menschen durch wiederholte Wahrnehmungen im Alltag aneignen. Bei neuen Situationen werden diese Strukturen verwendet, um die neuen Eindrücke zu verstehen oder zumindest eine für sich plausible Erklärung zu bilden. Dabei ist der Grad der Abstraktheit dieser neuen Situation entscheidend darüber, ob wir ein bereits vorhandenes Konzept (p-prim) anwenden können oder ob ein neues Konzept entwickelt werden muss.

„Je näher, desto stärker“ ist ein sehr anschauliches Beispiel dafür. Dieses p-prim beschreibt die Beobachtungen, dass die Intensität eines Einflusses mit dem Abstand zur Quelle abnimmt. Bereits im Kindesalter beobachtet man dies bei Geräuschen, die wesentlich lauter sind, wenn man näher an der Schallquelle sitzt oder steht. Auch bei der Heizung im Haus oder bei einem Lagerfeuer merken selbst Kinder den Zusammenhang von Intensität sehr schnell. Bei Lichtquellen erkennt man darüber hinaus ebenfalls einen solchen Zusammenhang, da eine Lichtquelle aus der Nähe wesentlich heller wirkt als aus größerer Distanz. All diese Beobachtungen gehen intuitiv und häufig von statten und setzen das Konzept „je näher, desto stärker“ tief in die Gedankenstrukturen fest. In vergleichbaren Situationen wird automatisch auf diese, für das Individuum fixe Regel, zurückgegriffen.

Dass dieses intuitive Konzept jedoch zu Fehlvorstellungen führen kann, zeigt sich, wenn man beim erwähnten Beispiel „je näher, desto stärker“ auf den Zusammenhang der Jahreszeiten, im Besonderen Sommer und Winter, mit der Entfernung der Erde zur Sonne eingeht. Auf Basis des p-prims gehen Lernende davon aus, dass die Erde im Sommer wesentlich näher an der Sonne sein muss als im Winter. Es wird also ein bekanntes p-prim auf eine scheinbar vergleichbare Situation angewendet.

Das zeigt, dass Lernendenvorstellungen durch die unreflektierte Anwendung von bekannten p-prims entstehen. Der Begriff „unreflektiert“ geht hierbei auf Smith III, diSessa und Roschelle (Smith III, DiSessa & Roschelle, 1994) zurück, welche damit meinen, dass das p-prim weder in ihrem physikalischen Kontext noch in ihrer ursprünglichen Intention interpretiert wird. Es wird einfach ohne kritische Prüfung angewendet.

1.5 Aktueller Forschungsstand

Während Lernendenvorstellungen zu Messen, Messunsicherheiten und Messfehlern bereits gut erforscht sind, wurden p-prims in diesem Bereich bislang kaum untersucht. Zahlreiche Studien belegen, dass Schüler:innen häufig naive oder intuitive Vorstellungen zum Charakter von Messprozessen und der Genauigkeit von Messwerten entwickeln, welche sich über die Schulzeit hinweg verfestigen können (vgl. (Allie, Buffler, Campbell & Lubben, 1998); (Buffler, Allie & Lubben, 2001a); (Lubben & Millar, 1996); (Sere, Journeaux & Larcher, 1993)).

Ein wesentlicher Ausgangspunkt ist die Unterscheidung zwischen dem sogenannten Point-Paradigma und dem Set-Paradigma (vgl. (Abbott, 2003); (Lubben, Campbell, Buffler & Allie, 2001)). Während das Point-Paradigma davon ausgeht, dass eine einzelne Messung bereits den „richtigen“ oder „wahren“ Wert liefert, liegt dem Set-Paradigma ein statistisches Verständnis zugrunde, bei dem mehrere Messwerte gemeinsam analysiert werden, um eine Aussage über deren Streuung und Relevanz treffen zu können. Studien zeigen, dass insbesondere jüngere Lernende tendenziell dem Point-Paradigma folgen, was zu einem übersteigerten Vertrauen in Einzelmessungen führt (vgl. (Davidowitz, Lubben & Rollnick, 2001); (Rollnick, Dlamini, Lotz & Lubben, 2001)).

In diesem Zusammenhang kommt dem Begriff des wahren Werts besondere Bedeutung zu. Die Annahme, es gäbe bei jeder physikalischen Größe einen exakt bestimmbaren Wert, ist weit verbreitet – sowohl unter Schüler:innen als auch unter Studienanfänger:innen (Fairbrother & Hackling, 1996). Diese Vorstellung steht jedoch im Widerspruch zu modernen wissenschaftstheoretischen Konzepten, welche betonen, dass der wahre Wert in der empirischen Wissenschaft prinzipiell nicht zugänglich ist, sondern nur angenähert werden kann (vgl. (Popper, 1934); (Nagel, 2017)).

Ergänzend liefern Arbeiten von Heinicke (Heinicke, 2012), Leach (Millar, Niedderer & et al., 1998) und Pillay (Pillay, Buffler, Lubben & Allie, 2008) wertvolle Einblicke in typische Fehlvorstellungen, die bei der Interpretation von Messdaten auftreten. Diese reichen von der unkritischen Übernahme gerundeter Zahlen als besonders exakt („schöne Zahlen“) bis hin zur irrigen Annahme, dass der häufigste Wert einer Messreihe automatisch der genaueste sei. Derartige Vorstellungen lassen sich unter Rückgriff auf diSessas Konzept der *phenomenological primitives* analysieren. P-prims wie „ähnlich ist gleich richtig“ oder „häufig heißt korrekt“ sind dabei typische Beispiele, die sich häufig in den Denkprozessen der Lernenden nachweisen lassen (vgl. (diSessa, 1993); (Coelho & Séré, 1998)).

Zudem existieren empirische Arbeiten, die die Verbindung zwischen Lernendenvorstellungen und p-prims explizit untersuchen. So zeigen beispielsweise Warwick (Warwick, Linfield & Stephenson, 1999) und Ibrahim (Ibrahim, 2006), dass viele Schüler:innen dazu neigen, eigene Vorerfahrungen und intuitive Erklärungsmuster auf experimentelle Situationen zu übertragen, auch wenn diese aus wissenschaftlicher Sicht unzutreffend sind. Dieses Verhalten unterstreicht die Bedeutung einer gezielten Auseinandersetzung mit p-prims im Unterricht.

Insgesamt zeigt sich, dass es zwar eine Vielzahl an Studien zu allgemeinen Schülervorstellungen in der Physik gibt, jedoch vergleichsweise wenige Untersuchungen, die spezifisch den Zusammenhang zwischen Zufall, Messunsicherheiten und p-prims thematisieren. Die vorliegende Arbeit greift diese Forschungslücke auf und leistet somit einen Beitrag zur systematischen Erhebung und Analyse intuitiver Denkstrukturen von Schüler:innen im Kontext von Messprozessen.

2 Methodik der Arbeit

2.1 Grundlagen qualitativer Forschung

Qualitative Forschung ist, dadurch dass sie sowohl explorativ als auch interpretativ ist, eine zentrale Methode, um komplexe und öfters auch vielschichtige Phänomene verständlich zu machen beziehungsweise ein tiefergehendes Verständnis dafür zu entwickeln (Mayring, 2016). Breite Anwendung findet sie vor allem in der Sozial- und Bildungswissenschaft, da mit dieser Methode soziale Zusammenhänge, individuelle Erfahrungen und menschliches Verhalten erfasst und analysiert werden können (Flick, von Kardorff & Steinke, 2000). Die methodischen Ansätze reichen von Beobachtungen über qualitative Interviews bis hin zur Analyse von Dokumenten und anderen schriftlichen Quellen. Qualitative Forschung zeichnet sich durch den Fokus auf das Erleben und Verstehen aus, welche auf Basis individueller Perspektiven erhoben werden (Friebertshäuser, Boller, Prengel & Langer, 2013).

Der wesentliche Unterschied zur quantitativen Forschung liegt darin, dass qualitative Verfahren in erster Linie nicht darauf aus sind, numerische Daten zu erhalten. Es ist möglich, dass die erhaltenen Ergebnisse gegliedert und abgezählt werden, dies ist aber nicht das primäre Ziel. Es geht vielmehr um das Nachvollziehen beziehungsweise das Verstehen von subjektiven Eindrücken und Verhaltensmustern (Lüders, 1993).

Auch die Flexibilität ist ein entscheidendes Merkmal qualitativer Forschung. Durch die Offenheit im Forschungsprozess ist es möglich, dass die Fragestellungen und das Forschungsdesign an die jeweils vorherrschenden Gegebenheiten angepasst werden können, was vor allem in dynamischen, wenig vorhersehbaren Kontexten einen großen Vorteil bietet. Die beobachteten Phänomene werden zudem in ihrem sozialen, kulturellen beziehungsweise historischen Kontext betrachtet, was dazu beiträgt, ein differenziertes Verständnis über die Einbettung und die Vielschichtigkeit der Phänomene in ihrer Umwelt zu gewinnen.

2.2 Leitfadengestütztes Interview

Bei der Durchführung eines leitfadengestützten Interviews wird eine qualitative Forschungsmethode zum Einsatz gebracht, die sich durch ihre besondere Mischung aus Struktur und Flexibilität auszeichnet. Wesentlich für die Umsetzung ist ein Leitfaden, der als Skript für das Interview dient. Dieser Leitfaden enthält die zentralen Fragestellungen und thematischen Schwerpunkte, welche das Interview abdecken soll. Das besondere daran ist jedoch, dass es keine starre Vorgabe ist, sondern der interviewenden Person genügend Freiraum lässt, um spontan auf Aussagen der befragten Person einzugehen und eventuell näher nachzufragen oder auch Fragen zu überspringen, wenn diese eigenständig beantwortet wurden. Zusätzlich können auch weitere Themen und Situationen behandelt werden, welche sich im Gespräch ergeben.

Der Leitfaden dient also als Orientierungshilfe, die eine Zielrichtung und eine Struktur vorgibt, ohne den Gesprächsverlauf zu starr vorzugeben. Dadurch können auch unvorhergesehene neue Erkenntnisse gewonnen werden, mit denen beim Erstellen des Leitfadens nicht gerechnet wurde. Diese Art der Interviewführung wird vor allem in naturwissenschaftlichen Forschungen verwendet.

Niebert und Gropengießer (Krüger, Parchmann & Schecker, 2014, S. 126) haben die wichtigsten Anforderungen an solch einen Interviewleitfaden definiert. Demnach sollte ein Leitfaden folgende Kriterien erfüllen:

- **Übersichtlichkeit und Lenkung ohne Ablenkung**

Der Leitfaden soll übersichtlich und klar gestaltet sein, so dass die interviewende Person sich einfach orientieren kann und nicht abgelenkt oder überfordert wird. Die Funktion des Leitfadens ist es, den Gesprächsverlauf zu lenken, diesen aber nicht zu dominieren oder die Atmosphäre des Gesprächs zu beeinträchtigen.

- **Vermeidung von Überladung und logische Struktur**

Der Leitfaden sollte sich auf die wesentlichen Themen beziehungsweise Fragestellungen konzentrieren und nicht mit Informationen oder Fragestellungen überladen sein. Die enthaltenen Fragen und Themen sollen in einer logischen Reihenfolge angeordnet sein, damit der Gesprächsfluss eher gefördert als behindert wird.

- **Flexibilität im Einsatz**

Sowohl beim Erstellen des Leitfadens, als auch bei der Verwendung dessen ist ein zentraler Punkt, dass dieser nicht als starres Gerüst verstanden werden soll, welches strikt abgearbeitet werden muss. Er soll viel mehr dazu dienen den natürlichen Gesprächsfluss durch Impulse und gezielte Fragestellungen zu fördern und als Inspirationsquelle für die interviewende Person fungieren. Die interviewte Person sollte niemals daran gehindert werden ihre Gedanken frei zu äußern.

2.3 Think aloud Interview

Das Think-aloud-Interview ist eine qualitative Erhebungsmethode, bei der die interviewte Person dazu animiert wird, jegliche Denkprozesse laut auszusprechen. Dies kann sowohl die Gedanken zu einem vorgeführten Video beinhalten oder auch den Gedankengang beim Lösen eines Rätsels. Das zentrale Ziel dieser Methode ist es, einen detaillierten Einblick in die kognitiven Prozesse der befragten Person zu erhalten. So soll es ermöglicht werden, Denkweisen, Entscheidungsprozesse, aber auch mögliche Fehlvorstellungen aufzudecken (Charters, 2003).

Das Besondere an dieser Form des Interviews ist, dass es weit über ein traditionelles Frage-Antwort-Interview hinausgeht und somit auch viel tiefgehendere Phänomene erforscht werden können. Während bei klassischen Interviews direkte Antworten auf vorgefertigte Fragen folgen, bietet diese Methode die Möglichkeit, auf Annahmen und Überlegungen einzugehen, die bei herkömmlichen Befragungen häufig unterdrückt werden. Befragte können ihren Gedanken freien Lauf lassen, ohne durch vorgefertigte

Strukturen oder feste Fragenformate eingeengt zu werden (Eccles & Aarsal, 2017).

Dadurch eignet sich das Think-aloud-Interview hervorragend zur Identifikation von Fehlvorstellungen in einem spezifischen Kontext. Es ermöglicht Forschenden, die Denkprozesse der Teilnehmenden detailliert nachzuvollziehen und somit ein tieferes Verständnis für deren Entscheidungsfindung zu entwickeln (van Someren, Barnard & Sandberg, 1994).

2.4 Transkribieren

Die Auswertung von Interviews, im Besonderen bei offenen Interviewformaten wie Think-aloud Interviews, ist eine präzise Dokumentation essenziell. Durch eine gründliche Dokumentation wird die Basis für eine aussagekräftige Analyse der Daten gelegt. Je qualitativ hochwertiger die erhobenen Daten sichergestellt werden, desto valider sind die Rückschlüsse und somit auch die Beantwortung der Forschungsfrage.

Entscheidend ist daher, dass bei der Datenaufbereitung der Rohdaten diese auf Genauigkeit und vor allem auf Vollständigkeit überprüft werden. Krüger, Parchmann und Stecker ((Krüger et al., 2014)) betonen, wie essenziell dieser Schritt ist um sicherzustellen, dass keine relevanten Informationen verloren gehen und die Daten korrekt erfasst wurden. Jegliche Ungenauigkeit in den Daten schadet der Qualität der Analyse und kann im schlimmsten Fall zu fehlerhaften Interpretationen führen.

Die erhaltenen Rohdaten müssen in ein geeignetes Analyseformat konvertiert werden. Dies ist notwendig um die Daten für die Analyse, beispielsweise durch Codierung, verwenden zu können.

Ein essenzieller Schritt, der im Zuge der Transkription der Daten erfolgt, ist die Anonymisierung der Daten. Dieser Prozess ist nicht nur wichtig für die Privatsphäre der interviewten Personen, sondern darüber hinaus auch ein elementarer Bestandteil wissenschaftlicher Forschung. Es dient dazu den ethischen Richtlinien in der Forschung zu entsprechen. Besonders in sensiblen Forschungsbereichen, in welchen persönliche Daten beziehungsweise intime Einblicke in die Handlungen und Gedanken der Teilnehmer:innen gewährt wird, hat die Anonymisierung dieser Daten und der Schutz der Privatsphäre oberste Priorität.

2.5 MAXQDA

Diese Software dient der Aufarbeitung von Daten ist auf qualitative beziehungsweise Mixed-Methods-Datenanalysen spezialisiert. Die zentrale Funktion ist das Codieren, also ein Prozess, bei dem die Rohdaten kategorisiert werden können. Im System werden diese Kategorien Codes genannt und repräsentieren einzelne kleine oder größere Segmente der Rohdaten. Zudem bietet MAXQDA auch eine umfangreiche Analyse von Text-, Audio- und Videodaten sowie Bildern an. So kann die Software beispielsweise eigenständige Transkriptionen von Audiodaten erstellen, welche in weiterer Folge codiert werden können.

MAXQDA unterstützt durch geteilte Projekte, gemeinsame Codes und das Erstellen thematisch strukturierter Veröffentlichungen die Zusammenarbeit im Team und darüber hinausgehend auch die Sicherstellung der Intercoder-Übereinstimmung (siehe Kapitel 3.1.3).

2.6 Kategorienbildung

Um die Aussagen in den Interviews einordnen zu können müssen Kategorien gebildet werden, in welche man die unterschiedlichen Antworten zuordnet. In der Forschung unterscheidet man zwischen zwei verschiedenen Arten, deduktive und induktive Kategorien.

2.6.1 Deduktive Kategorien

Deduktive Kategorien beruhen auf bereits vorhandener Literatur und sind somit auf bestehenden wissenschaftlichen Erkenntnissen, Theorien und Modellen aufgebaut. Sie werden bereits vor dem Erstellen des Leitfadens festgelegt, dadurch kann der Leitfaden auf diesen Kategorien aufgebaut werden und diese bestätigen oder widerlegen. Sie geben der Analyse aber auch den Befragungsinstrument ein strukturelles Gerüst, welches als Ausgangspunkt für weitere Erkenntnisse bietet.

Mit diesen deduktiven Kategorien wird auch bereits festgelegt, welche Themen beziehungsweise welche Aspekte im Rahme der Untersuchung von besonderem Interesse sind. Bei der Datenauswertung dienen sie als systematische Orientierungshilfe.

Obwohl sie bereits vorab definiert sind, sollen aber auch die deduktiven Kategorien nicht starr verwendet werden, vielmehr sollen sie im Verlauf der Forschung flexibel bleiben. Man kann also die Kategorien auch gegebenenfalls an die Erkenntnisse anpassen, falls sich bei der Datenerhebung oder auch bei der Analyse neue Themen auftun.

2.6.2 Induktive Kategorien

Induktive Kategorien werden direkt aus den erhobenen Daten entwickelt und sind eng an die Antworten der befragten Personen gekoppelt. Sie entstehen während der Datenanalyse und spiegeln die Inhalte und Muster wider, die sich aus den gegebenen Antworten ergeben. Diese Methode ermöglicht es, die Vielfalt und Nuancen der Daten möglichst unverfälscht zu erfassen und abzubilden.

Der Prozess der Erstellung und Definition induktiver Kategorien erfolgt in mehreren Schritten. Zunächst werden die Rohdaten, beispielsweise Transkripte von Interviews, detailliert analysiert, um offene Codes zu entwickeln. Diese offenen Codes sind kurze, prägnante Beschreibungen oder Schlagworte, die spezifische Inhalte, Aussagen oder Themen in den Daten zusammenfassen. In dieser Phase bleibt die Analyse bewusst ergebnisoffen, um den Daten Raum zu geben, eigene Strukturen und Bedeutungen hervorzubringen, ohne von vorab festgelegten Kategorien beeinflusst zu werden.

Anschließend werden diese offenen Codes untersucht, um Gemeinsamkeiten oder thematische Überschneidungen zu identifizieren. Wenn ähnliche oder zusammenhängende Codes festgestellt werden, können sie in übergeordnete Kategorien zusammengefasst werden. Dieser Schritt dient dazu, die Daten zu strukturieren und die Vielzahl individueller Codes in einem systematischen Rahmen zu organisieren, während die Vielfalt der Antworten erhalten bleibt. Die Zusammenführung in Kategorien erfolgt iterativ, das heißt, die Codes und Kategorien werden so lange überprüft und angepasst, bis eine sinnvolle, konsistente Struktur entsteht.

Durch diese induktive Herangehensweise wird eine tiefere und kontextbezogene Analyse der Daten gewährleistet. Der Fokus liegt darauf, die spezifischen Perspektiven der Befragten zu verstehen und neue, bisher unbekannte Muster oder Themen zu entdecken. Diese Offenheit ist besonders wertvoll in explorativen Forschungsprojekten, bei denen es darum geht, neue Erkenntnisse zu gewinnen oder unbekannte Phänomene zu untersuchen.

Laut der „Methoden naturwissenschaftlicher Forschung“ (Krüger et al., 2014) ist der induktive Ansatz besonders gut geeignet, um neue Hypothesen zu generieren. Da die Kategorien direkt aus den Daten abgeleitet werden, spiegelt die Analyse die tatsächlichen Gegebenheiten und Perspektiven wider, was die Grundlage für innovative und fundierte Hypothesenbildung bietet. Dadurch kann der induktive Zugang entscheidend dazu beitragen, die theoretische Weiterentwicklung eines Forschungsfeldes voranzutreiben und neue Forschungsfragen oder Ansätze zu identifizieren.

2.7 Codierleitfaden

Der Codierleitfaden ist eine klare Definition der einzelnen Kategorien, der klare Richtlinien für eine Zuordnung oder eine nicht Zuordnung in eine bestimmte Kategorie vorgibt. Dieser Leitfaden soll sicherstellen, dass jede unabhängige Person diese Zuordnung durchführen kann und im Idealfall auf dieselben Ergebnisse kommt. Mit Hilfe von Anker- beziehungsweise Abgrenzungsbeispielen werden die Grenzen der einzelnen Kategorien klar gesteckt. Bei der Erstellung eines Codierleitfadens werden zunächst die Kategorien mit aussagekräftigen Bezeichnungen benannt und anschließend genauer definiert um eine eindeutige Zuordnung zu einer Kategorie sicher zu stellen. Außerdem werden Ankerungs- und/oder Abgrenzungsbeispiele angegeben, welche entweder die Definition unterstreichen oder bei Formulierungen, die im Überschneidungsbereich zweier Kategorien liegen, eine richtige Zuordnung sicherstellen. Diese werden aus den gesammelten Antworten ausgewählt, um die Vielschichtigkeit der Aufgabenlösungen widerszuspiegeln und die Ausdrucksweise der Interviewten abzubilden (Krüger et al., 2014, S. 175). Die Qualität der erhobenen Daten steht in einem engen Zusammenhang mit einer klar abgegrenzten Interpretation und Zuweisung der Aussagen, was durch einen Codierleitfaden gegeben ist (Krüger et al., 2014, S. 187).

3 Qualitätssicherung

3.1 Erkenntnistheoretische Qualitätskriterien

3.1.1 Objektivität

Als eines der zentralen Gütekriterien wissenschaftlicher Studien beschreibt die Objektivität die Unabhängigkeit der Ergebnisse von den handelnden Personen. Eine Studie sollte also so gestaltet sein, dass unterschiedliche Personen, beim Durchführen der Studie unter gleichen Bedingungen auch zu den gleichen Ergebnissen kommen.

Im Fall dieser Studie wurden mehrere Aktionen gesetzt um diesem Gütekriterium zu entsprechen. Durch die Erstellung eines Leitfadens (vgl. Kapitel 2.2) wird bei der Durchführung der Interviews bereits ein wichtiger Schritt zur Objektivität gesetzt. Die Qualität des Leitfadens wurde überdies durch eine Expert:innenbefragung sichergestellt. Mit Hilfe des Codierleitfadens (vgl. Kapitel 2.7) werden auch die Grenzen bei der Auswertung klar abgesteckt. Zudem gilt die Intercoder-Reliabilität (vgl. Kapitel 3.1.3) auch als „Signierobjektivität“ womit bei gegebener Intercoder-Reliabilität auf die Objektivität geschlossen werden kann. Jedoch ist aufgrund des erforderlichen Interpretationsprozesses bei der Zuordnung der Kategorien keine vollständige Objektivität garantiert werden.

3.1.2 Validität

Die Validität bezieht sich auf die Richtigkeit der Messungen und gibt an wie genau die Untersuchung das misst, was sie zu messen vorgibt. Sie gibt somit an, ob die Schlussfolgerungen und Ergebnisse der Studie aussagekräftig und inhaltlich zutreffend sind. Aufgrund der Auswahl der Forschungsmethode „think aloud Interview“ werden die Daten, aufgrund des praxisnahen Erfassens der Daten, als hoch Valide bezeichnet. Die erhobenen Daten sind somit eng mit den untersuchten kognitiven Prozessen verknüpft. Auch der Stichprobenumfang ist ein wichtiges Kriterium der Validität.

3.1.3 Reliabilität

Dieses Gütekriterium bezieht sich auf die Zuverlässigkeit des Messinstruments und die Stabilität der Resultate. Um der Reliabilität zu genügen, müssen die Ergebnisse reproduzierbar sein. Das bedeutet, dass bei einer Wiederholung mit identen Methoden und gleichen Rahmenbedingungen auch andere Personen gleiche Ergebnisse erhalten. Festgestellt wird die Reliabilität auf mit Hilfe des Intercoder-Verfahrens.

Die Intercoder-Reliabilität (ICR) wird eingesetzt, um die Übereinstimmung von Codes bei nominal skalierten Daten zu quantifizieren. Ziel ist es, die Übereinstimmung der Code-Zuordnungen für einen Datensatz zu überprüfen, die von mehreren Codern vorgenommen wurde (O'Connor & Joffe, 2020). Wie bereits in Abschnitt 2.5 beschrieben, kann die Intercoder-Reliabilität direkt in MAXQDA bestimmt werden, wobei auch das Fleiss' κ berechnet wird. Jede Coder:in hat den gesamten Datensatz eines anderen Coders überprüft, wobei die jeweiligen Code-Zuordnungen berücksichtigt wurden. Durch dieses Verfah-

ren konnten alle Codes zu 100% auf ihre Reliabilität überprüft werden. Die Berechnung erfolgt wie folgt:

$$\kappa = \frac{(p_0 - p_c)}{1 - p_c} \quad (5)$$

wobei p_0 die beobachtete Übereinstimmung und p_c die erwartete Übereinstimmung darstellt. Da die Anzahl der Codes pro Segment in den Interviews variierte, wurde p_c mit der folgenden Formel berechnet:

$$p_c = \frac{\text{Anzahl der Codes}}{(\text{Anzahl der Codes} + 1)^2} \quad (6)$$

Für jede Coder:in wurden unterschiedliche Werte für p_0 und p_c ermittelt, die jeweils mit einer Unsicherheit von 10% gemittelt wurden. Aus den Mittelwerten von p_0 und p_c wurde dann das Fleiss' κ berechnet, wie in Formel 5 angegeben (Pelz & Schnögl, 2014). Das Ergebnis für κ war:

$$\kappa = 0,8891 \pm 0,0487 \quad (7)$$

Landis und Koch (1977) stellten eine Tabelle auf, anhand derer die Reliabilität des Fleiss' κ abgelesen werden kann. In diesem Fall liegt die Reliabilität im Bereich „fast perfekt“, mit einem Wert von 0,81 bis 1,00 (Landis & Koch, 1977).

3.2 Forschungsethische Qualitätskriterien

Die Einhaltung forschungsethischer Standards bildet eine zentrale Grundlage für die Durchführung empirischer Studien, insbesondere im schulischen Kontext. In Anlehnung an Huber (2019) lassen sich forschungsethische Probleme in zwei Kategorien unterscheiden: jene, die den Zweck der Forschung betreffen, und solche, die sich auf die konkrete Durchführung der Datenerhebung beziehen (vgl. Huber, 2019, S. 201ff.).

Hinsichtlich der Zielsetzung der vorliegenden Untersuchung ist festzuhalten, dass sie einen Beitrag zur Erweiterung des fachdidaktischen Wissens im Bereich der Schülervorstellungen zu zufälligen Ereignissen und Messunsicherheiten leisten soll. Aufgrund der begrenzten Forschungslage in diesem Bereich erscheint die Relevanz des Vorhabens sowohl im Hinblick auf wissenschaftlichen Erkenntnisgewinn als auch auf praxisorientierte Implikationen gerechtfertigt (vgl. Bortz & Döring, 2006, S. 32).

Die zweite Kategorie forschungsethischer Anforderungen bezieht sich auf die Durchführung der Interviews. Zur Vermeidung psychischer Beeinträchtigungen wurde ein Gesprächsrahmen geschaffen, der es den Schüler:innen ermöglichte, eigene Vorstellungen ohne Leistungsdruck zu äußern. Bereits zu Beginn der Interviews wurde betont, dass es keine „richtigen“ oder „falschen“ Antworten gebe. Zudem erhielten die Teilnehmenden den Hinweis, dass ausschließlich ihre subjektiven Gedanken im Mittelpunkt stünden. Dadurch sollte einer möglichen Beeinträchtigung des Selbstwertgefühls – etwa durch Unsicherheiten im Zusammenhang mit der Bearbeitung der Interviewfragen – entgegengewirkt werden.

Die physische Unversehrtheit der Schüler:innen wurde durch die Wahl eines ruhigen Raums sowie die Möglichkeit zu Pausen jederzeit gewährleistet. Die Erhebung fand nach Möglichkeit in Supplierstunden statt, sodass keine inhaltlich relevanten Unterrichtsinhalte versäumt wurden.

Eine Täuschung der Teilnehmenden im Sinne einer unbewussten Beobachtung oder unvollständigen Information fand nicht statt. Die Schüler:innen sowie deren Erziehungsberechtigte wurden im Vorfeld schriftlich über das Ziel und den Ablauf der Untersuchung informiert. Zusätzlich erfolgte vor Ort eine mündliche Erläuterung der Erhebungssituation.

Zur Vermeidung von Einflussnahme auf die Aussagen der Befragten wurde ein halbstrukturierter Leitfaden verwendet, dessen Fragen neutral formuliert waren. Persönliche Sichtweisen der Interviewer:innen wurden bewusst zurückgehalten. Ziel war es, die individuellen Vorstellungen der Schüler:innen möglichst authentisch zu erfassen.

Die Teilnahme an der Studie erfolgte auf freiwilliger Basis. Zwar wurde die Auswahl der Teilnehmenden teilweise von den jeweiligen Lehrpersonen unterstützt, jedoch erfolgte die Befragung nur nach ausdrücklicher Zustimmung der Schüler:innen sowie deren Erziehungsberechtigter. Der entsprechende Elternbrief diente der Information und der Einholung der Einverständniserklärung.

Schließlich wurde die Anonymität der Teilnehmenden durch die Verwendung von Pseudonymen in den Transkripten sowie durch die datenschutzkonforme Vernichtung der Tonaufnahmen sichergestellt. Die reine Aufzeichnung der Tonspur ermöglichte zudem eine Vermeidung identifizierender visueller Daten.

Insgesamt wurde auf eine ethisch verantwortungsvolle Gestaltung der Studie geachtet, um sowohl die Integrität der Forschung als auch den Schutz der teilnehmenden Personen zu gewährleisten.

3.3 Pragmatische Qualitätskriterien

Pragmatische Qualitätskriterien zielen darauf ab, Forschungsprozesse so zu gestalten, dass sie in der Praxis, insbesondere im schulischen Kontext, mit vertretbarem Aufwand umsetzbar sind. Nach Altrichter et al. (Altrichter, Posch & Spann, 2018, S. 107) gewährleisten diese Kriterien, dass sowohl der Forschungsprozess als auch die eingesetzten Erhebungsinstrumente so gestaltet sind, dass sie ohne erheblichen Mehraufwand von Praktiker:innen angewendet bzw. begleitet werden können.

Dabei werden zwei zentrale Teilaspekte unterschieden:

- Praktische und zeitökonomische Kompatibilität mit dem Unterricht
- Zeitökonomische Verträglichkeit mit der beruflichen Situation von Lehrpersonen

Die praktische Kompatibilität mit dem Unterricht bezieht sich auf die Eingliederung der Erhebung in den schulischen Alltag, ohne diesen wesentlich zu beeinträchtigen. In der vorliegenden Studie wurde darauf geachtet, die Interviews mit einzelnen Schüler:innen jeweils außerhalb des regulären Unterrichtsgeschehens durchzuführen, um Störungen des Klassenverbandes und des Lernprozesses möglichst gering zu halten. Pro Schule wurden nur wenige Interviews durchgeführt. Darüber hinaus wurde, sofern möglich, auf Supplierstunden zurückgegriffen, wodurch der reguläre Unterricht nicht unterbrochen werden musste.

Im Hinblick auf die berufliche Situation der beteiligten Lehrpersonen wurde das Verfahren so gestaltet, dass der zusätzliche organisatorische Aufwand minimiert werden konnte. Der Beitrag der Klassenlehrer:innen beschränkte sich auf das Verteilen von Informationsschreiben und das Einholen der Einverständniserklärungen. Sämtliche weiteren Materialien und die Durchführung der Interviews wurden von der forschenden Person vorbereitet und eigenständig durchgeführt. Das gewählte Vorgehen erfüllte somit in hohem Maße die Anforderungen an eine zeitökonomisch verträgliche Forschungsgestaltung.

4 Datenauswertung

Im Zuge der Bachelorarbeit von Frau Fleissner, Frau Hartmann, Frau Schaubmaier und Herrn Lechmann wurden insgesamt 26 Leitfadeninterviews durchgeführt. Im folgenden Kapitel geht um die Definition der Kategorien, welche gemeinsam gebildet wurden und um die daraus erhaltenen Ergebnisse.

4.1 Definition der Kategorien

Wie bereits in Kapitel 2.6 erklärt wurden sowohl induktiv als auch deduktiv Kategorien gebildet, welche bei der Auswertung der Interviews angewendet wurden. Aus der Theorie konnten, zum Wissensstand der Erstellung, drei wesentliche Kategorien deduktiv gewonnen werden, aus den Interviews ergaben sich weitere sechs induktiv gewonnene Kategorien. Mit Hilfe des Codierleitfadens (vgl. Kapitel 2.7) wurden die Kategorien definiert und eingegrenzt.

induktiv	deduktiv
1 – Ähnliches gehört zusammen	7 – Der am häufigsten vorkommende Wert ist der Genaueste.
2 – Der mittlere Wert ist der beste.	8 – Technik ist besser als der Mensch.
3 – Bekanntem vertraut man mehr.	9 – Messunsicherheit kommt von menschlichen Fehlern.
4 – „Schöne“ Zahlen sind genauer.	
5 – Messwerte situationsbedingt auswählen	
6 – Äußere Einflüsse als Erklärung der Schwankung	

Tabelle 1: Induktive und deduktive Annahmen im Messkontext

1 – Ähnliches gehört zusammen	
Definition:	Ähnliche Werte werden zu einer Gruppe zusammengefasst und im Vergleich zu Ausreißern als richtiger betrachtet.
Codierregel:	Antworten dieser Kategorie berücksichtigen Messwerte, die nahe beieinander bzw. in einem gewissen Bereich liegen. Werte in dieser selbstdefinierten Umgebung werden als genauer betrachtet und Ausreißer werden ausgeschlossen. Zu den wichtigen Schlüsselbegriffen zählen „Bereich“, „ähnlich“, „(großer) Unterschied“ oder „zusammen passen“. Nicht inkludiert werden Aussagen, welche sich auf denselben Wert beziehen.
Ankerungsbeispiele	Eine Schüler:innenaussage, die sich auf den Umgang mit Ausreißern bezieht, ist folgende: <i>„Also die anderen sind ziemlich beieinander und dann ist der eine so weit, also im Gegensatz zu einer anderen weit entfernt. Deswegen würde ich den hier ja. Rauskürzen“</i> (Interview 8). Eine ähnliche Denkweise ist auch in <i>„wobei ich jetzt das 0,63 weglasse, da ich das als falsche Wertung sozusagen für mich nehme“</i> (Interview 22) erkennbar. Es wird deutlich, dass die Lernenden Ausreißer entweder komplett aus ihrer Messung streichen oder ihnen eine geringere Bedeutung zuschreiben wie <i>„und deswegen hätte ich gesagt, dass sie nicht so wichtig sind“</i> (Interview 7) verdeutlicht. Viele Schüler:innen können sich Messwerte, die weiter entfernt liegen, nicht erklären und greifen folglich auf Erklärungsansätze zurück, die sich auf menschliche Messfehler, beispielsweise <i>„dass man da irgendwie versehentlich gedrückt hat“</i> (Interview 12) beziehen. Weitere wichtige Ankerungsbeispiele sind: <i>„Ja, weil alle anderen Werte sind unter 1 Sekunde. Und es ist halt komisch, dass 1,64 und 1,43 dabei sind“</i> (Interview 3) oder <i>„dass die Messwerte relativ in einem gleichen Bereich sind, also zwischen 1,7 und 1,8 zum Beispiel“</i> (Interview 4).
Abgrenzungsbeispiele	Aussagen, die im Zusammenhang mit der Akzeptanz unterschiedlicher Messwerte genannt werden und beispielsweise mit <i>„weil es fast immer dieselben Zahlen sind, ist das okay“</i> (Interview 5) begründet werden, zählen zu dieser Kategorie, obwohl sie im Überschneidungsbereich mit 7—Der am häufigsten vorkommende Wert ist der Genaueste liegen. Die Ähnlichkeit der Messwerte ist für viele Schüler:innen folglich die Voraussetzung, um die verschiedenen Messwerte zu legitimieren.

2 - Der mittlere Wert ist der Beste	
Definition:	Mittelwerte oder Mediane werden als richtig angesehen
Codierregel:	Antworten dieser Kategorie beziehen sich auf Werte, die in der Mitte einer Messreihe liegen bzw. durch eine Mittelwertsberechnung erhalten wurden. Schlüsselwörter sind beispielsweise „ <i>Mittelwert</i> “, „ <i>in der Mitte/mittig</i> “ oder „ <i>dazwischen</i> “.
Ankerungsbeispiele	Als wichtige Beispiele werden folgende genannt: „ <i>Weil der Wert dazwischen liegt. Man könnte sagen, dass der halbwegs genau ist</i> “ (Interview 6), „ <i>Es ist einfach wieder optisch die Mitte, also wirklich die Mitte</i> “ (Interview 9) oder „ <i>Also ich würde erstens alles zusammenzählen und dann es durch fünf dividieren, weil da fünf sind. Dann habe ich einen Durchschnitt und den würde ich halt nehmen</i> “ (Interview 2). Die Aussagen der Schüler:innen zeigen, dass sie einen Wert wählen, der optisch ungefähr in der Mitte liegt, jedoch bezieht sich die/der Lernende lediglich in einem Fall explizit auf den Median (Interview 22). In einigen Fällen begründen die Schüler:innen ihre Entscheidung mit einer Mittelwertsberechnung oder argumentieren mit „ <i>weil es nicht zu klein und nicht zu groß ist</i> “ (Interview 14). Das bedeutet, dass die Schüler:innen die Größe der Werte beurteilen und daraus ihre Entscheidung ableiten. Im Gegensatz zu den Ergebnissen der Analyse von 1–, „ <i>Ähnliches gehört zusammen</i> “ scheinen einige Schüler:innen Ausreißer durch das Auswählen eines Mittelwertes zu berücksichtigen wie „ <i>es gibt halt welche die irre drüber liegen</i> “ (Interview 8) zeigt.
Abgrenzungsbeispiele	Es ist bei dieser Kategorie nicht notwendig, ein Abgrenzungsbeispiel anzugeben, da die Aussagen der Schüler:innen unmissverständlich zugeordnet werden können.

3 - Bekanntem vertraut man mehr	
Definition:	Messwerte, die mit bekannten Messgeräten festgestellt werden, werden ausgewählt.
Codierregel:	Messgeräte, welche den Schüler:innen bekannt sind, werden als vertrauenswürdiger eingestuft. Wichtige Aussagen beziehen sich auf Beispiele aus dem Alltag und enthalten Formulierungen wie „ <i>das kenne ich</i> “, „ <i>das habe ich zuhause</i> “ oder „ <i>bekannt</i> “.
Ankerungsbeispiele	Schüler:innen-Antworten dieser Kategorie beziehen sich hauptsächlich auf die Temperaturmessung und spiegeln sich in Aussagen wie „ <i>Das ist das einzige Thermometer von denen, die ich zu Hause habe</i> “ (Interview 10), „ <i>Ähm, weil wir benutzen, glaube ich das am meisten</i> “ (Interview 14) oder „ <i>Also das kenne ich und das ist meistens dann richtig</i> “ (Interview 17) wider. Es wird deutlich, dass die Lernenden bei Erklärungen zu dieser Kategorie auf ihre Alltagserfahrungen zurückgreifen wie „ <i>weil ich dann genau gleich weiß ja, wenn der Wind geht, dann wird es halt ein bisschen kälter</i> “ (Interview 8) zeigt, oder die Lernenden Situationen aus ihrem praktischen Umfeld nennen, beispielsweise „ <i>weil das sind diese fürs Kochen die müssen auch sehr genau sein, weil sonst kann ja irgendwas anbrennen</i> “ (Interview 22). Außerdem ist die Vertrauenswürdigkeit eines Messgerätes größer, wenn dieses in ihrem Umfeld öfter verwendet wird, bzw. öfter vorkommt, was unter anderem aus „ <i>sind ganz oft irgendwo außen am Balkon</i> “ (Interview 8) hervorgeht.
Abgrenzungsbeispiele	

4 - „Schöne Zahlen“ sind genauer	
Definition:	Messwerte ohne Nachkommastelle(n) werden ausgewählt.
Codierregel:	Messwerte werden auf ganze Zahlen gerundet, bzw. ganze Zahlen werden ausgewählt. Schüler:innen erachten Messwerte, die keine Nachkommastelle(n) haben als genauer. Charakteristisch sind Aussagen, welche die Wörter „gerade“, „nächstgrößte“ und „ganze Zahlen“ beinhalten.
Ankerungsbeispiele	Einer eigenen Interpretation zufolge bezieht sich die Schüler:in mit der Aussage „gerade Zahl“ nicht auf die mathematische Definition gerader und ungerader Zahlen, sondern er/sie versteht darunter gerundete Werte wie das Beispiel <i>„Weil es genau ist. Und wenn zum Beispiel steht, dass ich etwas messen muss und es steht 10,04, dann würde ich eher die 10,05 nehmen. Weil es eine gerade Zahl ist“</i> (Interview 14) verdeutlicht. Das Auswählen gerundeter Werte zeigt auch die Antwort <i>„weil zehn die nächstgrößte Zahl ist“</i> (Interview 1), welche diese Schüler:in mit folgender Situation aus dem Schulalltag begründet: <i>„Zum Beispiel, wenn ich mit dem Lineal etwas abmesse. Oder wenn ich in Mathe etwas rechne, würde ich auch auf das Nächstgrößere runden“</i> (Interview 1). Außerdem wählen gewisse Interviewte Werte ohne Kommastelle(n) aus, was die Antwort <i>„Weil es eine genaue Zahl ist und da nicht so 10,4 oder so steht“</i> (Interview 2) zeigt. Interessant bei dieser Aussage ist, dass die Schüler:in diesen Wert zwar mit dem Begriff „genau“ verknüpft, auf Nachfrage aber behauptet, es gäbe keinen genauen Wert und die Entscheidung mit einem Bauchgefühl verbindet.
Abgrenzungsbeispiele	

5 - Messwerte situationsbedingt auswählen	
Definition:	Je nach Zweck der durchgeführten Messung wird der für die Situation geeignete Wert ausgewählt.
Codierregel:	Für die Situation passende Messwerte werden ausgewählt, je nach Einsatzbereich des Messgeräts, sprich Verwendungszweck des Thermometers oder Anwendung der Messwerte, beispielsweise Längen- oder Zeitmessung. Außerdem umfasst die Kategorie Aussagen, die keinem Prinzip folgen. Schlüsselwörter sind: „ <i>kommt drauf an</i> “, „ <i>je nach dem</i> “.
Ankerungsbeispiele	Antworten bei denen die Schüler:innen die Messwerte je nach Situation auswählen oder bei denen sie immer unterschiedliche Messwerte wählen, gehören dieser Kategorie an. Beispielsweise, dass sie je nach Verwendungszweck des Thermometers, einen anderen Wert auswählen wie „ <i>Für welchen Zweck, ja. Das für Erde. Das für Kuchen. Das würde ich für Körpertemperatur. Das für Luft.</i> “ (Interview 19) oder „ <i>Vielleicht ist eines fürs Fieber (zeigt auf Infrarotthermometer) messe und das andere ist vielleicht eben für die Industrie (Multimeter gemeint) oder so. Oder sowas habe ich schon mal beim Kochen gesehen oder so (rotes weißes digitales Thermometer)</i> “ (Interview 6) zeigen. Auch bei der Längen- und Zeitmessung wählen die Schüler:innen den für die Situation passenden Wert aus, entweder immer den ersten, größten oder kleinsten Wert, oder jedes mal einen anderen Wert. Eine Begründung für das situationsabhängige Auswählen wäre „ <i>Ja, ich habe mit meinem Maßband manchmal gemessen. Beispielsweise wenn wir jetzt einen neuen Kasten gekauft haben, habe ich immer auf die Länge geschaut, dass sie sich überhaupt in mein Zimmer oder so ausgeht. Weil wenn, wenn nicht, dann ist er zu groß und dann passt er nicht ins Zimmer, weil wenn ich den kleineren Wert nehme, kanns ein dass es nicht passt.</i> “ (Interview 19).
Abgrenzungsbeispiele	

6 - Äußere Einflüsse als Erklärung der Schwankung	
Definition:	Unterschiedliche Messwerte werden damit begründet, dass sich die Bedingungen während des Messvorgangs verändern.
Codierregel:	Zu dieser Kategorie zählen beispielsweise physikalische Erklärungen wie äußere Krafteinwirkungen, Ungenauigkeiten des Messgerätes oder der Versuchsanordnung. Schlüsselwörter sind: „ <i>Luftwiderstand</i> “, „ <i>Wind</i> “, „ <i>Reibung</i> “ oder „ <i>langsamer</i> “.
Ankerungsbeispiele	Schüler:innen begründen die Schwankung der Messwerte aufgrund von Veränderungen äußerer Parameter wie beispielsweise „ <i>Vielleicht...Weil...Bei einer frischen Luft wahr und bei anderen alten Luft.</i> “ (Interview 1), oder „ <i>Dass es bei der Erdanziehung so ganz leichte Schwankung gibt.</i> “ (Interview 4) zeigen. In diese Kategorie zählen auch Aussagen, welche sich auf Veränderungen der Versuchsanordnung beziehen wie „ <i>Eigentlich soll hier das Gleiche herauskommen, nur dass die Lichtschranke vielleicht ungenau ist oder auf der Kugel ein Staubkorn drauf ist oder sonst irgendwas.</i> “ (Interview 11). Weiters glauben die Schüler:innen, dass die Schwankung vom Luftwiderstand verursacht wird und begründen dies mit „ <i>Ja, schon, weil bei Dachziegeln, wenn die runterfallen, kommt drauf wie sie halt fliegen. Wenn sie so waagrecht runterfliegen, ist die Geschwindigkeit geringer, als wenn sie so senkrecht runterfliegen, dann sausen sie halt runter.</i> “ (Interview 19).
Abgrenzungsbeispiele	Aussagen, welche die Schwankung aufgrund der Unsicherheit des Messgerätes, gehören auch zu dieser Kategorie, wie beispielsweise „ <i>Ich denk mir halt, diese Schraube ist ein bisschen unförmig oder dieses Messgerät funktioniert nicht gscheit.</i> “ (Interview 11). Aussagen, die beinhalten, dass Messgeräte im Vergleich zum Mensch weniger Unsicherheiten mit sich bringen, gehören zu Kategorie 8.

7 - Der am häufigsten vorkommende Wert ist der Genaueste	
Definition:	Der Wert wird ausgewählt, weil er mehr als einmal vorkommt.
Codierregel:	Antworten dieser Kategorie enthalten Häufigkeiten zum Vorkommen eines Wertes oder eine genannte Anzahl. Wichtige Schlüsselwörter sind „ <i>mehrmals</i> “, „ <i>häufiger</i> “, „ <i>oft</i> “, „ <i>gleich</i> “ oder Variationen dieser Begriffe. Eine klare Abgrenzung zur Kategorie 1 – Ähnliches gehört zusammen ist gegeben, indem eine Antwort mit Hinweis auf die Ähnlichkeit der Messwerte unzureichend ist. Jedoch werden dieser Kategorie Aussagen zugeordnet, die sich auf ausgewählte Werte in der Nähe eines häufig vorkommenden Wertes beziehen oder aufgrund einer gleichen Ziffer der ersten Kommastelle ausgewählt werden.
Ankerungsbeispiele	Hierzu zählen die Beispiele „ <i>0,82, weil auch öfters vorkommt hätte ich jetzt gesagt, weil das ist dann quasi am öftesten</i> “ (Interview 7) oder „ <i>Weil der gleiche Wert dreimal vorkommt</i> “ (Interview 4), denn es wird die Definition der Kategorie klar unterstrichen und zeigt die Anwendung der Schlüsselbegriffe.
Abgrenzungsbeispiele	Als untypische Schüler:innenaussage dieser Kategorie ist beispielsweise „ <i>Weil es am nächsten bei 0,82 dran ist. Und weil ich glaube, dass 0,82 am nächsten beim richtigen dran ist</i> “ (Interview 4) zu nennen, denn der Lernende wählt den Wert einer Messreihe aus, der am nächsten beim häufigsten Wert einer anderen Messreihe liegt (in diesem Fall: 0,82 bei Messreihe 1 des Pendel-Versuches). Es findet bei einigen Interviewten somit ein Vergleichen und Kombinieren der Messreihen statt, das den Auswahlprozess beeinflusst. Außerdem beziehen sich einige Schüler:innen auf Werte, bei denen die erste Kommastelle gleich ist, sprich „ <i>ich würde 0,82 oder 0,8 nehmen, weil 0,8 am meisten vorangegangen ist</i> “ (Interview 2). Auffällig bei Aussagen dieser Kategorie ist, dass die Schüler:innen das mehrmalige Auftreten eines Wertes auf menschliche Ursachen wie „ <i>vermutlich hat man dann öfter richtig gestoppt</i> “ (Interview 4) zurückführen, was das Point-Verständnis unterstreicht.

8 - Technik ist besser als der Mensch	
Definition:	Messwerte, die ohne das Einwirken des Menschen erfasst werden, sind genauer.
Codierregel:	Diese Kategorie umfasst Aussagen, die Technik als weniger fehlerbehaftet darstellen als den Menschen. Sprich, durch das Heranziehen von computergestützten Messgeräten wie beispielsweise einer Lichtschranke wird ein genaueres Ergebnis erzielt als mit einer Stoppuhr. Es werden auch Aussagen, die beinhalten, dass modernere technische Geräte genauer messen, zu dieser Kategorie gezählt.
Ankerungsbeispiele	Beispielaussagen wie „ <i>Es müsste eigentlich die selbe Zeit rauskommen, weil die Maschine misst. Die misst sicher immer gleich</i> “ (Interview 12) und „ <i>Dem neuesten. Weil der ist irgendwie am genauesten. Weil immer desto neuer, misst es genauer.</i> “ (Interview 13) zählen zu dieser Kategorie.
Abgrenzungsbeispiele	Zu dieser Kategorie gehören keine Aussagen, welche die Schwankung aufgrund von menschlichen Fehlern erklären, sondern nur Aussagen, die behaupten, dass technische Messgeräte genauer sind oder keine Fehler machen. Ein Abgrenzungsbeispiel wäre „ <i>Weil diese Maschine schneller stoppt, als der Mensch.</i> “ (Interview 1).

9 - Messunsicherheiten kommen von menschlichen Fehlern	
Definition:	Messunsicherheit wird damit begründet, dass sich aufgrund von menschlichen Fehlern oder Ungenauigkeiten unterschiedliche Messwerte ergeben.
Codierregel:	Zu dieser Kategorie zählen Aussagen, die die verzögerte Reaktion des Menschen, die Konzentration und die unterschiedlichen Bedienungen der Messapparatur beinhalten. Schlüsselaussagen sind „zu früh/zu spät gedrückt“, „bessere Konzentration“, „zu fest gedrückt“, „anders gehalten“.
Ankerungsbeispiele	Schüler:innen erklären das Auftreten unterschiedlicher Messwerte mit ungenauem Messen, geringer Konzentration oder verzögerten Reaktionen des Menschen. Beispielaussagen wie „Weil es eben am Mensch liegt. Wegen der Reaktion. Und man sich da noch am meisten konzentriert.“ (Interview 2), „Weil man einfach nicht genau war beim Stoppen.“ (Interview 24), „Ja dass man es unterschiedlich ein dreht und die Schraube fester oder schwächer drückt.“ (Interview 17) und „Man kann ja nicht immer auf die Sekunde genau gleich Stopp drücken.“ (Interview 19) zählen zu dieser Kategorie.
Abgrenzungsbeispiele	Zu dieser Kategorie zählen nur Aussagen, die Schwankungen aufgrund von menschlichen Fehlern begründen.

Nach weiteren Gesprächen mit Experten wurden die Kategorien allerdings erweitert, da es bei einigen Interviews Erkenntnisse gab, die mit den bisherigen Kategorien nicht abgebildet werden konnten. Bei diesen Kategorien wurde explizit das Video zum freien Fall betrachtet und die Aussagen der Schüler:innen dazu dokumentiert.

4.1.1 Genauere Betrachtung des Fallversuch-Videos

Durch diese genauere Beobachtung der Antworten zum Video wurden weitere Kategorien gebildet, welche sich explizit mit diesen Antworten beschäftigen. Das Ziel dieser Kategorien ist es, herauszufinden wie die Schüler:innen auf den Einsatz von Technik reagieren und ob sich ihre Denkweisen dadurch verändern. Die Kategorien „Technik ist fehlerfrei“, „Technik ist nicht fehlerfrei“ und „Der mittlere Wert ist der Beste“ wurden in diesem Kontext nochmal explizit für das Video codiert und darauf geachtet welche Paradigmen die Schüler:innen anwenden.

Die Zuordnung zur Kategorie „Technik ist fehlerfrei“ wurde nur dann getätigt, wenn diese Aussage entweder direkt oder zumindest sinngemäß von den Schüler:innen getätigt wurde.

Dasselbe gilt für die Zuordnung zur Kategorie „Technik ist nicht fehlerfrei“.

Bei der Kategorie „Der mittlere Wert ist der Beste“ wurden alle Aussagen codiert, welche eine Berechnung des Mittelwerts anstreben oder einen möglichst zentral liegenden Wert auswählen.

4.2 Quantitative Analyse der Kategorien

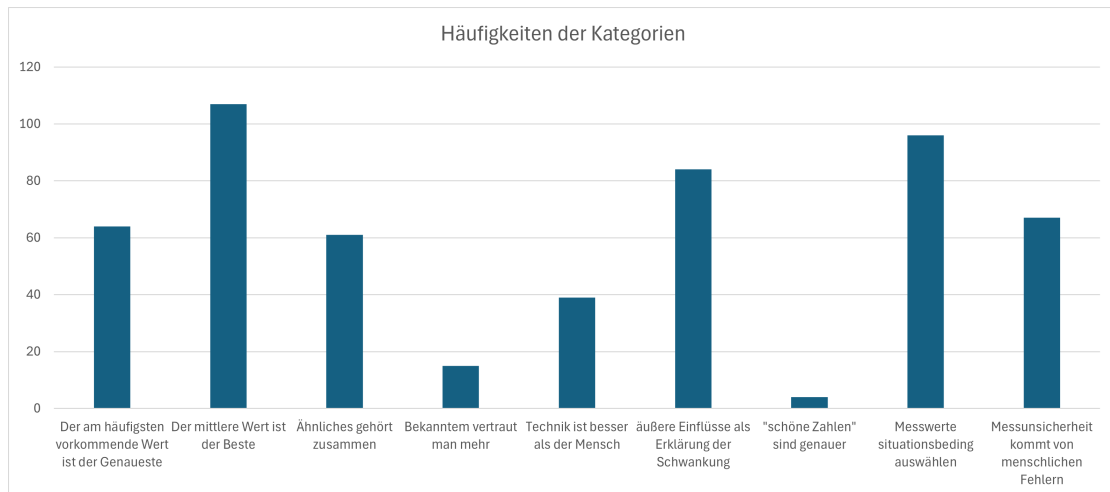


Abbildung 2: Häufigkeit der Kategorien

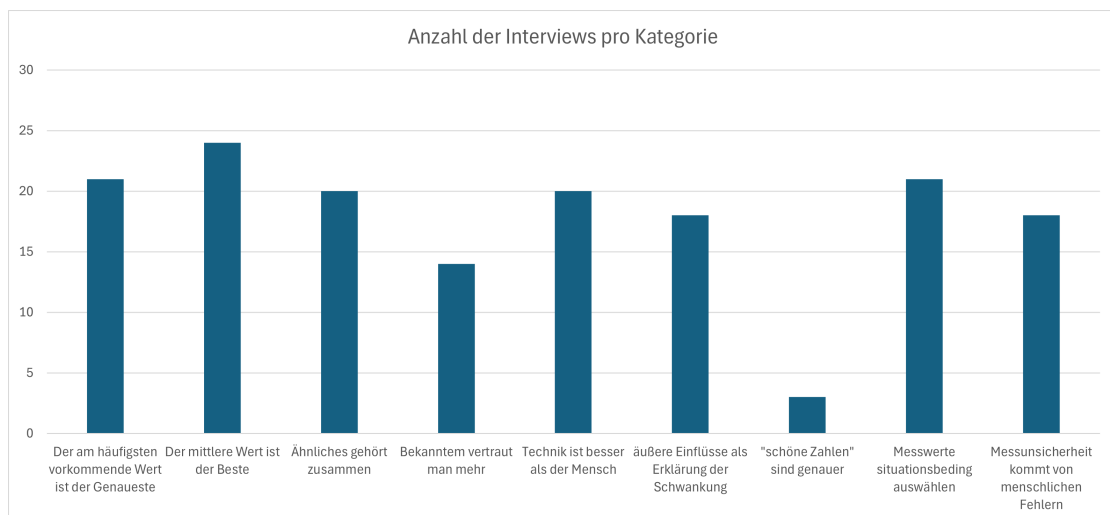


Abbildung 3: Anzahl der Interviews pro Kategorie

Der am häufigsten vorkommende Wert ist der Genaueste (n = 63)

Schüler:innen zeigten oft die Tendenz, Häufigkeit mit Richtigkeit gleichzusetzen. In 21 Interviews konnte dieses Konzept nachgewiesen werden, welches mit statistischem Denken verwandt ist jedoch nicht auf der kritischen Bewertung von Verteilungen beruht, sondern auf dem Eindruck von Wiederholung.

„Wenn ich zwei- oder dreimal denselben Wert hab, dann stimmt der bestimmt.“ (Interview 5)

„Wenn es öfter vorkommt, dann ist es eher richtig als die anderen.“ (Interview 11)

Der mittlere Wert ist der Beste (n = 107)

Eine der häufigsten Vorstellungen ist, dass der Mittelwert gleichbedeutend mit dem besten oder „richtigsten“ Wert sei. Dabei wird der Mittelwert nicht im mathematisch-statistischen Sinne verstanden, sondern vielmehr als intuitive Mitte, die gefühlt „ausgleichend“ wirkt.

„Der in der Mitte ist meistens der richtige. Die anderen sind eher zu hoch oder zu niedrig.“ (Interview 9)

„Wenn man viele Werte hat, muss man irgendwie schauen, was so der Durchschnitt ist. Das ist meistens der fairste Wert.“ (Interview 13)

Diese Denkweise, welche in 24 Interviews nachgewiesen werden konnte, spiegelt ein typisches Alltagsverständnis wider, in dem die Mitte als Maß für Ausgewogenheit dient – unabhängig davon, ob die Streuung groß oder klein ist. Der Mittelwert wird somit nicht als Ergebnis einer mathematischen Operation interpretiert, sondern als Kompromiss.

Ähnliches gehört zusammen (n = 61)

In 20 der 26 Interviews wurde deutlich, dass Schüler:innen stark auf Kohärenz von Messwerten achten. Werte, die „optisch“ oder logisch nicht zur Mehrheit passen, werden intuitiv ausgeschlossen:

„Die Zahl ist irgendwie komisch, die passt gar nicht zu den anderen – das war sicher ein Fehler.“ (Interview 2)

„Die drei Werte sind gleich, aber der vierte ist ganz anders – der zählt nicht.“ (Interview 10)

Abweichungen werden dabei automatisch als Fehler interpretiert – unabhängig von deren statistischer Bedeutung oder den Messbedingungen.

Bekanntem vertraut man mehr (n = 15)

Auch Vertrautheit spielt eine Rolle bei der Bewertung von Messergebnissen. Geräte, die aus dem Alltag bekannt sind, genießen höheres Vertrauen als unbekannte Apparaturen. Dieser Punkt wurde bei der Frage zum Thermometer thematisiert, wie man auch an den Zitaten merkt:

„Das Thermometer ist mir lieber – das kenn ich von zu Hause.“ (Interview 4)

„Wenn's so aussieht wie im Physiksaal, dann stimmt's meistens auch.“ (Interview 6)

Diese Kategorie konnte in 14 Interviews bestätigt werden.

Technik ist besser/genauer als der Mensch (n = 39)

Ein besonders spannendes Ergebnis ist das ausgeprägte Vertrauen in technische Systeme gegenüber menschlicher Leistung. In 20 Interviews zeigt sich eine starke Ausprägung dieser Kategorie. Dieses Vertrauen äußert sich durch Zuschreibungen wie „objektiver“, „exakter“ oder „fehlerfrei“:

„Die Maschine hat's halt richtig gemacht. Ich hätte sicher wieder falsch gedrückt.“ (Interview 8)

„Wenn das Gerät Kommastellen hat, dann weiß man, dass es genau ist.“ (Interview 12)

Diese Vorstellungen zeigen, dass Schüler:innen Technik mit höherer Autorität und Präzision verbinden. Kritisches Reflektieren über Gerätegenauigkeit oder systematische Fehlerquellen scheint kaum stattzufinden.

Äußere Einflüsse als Erklärung (n = 84)

In 18 der 26 Interviews benannten Schüler:innen, oftmals wiederholt, externe Ursachen für Abweichungen:

„Vielleicht war Wind oder der Tisch hat gewackelt – dann ist es anders geworden.“ (Interview 3)

„Die Kugel rollt nicht immer gleich, weil vielleicht was auf der Bahn war.“ (Interview 7)

Diese Aussagen zeigen ein wachsendes Verständnis dafür, dass Störfaktoren eine Rolle spielen können. Da in den durchgeführten Versuchen jedoch alle äußeren Störfaktoren so gut es ging eliminiert wurden, zeigt es auch, dass die Schüler:innen Erklärungen für Zufällige Abweichungen suchen.

Schöne Zahlen sind genauer (n = 4)

Ein seltener auftretendes Ergebnis ist das Vertrauen in „ästhetisch ansprechende“ Zahlen:

„10,00 sieht einfach richtiger aus als 9,87 – das merkt man doch.“ (Interview 6)

„Wenn's glatt ist, passt's auch meistens.“ (Interview 9)

Diese Form der Wahrnehmung verweist auf ein numerisches Schönheitsideal, das mit Präzision verwechselt wird. Wobei dazu gesagt werden muss, dass diese Aussage nur vier Mal, von insgesamt drei Schüler:innen getätigt wurde.

Messwerte situationsbedingt auswählen (n = 95)

Eine große Anzahl von Schüler:innen gaben an, ihre Auswahl von Messwerten situationsabhängig zu treffen:

„Wenn ich zeigen will, dass es schnell war,nehm ich den kleinsten Wert.“ (Interview 5)

„Je nachdem was rauskommen soll, kann man halt auch einen anderen nehmen.“ (Interview 13)

Die Auswahl erfolgt hier subjektiv und zielorientiert, nicht methodisch fundiert. Ein mehrmals erwähnter Vergleich ist der Kasten im Kinderzimmer, bei welchem das größtmögliche Maß des Kastens in die kleinste Abmessung des Zimmers passen muss.

Messunsicherheit als menschliches Versagen (n = 66)

18 Schüler:innen gehen davon aus, dass Messfehler primär auf eigene Unzulänglichkeiten zurückzuführen sind:

„Ich hab einfach zu spät gestoppt – also war der Wert dann falsch.“ (Interview 1)

„Wenn's nicht passt, hab ich was falsch gemacht, weil das Experiment war ja gleich.“ (Interview 14)

Solche Aussagen offenbaren ein Fehlermodell, das wenig Raum für die systemische oder zufällige Natur von Messunsicherheiten lässt. Die Schüler:innen empfinden Abweichungen nicht als „normal“, sondern als Ausdruck ihrer eigenen Unfähigkeit.

4.2.1 Detailanalyse zur Identifikation möglicher neuer p-prims

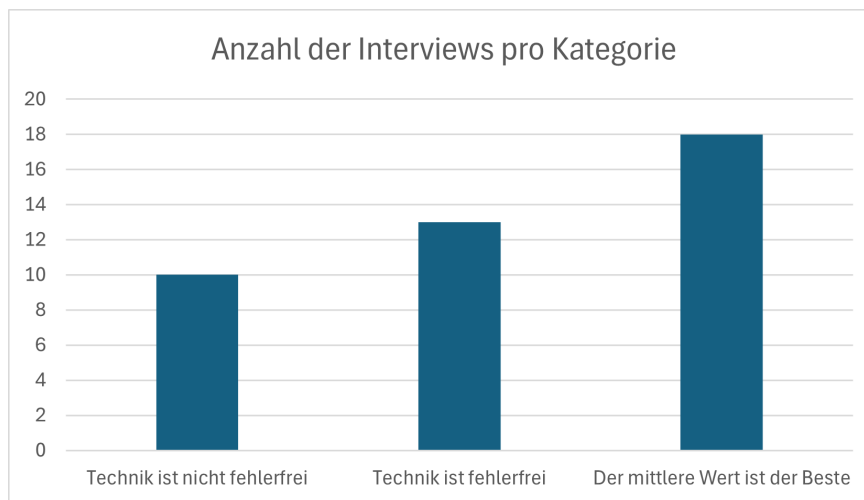


Abbildung 4: Anzahl der Interviews pro Kategorie

In 13 der 26 Interviews zeigen sich Evidenzen dafür, dass Schüler:innen davon ausgehen, dass die Technik – zumindest bei Messungen – fehlerlos arbeiten.

„Also, dass die Maschinen den genauen Wert misst. Also wirklich, genau ist... Also dass das ein sicherer Wert ist. Also ich würde persönlich gerne diesen Wert nehmen, weil ich genau da weiß, dass das richtig ist.“ (Interview 2)

„Ich müsste eigentlich dieselbe Zeit rauskommen, weil die Maschine misst. Die misst sicher immer gleich.“ (Interview 12)

Diese expliziten Aussagen zeigen, dass die Schüler:innen erwarten, dass bei wiederholten Messungen stets dasselbe Ergebnis liefern.

In 10 dieser 13 Interviews wurde in weiterer Folge trotzdem der Mittelwert der Messreihe gewählt. Was interessant ist, da dies impliziert, dass das set-paradigma akzeptiert wird. Denn nur so ist es möglich, dass man die Messwerte als Messreihe betrachtet und mit Hilfe eines Mittelwertes den Wert auswählt. Dies ist besonders spannend, weil diese Schüler:innen durch ihre Aussagen, welche von gleichbleibenden Ergebnissen bei mehreren Messungen ausgehen, eigentlich dem point-Paradigma zugewendet sind/waren.

In zehn Interviews ist aus den Aussagen der Schüler:innen herauszulesen, dass die Technik als fehlerlos angesehen wird. Dies zeigt sich daran, dass die Schüler:innen nicht davon ausgehen, dass bei der Messung mit Lichtschranken auch Differenzen der Messwerte auftreten werden. Diese Abweichungen werden vorwiegend auf äußere Einflüsse zurückgeführt.

„Ich glaube vielleicht wirklich, dass der Luftwiderstand bei einem bissl höher war als bei den anderen.“ (Interview 9)

„Eigentlich soll hier das Gleiche rauskommen, nur dass die Lichtschranke vielleicht ungenau ist oder auf der Kugel ein Staubkorn drauf ist oder sonst irgendwas.“ (Interview 11)

Die beiden Aussagen zeigen zwei mögliche Erklärungen der Schüler:innen.

In drei Interviews zeigt sich wiederum, dass die Schüler:innen der Messtechnik zwar eine höhere Präzision und Richtigkeit zuschreiben und somit davon ausgehen, dass die Messungen geringere Streuungen aufweisen als Messungen, welche von Menschen durchgeführt wurden. Der wesentliche Punkt ist, dass diese drei Schüler:innen nicht davon ausgehen, dass die Messungen vollkommen fehlerfrei sind.

Insgesamt kommt es in 18 der 26 Interviews zu der Aussage, dass der mittlere Wert der Beste ist. Dies zeigt sich daran, dass bei Messreihen entweder der Mittelwert gebildet wird oder ein möglichst zentraler Wert ausgewählt wird.

4.3 Interpretation der Ergebnisse

Anhand der in 4.2 diskutierten Datensätze erkennt man, dass ein Großteil der Kategorien mit einem sehr hohen Anteil von Schüler:innen angewendet werden. Im weiteren werden bei den einzelnen Kategorien mögliche p-prims identifiziert oder bereits bekannte p-prims erläutert.

Ähnliches gehört zusammen

Es konnte gezeigt werden, dass Schüler:innen dazu neigen ähnliche Messwerte als zusammengehörig betrachten und Ausreißer zu vernachlässigen. Dieses Phänomen hängt eng mit der Kategorie „Der am häufigsten vorkommende Wert ist der genaueste“ zusammen, da Lernende versuchen wiederkehrende Werte zu erzielen. Nur dann gilt ein Experiment für die Schüler:innen als erfolgreich. Das dahinterstehende p-prim kann als Ähnlichkeits-Prinzip - oder wie es bei diSessa heißt „Similarity Clustering“ (diSessa, 1993) - definiert werden. Dieses Prinzip ist eine universelle Strategie, die sowohl in der Optik, Akustik als auch der Mechanik ihre Anwendung findet. Vor allem aber findet man es allerdings im Alltag, wenn ähnliche Objekte oder Meinungen zusammengefasst werden.

Bekanntem vertraut man mehr

Wenn Schüler:innen bei der Messung von Daten die Wahl gelassen wird, dann wählen diese vermehrt jene Messgeräte, im Fall dieser Studie jenes Thermometer, welches sie kennen. Nach weiteren Literaturrecherchen zu diesem durchwegs alltäglichen Denkmuster konnte dieses als Vertrautheitsprinzip definiert werden kann, was jedoch nicht als p-prim im eigentlichen Sinn angesehen wird (diSessa, 1993). Es be-

schreibt, dass bekannte Objekte, Methoden und Werte als verlässlicher und richtiger angesehen werden als Unbekannte. Diese Vertrautheit spielt eine wesentliche Rolle bei der Aktivierung der p-prims.

„Schöne“ Zahlen sind genauer

Die bereits in der Literatur definierte Vertrautheit in „schöne“ Zahlen konnte bei dieser Studie nur in wenigen Interviews identifiziert werden. In der Literatur ist diese Neigung meist als Heuristik oder Bias definiert (vgl. Honda, Kagawa & Shirasuna, 2022). Ein p-prim, welches dieses Phänomen beschreibt, ist in der Literatur noch nicht beschrieben. Was daran liegt, dass diese Vorstellung viel mehr zum allgemeinen Erfahrungsschatz zählt, den die Lernenden mitbringen. Im Alltag, aber auch im Schulkontext wird häufig auf „schöne Zahlen“ gerundet um diese einfacher vergleichen zu können beziehungsweise einfacher mit diesen rechnen zu können.

Messwerte situationsbedingt auswählen

Schüler:innen gehen nachweislich selektiv mit Messdaten um. Je nach Situation werden die Messwerte unterschiedlich ausgewählt. Dieses Erklärungsmuster ist noch nicht explizit als p-prim definiert, jedoch wird es im Rahmen der Coordination-Class-Theorie (diSessa & Sherin, 1998) als eine der Readout-Strategien erwähnt. Dort wird diese Denkweise als Situational-Sampling p-prim definiert, was so viel heißt wie Situative Messwertauswahl.

Äußere Einflüsse als Erklärung für Schwankungen

Anstatt Schwankungen in der Messreihe als natürlich anzusehen, suchen Lernende nach konkreten äußeren Einflüssen und Erklärungen für diese. Diese Vorstellung beruht unter anderem darauf, dass man bei Messungen versucht die äußeren Einflüsse durch geeignete Messtechniken zu reduzieren. Die Schüler:innen gehen in weiterer Folge davon aus, dass dies bei den Messungen nicht ideal gelingt. In diesem Fall steckt das p-prim „von nichts kommt nichts“ hinter dem Erklärungsmuster. DiSessa bezeichnete dieses p-prim als „Actuating agency“ (diSessa, 1993).

Der am häufigsten vorkommende Wert ist der Genaueste

Wie bereits bei „Bekanntem vertraut man mehr“ erklärt haben Schüler:innen den Drang zu möglichst einheitlichen Ergebnissen. Dabei spielt auch die Vorstellung des „wahren Wertes“ (Referenz) mit hinein. In der Studie konnte gezeigt werden, dass diese Vorstellung unter den Schüler:innen weit verbreitet ist. Ein klares p-prim ist für diese Strategie nicht definiert, als Mustererkennung und Mehrheitsheuristik ist sie jedoch bereits bekannt.

Es handelt sich allerdings um ein intuitives Wissenselement, welches darauf hinweist, dass Lernende häufig auftretenden Ereignissen mehr vertrauen, weil es als stabiles Muster erscheint.

Eine weiterer Erklärungsansatz könnte ein Zusammenhang mit „Ohm’s p-prim“ von diSessa (diSessa, 1993), bei welchem die Hauptaussage *More effort implies more result*. ist, welche auf „Viel hilft viel“

vereinfacht werden kann, was in diesem Kontext durchaus passend ist.

Aus diesem Grund ist eine Definition eines neuen p-prims nicht zielführend, vielmehr wirkt das „Ohm's p-prim“ im gegebenen Kontext als „Häufigkeit gleich Wahrheit“.

Messunsicherheit kommt von menschlichen Fehlern

Gerade bei Schüler:innen ist diese Vorstellung weit verbreitet, weil diese in vielerlei Hinsicht das richtige Messen erst lernen und somit auch noch mehr Ungenauigkeiten im Messprozess haben. Die Lernenden denken, dass bei perfekten Bedingungen und einer sorgfältigen Durchführung ein perfektes Ergebnis erhalten wird. Dies kann auch daran liegen, dass der Begriff „Messfehler“ als „Fehler beim Messen“ verstanden wird. Das hinter dieser Vorstellung stehende p-prim wird meist als „von nichts kommt nichts“ bezeichnet und steht dafür, dass jede Wirkung eine Ursache haben muss – in diesem Fall eben der menschliche Fehler (Heinicke, 2012).

Technik ist besser als der Mensch

Im eigentlich definierten Sinn dieser Kategorie, also dass Technik genauer ist als Menschen, zeigt sich, dass diese Vorstellung in der Literatur bereits mehrmals thematisiert wird. Auch hier spielt wieder die bereits erwähnte Vorstellung des „wahren Werts“ mit hinein und auch der Gedanke, dass man bei sorgfältiger Messung mit geeigneter Messtechnik beliebig genau messen kann. Schüler:innen gehen davon aus, dass Maschinen bei reproduzierbaren Tätigkeiten die menschlichen Schwächen eliminieren kann.

In der tiefergehenden Analyse des Videos zum Fallversuch wird allerdings auch in dieser Kategorie eine extremere Richtung eingeschlagen. Es wird nicht mehr darauf geachtet ob die Schüler:innen den rein technischen Messaufbau als genauer als die menschliche Messung betrachten, sondern ob sie die (digitale) Technik generell als fehlerlos ansehen.

Diese These ist in dieser Form in der Literatur nicht als p-prim definiert und wird in dieser Interview Studie, wie im Kapitel 4.2 zu sehen ist jedoch sehr gut bestätigt. Die Alltagserfahrungen zeigen, dass Lernende intuitiv das Schema „Ein Messgerät zeigt den richtigen Wert an“ anwenden und der Meinung sind, dass das Gerät, das tut, was es tun soll, und zwar fehlerfrei, vor allem wenn das Messgerät mehrere Nachkommastellen vorweist (Goertz & Heinke, 2021, S. 665). Aus diesen Aspekten liegt eine Definition eines neuen p-prims nahe, welches hier als „(digitale) Technik ist fehlerlos“ definiert wird. Dieses p-prim soll die bereits erwähnten Vorstellungen Rechnung tragen und auch jene Alltagserfahrung einschließen, dass Maschinen konsistent und ohne Ermüdung arbeiten.

Der mittlere Wert ist der Beste

Eine häufig beobachtete Vorstellung ist jene, dass der mittlere Wert der Beste ist. Viele Schüler:innen verwenden den Durchschnitt intuitiv als „wahren Wert“ ohne das Konzept der Messunsicherheit zu verstehen. In den Interviews wurde selten der Mittelwert berechnet, viel mehr wurde ein möglichst zentraler Wert der Messreihe gesucht. Dieses Prinzip ist in dieser Form nicht als p-prim definiert. Ein möglicher

Querverweis mit einem bekannten p-prim ist jenes zur Balance. Dieses p-prim beschreibt, dass Systeme im Gleichgewicht nach einer Störung von selbst wieder in die Stabilität – umgangssprachlich auch oft in die „Mitte“ genannt – zurück. Stabilität entsteht demnach durch das Ausbalancieren entgegengesetzter Einflüsse. Eine Erweiterung dieses Denkmusters auf Messungen könnte lauten, dass sich „entgegengesetzte Fehler ausgleichen“.

Im Kontext der Interviews kann jedoch davon ausgegangen werden, dass die Schüler:innen, welche die Technik als unfehlbar anerkennen, wohl davon ausgehen, dass diese entgegengesetzten Fehler zufällig entstehen. Diese Schlussfolgerung kann gezogen werden, da die Schüler:innen keine anderwertigen schlüssigen Erklärungen für diese „Fehler“ nennen können. Und auf Basis dessen kann schlussgefolgert werden, dass Schüler:innen beim Auftreten von zufälligen Ereignissen dazu tendieren den mittleren Wert zu wählen.

Darüber hinaus haben Menschen einen inneren Drang nach Symmetrie, welcher bei der Auswahl des mittleren Wertes sicher auch mitspielt. Dies kann jedoch als Heuristik angesehen werden.

Auf Basis der erhaltenen Daten und der Literaturrecherche liegt es nahe ein neues p-prim zu benennen, welches die Intuition „der mittlere Wert ist der beste Wert“ abbildet. Dieses wird in weiterer Folge „die Mitte ist am Besten“ benannt. Es dient somit als Gegenpol zum bereits bekannten p-prim „mehr ist besser“.

5 Diskussion und Ausblick

Es zeigt sich, dass Schüler:innen den Zufall bei Messunsicherheiten nicht (ohne gezielte Konfrontation) als Konzept zur Erklärung der subjektiven Realität heranziehen und versuchen verschiedenste Erklärungen für die Schwankungen zu finden. Zum einen erkennt man, dass es eine weit verbreitete Vorstellung ist, dass Abweichungen durch menschliche Fehler entstehen. Schüler:innen davon ausgehen, dass sie selbst für die Abweichungen verantwortlich sind. Im Umgang mit Messtechnik zeigt sich, dass die Lernenden den erhaltenen Messdaten mehr vertrauen als den eigenhändig gemessenen Werten. Ein großer Anteil geht sogar davon aus, dass Maschinen fehlerlos sind und somit keine potenzielle Quelle für Streuungen sein kann. Diese, aber auch weitere Schüler:innen suchen Erklärungen für die Abweichungen in äußeren Einflüssen.

Beim Umgang mit den erhaltenen Messwerten zeigen sich zwei Strategien. Zum einen werden Ausreißer als grundlegend falsche Werte identifiziert und somit ignoriert. Zum anderen erkennt man eine starke Tendenz dazu den mittleren Wert als den „repräsentativen Wert“ der Messung anzusehen. Es werden Messwerte teilweise aber auch abhängig von der Situation ausgewählt. Das bedeutet, dass bei Zeitmessungen häufig der kürzeste Wert gewählt wird und bei Längenmessungen tendenziell der längere Wert verwendet wird.

Der Kontrast aus eigenem Experiment und dem Video erwies sich als sehr guter Ansatz um die Gedankengänge der Schüler:innen nachzuvollziehen. So konnte gezeigt werden, dass die Schüler:innen beim eigenen Experiment die Abweichungen häufig bei sich selbst suchten. Beim Video konnten die Schüler:innen sich die Abweichungen nicht erklären, weil für viele von ihnen die Technik als fehlerlos ansehen.

Aus der Kategorie „Ähnliches gehört zusammen“ konnte auf das bereits bekannte Ähnlichkeitsprinzip oder auch „Similarity Clustering“ (diSessa, 2018) geschlossen werden, welches im Alltag vielfältig zum Einsatz kommt.

Die der Kategorie „Bekanntem vertraut man mehr“ zugeordneten Antworten zeigen, dass das Vertrauensprinzip (diSessa & Sherin, 1998) bei den Schüler:innen stark ausgeprägt ist.

Ein -prim zur Auswahl von „schönen Zahlen“ konnte in der Literatur nicht gefunden werden. Dort ist diese Auswahl meist als Heuristik oder Bias erwähnt, welcher laut dieser zum allgemeinen Erfahrungsschatz der Schüler:innen zählt.

Die situative Auswahl von Messwerten, welche aus den Daten der Kategorie „Messwerte situationsabhängig auswählen“ hervorgehen kann laut Literatur mit der „Coordination-Class-Theorie“ (diSessa & Sherin, 1998) erklärt werden. Genauer gesagt spricht DiSessa vom „Situational-Sampling p-prim“.

Die Kategorie „Äußere Einflüsse als Erklärung für Schwankungen“, aber auch die Kategorie „Messunsi-

cherheit kommt von menschlichen Fehlern“ liefernwertvolle Daten, welche das bereits bekannte p-prim „von nichts kommt nichts“, welches bei DiSessa als „Actuating agency“ (diSessa, 1993) bekannt ist bekräftigen.

Auch „Ohm´s p-prim“ konnte bestätigt werden. Dieses besagt: „viel hilft viel“, was mit Hilfe der Daten in der Kategorie „Der am häufigsten vorkommende Wert ist der Genaueste“ gezeigt wird. Auch wenn dieses p-prim im Kontext von Messunsicherheiten eher als „Häufigkeit gleich Wahrheit“ zu verstehen ist.

Neu definierte p-prims

Darüber hinaus konnten zwei p-prims identifiziert werden, welche in der Literatur noch nicht bekannt sind.

Auf Basis der erhaltenen Werte aus der Kategorie „Technik ist besser als der Mensch“ und der tiefergehenden Analyse dieser Daten im Kontext des Videos konnte ein neues p-prim „(digitale) Technik ist fehlerlos“ definiert werden. Vor allem beim Betrachten der Antworten zum Video ist klar zu erkennen, dass diese Intuition weit verbreitet ist.

Das zweite p-prim, welches bis dato noch nicht näher definiert wurde folgt aus den Daten der Kategorie „Der mittlere Wert ist der Beste“. Es lautet: „die Mitte ist am Besten“ und kann eventuell auch dem p-prim zur Balance zu Grunde liegen.

Limitationen

Die Stichprobe für die leitfadengestützten Interviews umfasst 26 Schüler:innen, wobei alle Interviews, abgesehen von drei, die in Gmunden durchgeführt wurden, an Wiener Schulen stattfanden. Somit ist die Abdeckung von unterschiedlichen Bundesländern, aber auch von städtlichen beziehungsweise ländlichen Schulen nicht ideal ausgeglichen. Zudem wurden viele der interviewten Personen von den Lehrkräften ausgewählt, was dazu führen kann, dass vorwiegend Schüler:innen ausgewählt werden, welche ein größeres physikalisches Interesse vorweisen. Auch dadurch wird die Allgemeingültigkeit dieser Studie ein Stück weit eingeengt.

Ein weiterer limitierender Faktor ist, dass die Intercoder-Reliabilität von internen Personen durchgeführt wurden. O’Connor und Joffe argumentieren, dass es effektiver wäre, diese Überprüfungen von externen Personen vornehmen zu lassen (vgl. O’Connor & Joffe, 2020).

Die Anwendung der Think-Aloud-Methode und das Erheben von p-prims haben ebenfalls ihre Einschränkungen, da die Schüler:innen oft Schwierigkeiten haben, ihre Gedanken klar und strukturiert auszudrücken. Zudem können nicht alle kognitiven Prozesse vollständig kommuniziert werden (von der Oelsnitz & Busch, 2008, S. 264). P-prims sind kleine und spezifische kognitive Elemente, die oft nur in bestimmten Situationen und für kurze Zeit aktiv sind. Sie können nicht bewusst erfasst und beschrieben

werden, da sie durch persönliche Erfahrungen geprägt sind. Daher können sie nicht im Voraus durch eine Theorie festgelegt werden (diSessa, 1993, S. 119).

Die wesentlichste Limitation ist jedoch der Fakt, dass die p-prims, welche möglicherweise neu entdeckt wurden, nur unter der Voraussetzung gefunden werden konnten, dass es zuvor zur Konfrontation von vermeintlich fehlerfrei messender Technik und den zufälligen Streuungen kommt. Diese „Stecknadel im Heuhaufen“ konnte gefunden werden, limitiert jedoch die Stichprobe drastisch.

Ausblick

Die Ergebnisse dieser Untersuchung liefern wertvolle Hinweise für zukünftige Entwicklungen im Bereich der Unterrichtskonzepte. Es ist zu erkennen, dass es einige Lernendenvorstellungen und p-prims in diesem Gebiet der Naturwissenschaften gibt, welche mit darauf abgestimmten Unterrichtskonzeptionen aufgegriffen werden sollten. Zudem liefert die Studie ansätze für weitergehende beziehungsweise vertiefende Forschungen im Bereich der Messunsicherheiten und der Lernendenvorstellungen. Eine genauere Auseinandersetzung mit der Frage, wie Schüler:innen mit den Konzepten von Messunsicherheit und Zufall umgehen, führt hoffentlich zu einem besseren Verständnis der Fehlvorstellungen und damit auch zu einer zielgerichteten Entwicklung von Unterrichtskonzepten.

Diese Entwicklung von neuen Unterrichtskonzepten beziehungsweise die Weiterentwicklung bereits vorhandener Konzepte (siehe (Bärenthaler-Pachner & Nagel, 2024) bzw. (Loidl & Nagel, 2024)) ist bereits im Gange und wird durch diese Studie hoffentlich noch weiter verbessert. Durch die oftmalige Durchführung von Messungen kann in weiterer Folge die Zufälligkeit von Messunsicherheiten thematisiert werden.

Zukünftige Studien könnten untersuchen, wie die Integration von praktischen Beispielen aus dem Alltag – wie etwa Messungen zur Wettervorhersage oder zur Bestimmung von Sportrekorden – den Schüler:innen helfen kann, das Konzept der Messunsicherheit besser zu begreifen und den Zufall als Teil von Messprozessen zu akzeptieren. Solche Alltagsbeispiele könnten eine Brücke schlagen zwischen abstrakten theoretischen Konzepten und den praktischen Anwendungen von Messungen.

Darüber hinaus können und sollen zukünftige Forschungen in diesem Gebiet versuchen einen größeren Stichprobenumfang zu erreichen um die erwähnten Limitationen zu verringern. Gerade der Ansatz mit dem Video eines Versuchs erscheint hierbei besonders vielversprechend zu sein, da er auch verhältnismäßig zeitsparend umgesetzt werden kann.

Durch eine frühzeitige und kontinuierliche Auseinandersetzung mit diesen Themen könnten Missverständnisse verringert und das wissenschaftliche Verständnis der Schüler:innen auf eine feste Grundlage gestellt werden.

Zusammenfassend lässt sich sagen, dass die Auseinandersetzung mit Messunsicherheiten und den damit verbundenen kognitiven Prozessen der Schüler:innen ein vielversprechendes Forschungsfeld darstellt, das

in zukünftigen Studien weiter vertieft werden sollte. Die in dieser Studie bestätigten und neu gefundenen p-prims sollen hierbei eine gute Basis schaffen, auf der diese weiterführenden Studien aufbauen können.

Abbildungsverzeichnis

1	Die drei Hauptkomponenten der Typ B Messunsicherheit (Nagel et al., 2021)	6
2	Häufigkeit der Kategorien	35
3	Anzahl der Interviews pro Kategorie	35
4	Anzahl der Interviews pro Kategorie	39

Anhang

Inhaltsübersicht Anhang

Transkripte	S. 49
Leitfaden Durchführende	S. 50
Leitfaden Schüler:innen	S. 55

Transkripte

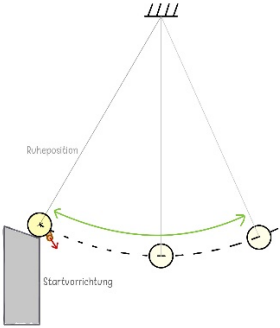
Die vollständigen Transkripte der geführten Interviews sind aus Platzgründen nicht direkt in dieser Arbeit abgedruckt. Sie stehen jedoch über den folgenden QR-Code zum Nachlesen zur Verfügung:



Scannen Sie den QR-Code und öffnen Sie den Link im Browser, um auf die Transkripte zuzugreifen.

Alternativ ist der Ordner auch über den folgenden Link erreichbar:

https://drive.google.com/drive/folders/1qfewNqzHQrN2hZmRsja72_26YuOV7djr?usp=sharing

Fragen und Input für die Schülerinnen und Schüler	Infos für die interviewende Person
1. Organisatorisches Dieses Interview soll als think aloud Interview durchgeführt werden. Das bedeutet, dass die interviewte Person laut denken soll und einfach alles, was ihr einfällt, sagen soll. Die Fragen sind als Unterstützung für den Interviewenden gedacht, damit die Information auch gesichert eingeholt wird. Gibt die Interviewte Person die Information von selbst preis, kann auf die Nachfragen verzichtet werden.	Vorzubereiten für das Interview: Ein Sekundenpendel (mit Startvorrichtung, wenn möglich) Ein Laptop (Mit dem das Video abgespielt werden kann) Eine Kamera mit Stativ zum Mitfilmen (kann auch das Handy sein)  Mikrometerschraube und Bolzen Allgemein: bei allen Aussagen der Testperson immer möglichst genau nachfragen, um das Vorstellungsbild so genau wie möglich reproduzieren zu können.
2. Eigener Versuch Versuchsbeschreibung: Mit Hilfe eines Stativs, einer Schnur und einem Gewicht, wird ein Pendel aufgebaut. Das Pendel soll immer gleich weit ausgelenkt werden und die Periodendauer mit Hilfe einer Stoppuhr (Handy) bestimmt werden. Führe den Versuch fünfmal durch und dokumentiere deine Ergebnisse. • Welche Messwerte hast du herausbekommen? • Denkst du, dass du die Messung richtig gemacht hast? • Warum glaubst du, dass unterschiedliche Werte herausgekommen sind? • Stört dich das? Oder ist das normal? • Denkst du, dass deine gemessenen Werte richtig sind? • Wieso denkst du das? • Welchen deiner Werte würdest du wählen, wenn du dich auf einen Wert festlegen müsstest? • Wieso würdest du diesen Wert wählen? • Machst du das immer schon so? Hast du das irgendwo gelernt?	Das Pendel ca. 5-6° auslenken und mit der Startvorrichtung fixieren. Danach mithilfe der Startvorrichtung bzw. ohne menschlichen Einfluss zum Schwingen bringen erst nach allen Messungen stoppen. Wie könnte man die Messung verbessern? Bei allen erwähnten Punkten möglichst oft nachfragen Alltagserfahrungen erfragen Auf potenzielle Schüler*innenvorstellungen achten Was passiert mit eventuellen Ausreißern?

Jetzt miss die Periodendauer noch fünf Mal.

- Inwiefern haben sich die Ergebnisse verändert?
- Würdest du, wenn du die beiden Messreihen zusammennimmst, jetzt einen anderen Wert wählen?
- Wieso würdest du das machen? /Wieso würdest du das nicht machen?
- Was denkst du wird passieren, wenn wir den Versuch noch ein paar Mal durchführen?

Weitere Messreihen:

n	1	2	3	4	5	6	7	8	9
t(s)	0,8	0,98	0,82	1,03	0,97	0,82	0,82	1,10	1,06

- Wie sieht es bei dieser Messreihe aus? Welchen Wert würdest du hier wählen?
- Wieso würdest du diesen Wert wählen?

n	1	2	3	4	5	6	7	8	9
t(s)	0,86	0,90	0,83	0,81	0,87	0,84	0,82	0,85	0,89

- Wie sieht es bei dieser Messreihe aus? Welchen Wert würdest du hier wählen?
- Wieso würdest du diesen Wert wählen?

n	1	2	3	4	5	6	7	8	9
t(s)	0,90	1,64	0,83	0,79	0,87	0,89	0,95	0,85	1,43

- Was fällt dir auf, wenn du diese Messreihe ansiehst?
- Wieso sind da manche Werte so weit weg von den anderen?
- Wie kommst du auf diese Idee? Hast du das schon mal irgendwo gesehen?
- Was denkst du macht man mit solchen Werten?

Fragen möglichst einfach halten (auch Nachfragen)

Bei erkannten Routinen/p-prims nachfragen, ob das immer so gemacht wird (z.B. dass der Mittelwert oder häufigste Wert gewählt wird)

Bei der Messreihe geht es darum zu beobachten, wie die SuS darauf reagieren, wenn ein Wert häufiger auftritt, die anderen Werte aber wesentlich größer sind.

Hier vermuten wir das p-prim: Je häufiger ein zufälliger Wert auftritt, desto richtiger/vertrauenswürdiger ist das Ergebnis.

Nachfragen!

Alltagserfahrungen erfragen

Bei dieser Messreihe sind die Zahlen sehr ausgeglichen, auch hier geht es darum herauszufinden, wie die SuS darauf reagieren.

Hier vermuten wir das p-prim: Der mittlere Wert ist der Beste.

Nachfragen!

Alltagserfahrungen erfragen

Hier geht es um das intuitive Verhalten bei Ausreißern.

Hier vermuten wir das p-prim: Nur ähnliches gehört zusammen/ Ausreißer gehören nicht dazu.

3. Video eines Versuchs

Vor der 2. Durchführung stoppen (0:48) und fragen:

Diese Stoppuhr kann 1000 Mal genauer messen als deine Handy Stoppuhr. Er verwendet eine Laserlichtschrankenmessung, das Auslösen und Stoppen der Messung passiert fast mit Lichtgeschwindigkeit (70%)

- Was glaubst du, kommt jetzt nochmal der gleiche Wert heraus?
- Warum denkst du das?

Video weiter abspielen bis 1:07

- Wie kann es deiner Meinung nach zu den unterschiedlichen Messwerten kommen?
- Ist das normal/ok?
- Wie kommst du auf diese Antwort? Hast du dazu schon etwas gehört oder gelernt?
- Wie würdest du dich für einen Wert entscheiden?
- Wieso entscheidest du dich für diesen Wert?

4. Messreihe

n	1	2	3	4	5	6	7	8	9	10	11	12
l (mm)	10,19	9,99	9,90	10,05	10,01	10,12	9,87	9,94	10,00	10,04	9,93	9,23

Es wurde immer derselbe Metallbolzen mit einer Mikrometerschraube gemessen.

- Auch hier haben wir wieder unterschiedliche Messwerte;
- Was glaubst du steckt da dahinter, wenn hier doch 12 Mal der selbe Bolzen gemessen wurde (immer an der gleichen Stelle, mit dem selben Messinstrument)
- Wie kommst du auf diese Antwort?
- Hast du dazu etwas im Unterricht gelernt?
- Wie gehst du damit um?
- Kannst du das in eigene Worte fassen?
- Hast du das in der Kindheit/im Alltag erlebt? Wenn ja, wo?
- Welchen Wert würdest du wählen, wenn du dich auf einen Wert fixieren müsstest?
- Wieso würdest du gerade diesen Wert wählen?



Genau erklären, dass durch diese Methode viel genauer gemessen wird und der Mensch keinen Einfluss darauf hat.

Nachfragen!

Fragen einfach halten (auch Nachfragen)

Erfahrungen aus dem Alltag?

Habe ich richtig verstanden, dass du das so meinst?

Mikrometerschraube und Bolzen herzeigen

SuS formulieren lassen

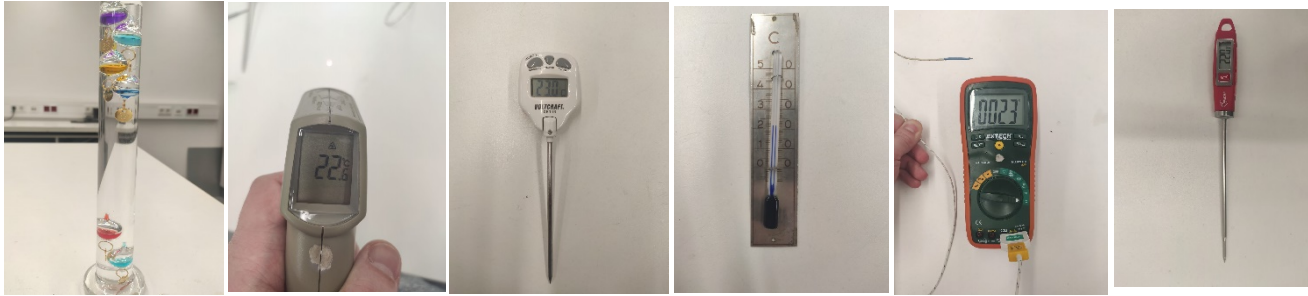
Nachfragen!

Alltagsbezug erfragen

Habe ich richtig verstanden, dass du das so meinst?

5. Thermometer

Wir haben hier mehrere verschiedene Arten von Thermometer, mit denen ich die Temperatur gemessen haben:



Was fällt dir auf, wenn du dir die Bilder ansiehst?

Wieso könnten da unterschiedliche Werte herauskommen?

Wie kommst du darauf?

Kannst du dich an eine (oder mehrere) Situationen im Alltag erinnern, wo du das schon erlebt hast?

Welchen Wert würdest du wählen?

Wieso würdest du diesen Wert wählen?

Wie gehen die SuS mit der Streuung der Werte um?

Nachfragen!

Alltagserfahrungen erfragen

Vertrauen SuS einer bestimmten Art von Thermometer mehr als anderen?

Nachfragen!

6. Allgemeiner Part (Vorwissen)

- Hast du schon mal etwas von einem Durchschnitt gehört?
- Kannst mir erklären was du darunter verstehst?
- Wo hast du davon gehört?
- Hast du den Durchschnitt schon mal berechnet?

- Hast du schon mal etwas von einem Median gehört?
- Kannst mir erklären was du darunter verstehst?
- Wo hast du davon gehört?
- Hast du den Median schon mal bestimmt?

- Hast du im Alltag schon einmal vom „Zufall“ gehört?
- Wo hast du davon gehört?
- Was hat dieser Zufall bewirkt?

- Hast du in der Schule schon etwas zum Thema „Zufall“ gelernt?
- In welchem Fach?
- Was genau hast du darüber gelernt?

- Was hältst du von der Idee, dass unsere Messwerte zufällig waren? Hältst du das für möglich?
- Wieso könnte das Zufällig sein (oder nicht)?

- Hast du in der Schule oder im privaten Bereich schon etwas von Quantenmechanik gehört?
- Wo hast du davon gehört?
- Was genau weißt du dazu?
- Gibt es bei der QM den Zufall?

- Hast du in der Schule oder im privaten Bereich schon einmal von radioaktiven Zerfallsprozessen gehört?
- Wo hast du davon gehört?
- Was genau weißt du dazu?
- Gibt es bei radioaktiven Zerfallsprozessen einen Zufall?

Alltagserfahrungen erfragen

Nachfragen!
Alltagserfahrungen erfragen

In der Oberstufe fragen!
In der Unterstufe nicht unbedingt notwendig.

Interview Masterarbeit

1. Eigene Messung

Mit Hilfe eines Stativs, einer Schnur und einer Masse, wird ein Pendel aufgebaut. Das Pendel hat eine Startvorrichtung, welche verhindern soll, dass man das Pendel unterschiedlich anstoßen kann. Wenn das Pendel schwingt, soll die Zeit zwischen zwei aufeinander folgende „Ruhelage-Durchgänge“ gemessen werden.

Führe den Versuch fünfmal durch und dokumentiere deine Ergebnisse.

Messwerte (s)

Wiederhole deine Messungen noch fünf Mal.

Messwerte (s)

2. Video



3. Messreihe

n	1	2	3	4	5	6	7	8	9	10	11	12
l (mm)	10,19	9,99	9,90	10,05	10,01	10,12	9,87	9,94	10,00	10,04	9,93	9,23

n	1	2	3	4	5	6	7	8	9
t(s)	0,8	0,98	0,82	1,03	0,97	0,82	0,82	1,10	1,06

n	1	2	3	4	5	6	7	8	9
t(s)	0,86	0,90	0,83	0,81	0,87	0,84	0,82	0,85	0,89

n	1	2	3	4	5	6	7	8	9
t(s)	0,90	1,64	0,83	0,79	0,87	0,89	0,95	0,85	1,43

4. Thermometer



Literatur

- Abbott, D. S. (2003). *Assessing student understanding of measurement and uncertainty* (Unveröffentlichte Dissertation). North Carolina State University.
- Abl, E. (2024). *Klärung des alltagsweltlichen konzeptes der vertrauenswürdigkeit (von messdaten) bei lernenden der sekundarstufe 1 und prüfung der langzeitwirksamkeit von darauf basierenden unterrichtskonzeptionen* (Diplomarbeit). (Zentrum für LehrerInnenbildung, Unterrichtsfächer Mathematik und Physik) doi: 10.25365/thesis.76909
- 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. doi: 10.1080/0950069980200405
- Altrichter, H., Posch, P. & Spann, H. (2018). *Lehrerinnen und lehrer erforschen ihren unterricht*. Verlag Julius Klinkhardt.
- Bortz, J. & Döring, N. (2006). Quantitative methoden der datenerhebung. In *Forschungsmethoden und evaluation: für human- und sozialwissenschaftler* (S. 137–293). Berlin, Heidelberg: Springer Berlin Heidelberg. Zugriff auf https://doi.org/10.1007/978-3-540-33306-7_4 doi: 10.1007/978-3-540-33306-7_4
- Buffler, A., Allie, S. & Lubben, F. (2001a). 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. doi: 10.1080/09500690110039567
- Buffler, A., Allie, S. & Lubben, F. (2001b). 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. doi: 10.1080/09500690110039567
- Bundesminister für Unterricht und Kunst. (1985, 14. 11). *Lehrpläne – allgemeinbildende höhere schulen* (Verordnung Nr. BGBl. Nr. 88/1985). <https://www.ris.bka.gv.at/eli/bgbl/1985/88/P0/NOR11008719>. (Verordnung des Bundesministers für Unterricht und Kunst vom 14. November 1984 über die Lehrpläne der allgemeinbildenden höheren Schulen; Bekanntmachung der Lehrpläne für den Religionsunterricht an diesen Schulen)
- Bärenthaler-Pachner, R. (2022). *Entwicklung und evaluation einer lernumgebung zum thema messunsicherheit in der sekundarstufe ii* (Master's Thesis). Universität Wien.
- Bärenthaler-Pachner, R. & Nagel, C. (2024). Development and evaluation of a gum-based curriculum on measurement uncertainty for upper secondary. *Progress in Science Education (PriSE)*, 7 (1), 56–77. Zugriff auf <https://doi.org/10.25321/prise.2024.1428> doi: 10.25321/prise.2024.1428
- Charters, E. (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education Journal*, 12 (2), 68–82. Zugriff auf <https://journals.library.brocku.ca/brocked/index.php/home/article/view/38>
- 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. doi: 10.1080/0263514980160107
- Davidowitz, B., Lubben, F. & Rollnick, M. (2001). Undergraduate science and engineering students'

- understanding of the reliability of chemical data. *Journal of Chemical Education*, 78 (2), 247. doi: 10.1021/ed078p247
- diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Hrsg.), *Constructivism in the computer age* (S. 49–70). Hillsdale, NJ: Lawrence Erlbaum Associates.
- diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, 10 (2-3), 105–225. doi: 10.1207/s1532690xci1002&3_2
- diSessa, A. A. (2018). A friendly introduction to “knowledge in pieces”: Modeling types of knowledge and their roles in learning. In G. K. et al. (Hrsg.), *Invited lectures from the 13th international congress on mathematical education* (S. 65–84). Springer. doi: 10.1007/978-3-319-72170-5_5
- diSessa, A. A. & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education*, 20 (10), 1155–1191. Zugriff auf <https://doi.org/10.1080/0950069980201002> doi: 10.1080/0950069980201002
- Eccles, D. W. & Aarsal, G. (2017). The think aloud method: what is it and how do i use it? *Qualitative Research in Sport, Exercise and Health*, 9 (4), 514–531. Zugriff auf <https://www.tandfonline.com/doi/full/10.1080/2159676X.2017.1331501> doi: 10.1080/2159676X.2017.1331501
- Fairbrother, R. & Hackling, M. (1996). Helping students to do open investigations in science. *Australian Science Teachers Journal*, 42 (4), 26–33.
- Flick, U., von Kardorff, E. & Steinke, I. (Hrsg.). (2000). *Qualitative forschung: Ein handbuch*. Reinbek bei Hamburg: Rowohlt.
- Friebertshäuser, B., Boller, H., Prengel, A. & Langer, A. (Hrsg.). (2013). *Handbuch qualitative forschungsmethoden in der erziehungswissenschaft* (4. Aufl.). Weinheim: Beltz Juventa.
- Gigerenzer, G. (2015). *Das einmaleins der skepsis: Über den richtigen umgang mit zahlen und risiken*. München: Piper Verlag.
- Goertz, S. & Heinke, H. (2021). Module zum umgang mit messdaten als unterstützungsangebot für lehrkräfte. In S. Bernholt (Hrsg.), *Inquiry-based learning – forschendes lernen. tagungsband zur frühjahrstagung der gdcp 2021* (S. 665–668). Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN).
- Heinicke, S. (2012). *Aus fehlern wird man klug: Eine genetisch-didaktische rekonstruktion des messfehlers*. Logos Verlag Berlin GmbH. Zugriff auf https://books.google.com/books?hl=en&id=pPN19uKfCpIC&oi=fnd&pg=PA1&dq=Susanne+Heinicke+%22Aus+Fehlern+wird+man+klug,+Eine+Genetisch-Didaktische+Rekonstruktion+des+Messfehlers%22&ots=-CcYx-NYXI&sig=8CR3X6xJeorsx15kAR2a0bPrR_0
- Honda, H., Kagawa, R. & Shirasuna, M. (2022). On the round number bias and wisdom of crowds in different response formats for numerical estimation. *Scientific Reports*, 12 (1), 8167. doi: 10.1038/s41598-022-11900-7
- Huber, O. (2019). *Das psychologische experiment. eine einföhrung*. Hogrefe Verlag.
- Ibrahim, B. B. (2006). *The relationship between views of the nature of science and views of the nature of scientific measurement* (Unveröfentlichte Dissertation). University of Cape Town.
- Krüger, D., Parchmann, I. & Schecker, H. (Hrsg.). (2014). *Methoden in der naturwissenschaftsdidakti-*

- schen forschung*. Berlin Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-37827-0
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.
- Loidl, H. (2021). *Entwicklung und evaluation von unterrichtseinheiten zum thema messunsicherheiten* (Masterarbeit, Universität Wien, Wien). (Betreuer: Univ.-Prof. Dr. Martin Hopf; Mitbetreuer: Mag. Dr. Clemens Nagel) doi: 10.25365/thesis.70616
- Loidl, H. & Nagel, C. (2024). Introduction to measurement uncertainties for middle school classes. In *Challenges in physics education*. Springer. (im Erscheinen)
- 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. doi: 10.1002/sce.1012
- Lubben, F. & Millar, R. (1996). Children’s ideas about the reliability of experimental data. *International Journal of Science Education*, 18 (8), 955–968. doi: 10.1080/0950069960180807
- Lüders, C. (1993). Grundlagen und methoden qualitativer sozialforschung: Ein Überblick über neuere veröfentlichungen. *Zeitschrift für Pädagogik*, 39 (2), 335–348.
- Majiet, N. & Allie, S. (2018). Student understanding of measurement and uncertainty: probing the mean. *Physics Education Research Conference Proceedings*, 264–267. (Available at <https://www.per-central.org/items/perc/4967.pdf>)
- Mayring, P. (2016). *Einführung in die qualitative sozialforschung: Eine anleitung zu qualitativem denken* (6. Aufl.). Weinheim: Beltz.
- Millar, R., Niedderer, H. & et al., V. T. (1998). *Survey 2: Students’ images of science as they relate to labwork learning* (Bericht). Brussels, Belgium: European Commission – Labwork in Science Education (LSE) Project. (Technical Report)
- Mittelstraß, J. (1996). Enzyklopädie philosophie und wissenschaftstheorie. In (Bd. 1, S. 855). Stuttgart: Metzler.
- Nagel, C. (2017). *Auswertung und dokumentation experimenteller daten. der umgang mit messunsicherheiten von der planung eines experiments bis zur publikation analysierter messdaten*. Zugriff auf https://moodle.univie.ac.at/pluginfile.php/3792555/mod_resource/content/6/AuswertungSkript.pdf (Skriptum zur gleichnamigen Vorlesung des Autors aus dem Sommersemester 2016 – Universität Wien)
- Nagel, C., Lux, B. & Steindl, S. (2021). Die thematisierung von messunsicherheiten im physikunterricht – eine umfrage. *Plus Lucis*, 1 (4), 4–6.
- O’Connor, C. & Joffe, H. (2020). Intercoder reliability in qualitative research: debates and practical guidelines. *International Journal of Qualitative Methods*, 19.
- Pelz, R. & Schnögl, T. (2014). Vertrauenswürdigkeit – teil 1: Qualitative forschung. *Methoden in der naturwissenschaftsdidaktischen forschung*.
- 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, 647–659. doi: 10.1088/0143-0807/29/3/024

- Pollard, B., Hobbs, R., Stanley, J. T., Dounas-Frazer, D. R. & Lewandowski, H. J. (2017). Impact of an introductory lab on students' understanding of measurement uncertainty. *arXiv preprint arXiv:1703.00868*. Zugriff am 2025-04-20 auf <https://arxiv.org/abs/1703.00868>
- Popper, K. (1934). *Logik der forschung, nachdruck der 10. auflage (2002)*. Mohr Siebeck Verlag.
- Posner, G. J., Strike, K. A., Hewson, P. W. & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66 (2), 211–227. doi: 10.1002/sce.3730660207
- 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. doi: 10.1023/A:1013102108541
- Schecker, H., Wilhelm, T., Hopf, M. & Duit, R. (Hrsg.). (2018). *Schülervorstellungen und physikunterricht: Ein lehrbuch für studium, referendariat und unterrichtspraxis*. Springer Spektrum. doi: 10.1007/978-3-662-57270-2
- Sere, M.-G., Journeaux, R. & Larcher, C. (1993). Learning the statistical analysis of measurement errors. *International Journal of Science Education*, 15 (4), 427–438. doi: 10.1080/0950069930150406
- Sill, H.-D. & Kurtzmann, G. (2012). *Didaktik der stochastik in der primarstufe*. Wiesbaden: Springer Spektrum.
- Smith III, J. P., DiSessa, A. A. & Roschelle, J. (1994). Misconceptions re conceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3 (2), 115–163.
- van Someren, M. W., Barnard, Y. F. & Sandberg, J. A. (1994). *The think aloud method: A practical approach to modelling cognitive processes*. London: Academic Press. Zugriff auf https://pure.uva.nl/ws/files/716505/149552_Think_aloud_method.pdf
- von der Oelsnitz, D. & Busch, M. W. (2008). Die think-aloud-methode. *Die Betriebswirtschaft*, 68 (2), 260–265. Zugriff auf <https://www.econbiz.de/Record/die-think-aloud-methode-oelsnitz-dietrich-von-der/10007992214> doi: 10.1007/s11628-008-0013-0
- 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. doi: 10.1080/095006999290318
- Weßling, B. (2022). *Was für ein zufall! Über unvorhersehbarkeit, komplexität und das wesen der zeit*. Wiesbaden: Springer Vieweg.